## Human Y chromosomes and the origins of modern European populations

Thesis submitted for the degree of Doctor of Philosophy at the University of Leicester

by
Zoë H Rosser (BSc, MSc)
Department of Genetics
University of Leicester
November 2000

All rights reserved

## INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.
In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.


UMI U139992
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code.


ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346

Ann Arbor, MI 48106-1346

This work is dedicated to my brother, Simon F. Lenehan, who we miss everyday.

12. 1. 1969-30.3. 1994

# Human Y chromosomes and the origins of modern European populations 

Zoë H. Rosser


#### Abstract

The $Y$ chromosome is haploid, paternally inherited, and the majority of the chromosome does not undergo recombination at meiosis, and so it contains a comparatively simple record of its past, which tells us about our male ancestors. Contained within its non-recombining portion are different polymorphisms with different mutation rates, including biallelic markers with low mutation rates, and multiallelic markers with higher mutation rates; the combination of these allows evolutionary events on different timescales to be examined. Three different studies examining Y-chromosomal diversity have been undertaken: a large continent-wide study examining the distribution of biallelic haplogroup diversity within 48 European and circum-European populations; a micro-geographic study examining the haplogroup distribution and diversity within a linguistic isolate of the Italian Alps; and finally, a global study of a single lineage, employing biallelic and multiallelic markers, to try to explain the unusual global distribution of this lineage.

Within the European study, five haplogroups showed significant clinal patterns. Two major continent-wide clines are consistent with the migration of farmers from the Near East. Statistical analyses confirm that populations are related primarily on the basis of geography rather than language. The second study shows that Ladin-speakers have low haplogroup diversity when compared to neighbouring populations however, a high level of internal diversity was observed using the multiallelic markers. Minisatellite MSY1 examination shows unique alleles within Ladin valleys which suggests either in situ differentiation of an isolated population, or a small, already diverse founding population. Within the final study additional biallelic markers were incorporated using a new typing methodology, enabling the further sub-division of the YAP branch, and some new haplogroups showed population specificity. Increasing the number of microsatellite loci from seven to seventeen allowed improved discrimination within the haplogroups. The dating of mutations enabled historical events to be proposed to explain the unusual geographical distribution of this lineage.


## Acknowledgements

I would like to thank my parents and Matt for their emotional and financial support during all my studies, particularly during the PhD. I feel it has been invaluable. To all my friends, I would like to thank them for constantly asking me when am I going to get a proper job!

I would especially like to thank Mark for his support, eternal patience and good humour during my time in his lab, and for listening to many hours of practice talks. I would also like to thank him for his encouragement for me to attend international conferences (even if it meant giving a talk!), particularly in places where the shopping was so good!

I would like to say a big thank you to the other members of the $Y$ group, in particular Matt Hurles, for being a fun person to share a bench with, and for spending a huge amount of his time helping me with statistical analyses! And thanks to the newer members, Elena, Turi and Andy, who all make the lab a really enjoyable place to work in.

Thanks to the members of lab 136 (too many to mention by name), who have all generated a great deal of humour, and amusement during my time in Leicester. To all the members of G19, a big thank you, in particular to Celia, who has endless patience with regards to me asking her questions on a variety of things. The Department of Genetics was a great place to spend (a total of 4 years) 3 years during my PhD , because the people are so friendly and welcoming. I feel I have made some extremely good friends and will really miss you all.

## Abbreviations

| APS | Ammonium persulphate |
| :---: | :---: |
| ASD | Average squared distance |
| ASO | Allele-specific oligonucleotide |
| bp | base pairs |
| BPB | Bromophenol blue |
| BSA | Bovine serum albumin |
| DNA | Deoxyribonucleic acid |
| DTT | Dithiothreitol |
| dATP | 2'-deoxy adenosine 5'-triphosphate |
| dCTP | 2'-deoxy cytosine 5'-triphosphate |
| dGTP | 2'-deoxy guanosine $5^{\prime}$-triphosphate |
| dTTP | 2'-deoxy thymidine 5'-triphosphate |
| EDTA | Ethylene-diamino-tetra-acetic acid |
| hg | Haplogroup |
| kb | kilobase |
| MRCA | Most recent common ancestor |
| MVR-PCR | Minisatellite variant repeat mapping using PCR |
| PCA | Principal components analysis |
| PCR | Polymerase chain reaction |
| RFLP | Restriction fragment length polymorphism |
| SAAP | Spatial autocorrelation analysis program |
| SDS | Sodium dodecyl sulphate? |
| SNP | Single nucleotide polymorphism |
| TBE | Tris-borate EDTA |
| TEMED | $\mathrm{N}, \mathrm{N}, \mathrm{N}^{\prime}, \mathrm{N}^{\prime}$-Tetramethylethylenediamine |
| TMAC | Tetramethylammonium chloride |
| Tris | Tris (hydroxymethyl) aminomethane |

## Contents

Page
Chapter 1: The human $Y$ chromosome as a tool to examinediversity within modern populations
1.1. General introduction ..... 1
1.2. Evolution of modern humans ..... 1
1.2.1. Australopithecines ..... 2
1.2.2. Homo habilis ..... 2
1.2.3. Homo erectus ..... 2
1.2.4. 'Archaic Homo sapiens' ..... 3
1.2.5. Homo neanderthalensis ..... 3
1.2.6. Homo sapiens ..... 4
1.3. Two hypotheses proposed to explain the origins of modern humans ..... 4
1.4. Linguistics as a tool ..... 5
1.5. Genetics ..... 6
1.5.1. Autosomal DNA ..... 6
1.5.2. Mitochondrial DNA ..... 7
1.5.3. $X$ and $Y$ chromosomes ..... 8
1.6. Structure of the $Y$ chromosome ..... 9
1.6.1. Human $Y$-chromosomal diversity ..... 9
1.6.2. Y-chromosomal genes ..... 11
1.6.3. Y-chromosomal biallelic polymorphic markers ..... 11
1.6.4. Microsatellites ..... 13
1.6.5. Minisatellites ..... 14
1.7. A genealogical approach to Y -chromosomal diversity ..... 16
1.7.1. Trees of Y -chromosomal multiallelic haplotypes ..... 16
1.7.2. The dating of DNA lineages ..... 17
1.7.3. Assigning a root ..... 21
1.8. Genetic evidence for place of origin ..... 21
1.9. Ancient and modern DNA comparisons ..... 22
1.10. The origin and relationship between modern populations ..... 23
1.11. The scope of this thesis ..... 23

## Chapter 2: Materials and Methods

### 2.1. Materials

2.1.1. Chemical and molecular biology reagents ..... 24
2.1.2. Oligonucleotides ..... 24
2.1.2.1. Fluorescently labelled oligonucleotides ..... 24
2.1.3. Genomic DNA samples ..... 24
2.1.4. Commonly used solutions ..... 24
2.2. Methods
2.2.1. Extraction of DNA from blood ..... 26
2.2.1.1. GuSCN purification method using a silica suspension ..... 26
2.2.1.2. QIA amp blood mini kit ..... 27
2.2.2. Precipitating oligonucleotide primer stocks ..... 28
2.2.2.1. Re-hydrating lyophilised oligonucleotide primers ..... 29
2.2.3. DNA amplification ..... 29
2.2.4. Primer annealing temperature ..... 29
2.2.5. Restriction enzyme digestion ..... 29
2.2.6. Agarose gel electrophoresis ..... 30
2.2.7. Biallelic analysis ..... 30
2.2.8. Allele-specific oligonucleotide analysis ..... 30
2.2.9. Y-chromosome specific microsatellite analysis ..... 30
2.2.10. Minisatellite variant-repeat mapping via PCR ..... 30
Chapter 3: A low resolution, continental study, examining
Y chromosome diversity within European and circum-
European populations
3.1. Introduction
3.1.1. The development and dispersal of agriculture across Europe ..... 31
3.1.2. Physical geography and palaeoclimatology ..... 33
3.1.2.1. The definition of populations used within this study ..... 34
3.1.3. European linguistic comparison ..... 34
3.1.4. Previous European genetic studies ..... 36
3.1.5. Previous Y-chromosomal studies ..... 38
3.1.6. Questions to be addressed ..... 39
3.2. Materials and Methods
3.2.1. Genomic DNA samples ..... 40
3.2.1.1. DNA samples typed at Leicester ..... 40
3.2.1.2. DNA samples typed elsewhere ..... 41
3.2.2. Experimental procedure for biallelic typing ..... 41
3.2.3. Biallelic markers ..... 42
3.2.4. Statistical analysis ..... 43
3.2.4.1. Principal components analysis ..... 43
3.2.4.2. Spatial autocorrelation analysis ..... 44
3.2.4.3. Mantel analysis ..... 45
3.2.4.4. Genetic barrier analysis ..... 46
3.3. Results
3.3.1. General observations ..... 49
3.3.2. Clinal distribution of $Y$-chromosomal lineages within Europe ..... 50
3.3.3. A major cline appears to be consistent with the demic diffusion hypothesis ..... 52
3.3.4. A north east to south west cline may represent a signal of an expansion from north of the Black Sea ..... 53
3.3.5. A north south cline may represent influence from North African populations ..... 54
3.3.6. A haplogroup that appears localised in the north east of Europe, may reflect the influence of Uralic speakers ..... 54
3.3.7. Which has the strongest affect on European Y-chromosomal genetic differentiation, geography or language? ..... 55
3.3.7.1. Population comparisons through principal components analysis ..... 55
3.3.7.2. Correlating geography, language and genetics through Mantel testing ..... 57
3.3.7.3. Locating Y-chromosomal genetic barriers within Europe ..... 58
3.4. Discussion
3.4.1. General discussion ..... 61
3.4.2. Clines of $Y$-chromosomal haplogroups ..... 62
3.4.3. Language, geography and $Y$-chromosomal diversity ..... 64
3.4.4. Conclusion ..... 66

## Chapter 4: Micro-geographic study examining the genetic diversity of linguistic isolates of the Italian Alps

4.1. Introduction
4.1.1. Linguistic isolates ..... 67
4.1.2. The Ladin-speaking population ..... 69
4.1.3. Mitochondrial DNA analysis ..... 70
4.1.3.1. Models proposed to explain the high level of genetic diversity within the Ladin speakers ..... 70
4.1.4. Linguistic comparisons ..... 72
4.1.5. Questions to be addressed ..... 73
4.2. Materials and Methods
4.2.1. Genomic DNA samples ..... 75
4.2.2. Biallelic analysis ..... 75
4.2.3. MSY1 flanking DNA amplification via PCR ..... 75
4.2.4. MSY1 flanking PCR conditions ..... 75
4.2.5. Minisatellite variant repeat mapping by PCR (MVR-PCR) ..... 76
4.2.6. Radioactively end labelling of PCR primers ( ${ }^{33} \mathrm{P} \gamma$-ATP) ..... 76
4.2.7. Three-state Minisatellite Variant Repeat mapping by PCR ..... 77
4.2.8. MVR-PCR conditions ..... 77
4.2.9. Polyacrylamide gel preparation ..... 78
4.2.10. Acrylamide gel running conditions ..... 79
4.3. Results
4.3.1. Biallelic analysis ..... 80
4.3.2. Examination of haplotype diversity ..... 81
4.3.3. Statistical analysis of the biallelic data ..... 82
4.3.3.1. Principal components analysis ..... 82
4.3.3.2. Spatial autocorrelation analysis ..... 82
4.3.3.3. Genetic barrier analysis ..... 83
4.3.4. Microsatellite analysis ..... 83
4.3.5. MSY1 results ..... 84
4.3.5.1. Unusual haplogroup 1 MSY1 modular structures ..... 85
4.3.5.2. Unusual haplogroup 9 MSY1 modular structures ..... 85
4.3.5.3. Unusual haplogroup 21 MSY1 modular structures ..... 85
4.3.6. Examination of different MS1Y lineages within the Ladin speakers ..... 86

### 4.4. Discussion

4.4.1. General observations 87
4.4.2. Haplogroup examination 87
4.4.3. Multiallelic analysis 87
4.4.4. Conclusions 88

## Chapter 5: A high resolution, global study of a single lineage, YAP (Y Alu polymorphism), and the implications for human evolution

### 5.1. Introduction

### 5.1.1. Repetitive elements <br> 89

5.1.2. Alu elements as polymorphic markers ..... 91
5.1.3. The YAP polymorphism ..... 91
5.1.4. The YAP lineage global distribution ..... 91
5.1.4.1. Diversity within YAP+ lineages ..... 92
5.1.5. The Bantu expansion ..... 94
5.1.6. Asian migrations ..... 95
5.1.7. Previous population studies ..... 96
5.1.8. Questions to be addressed ..... 96
5.2. Materials and Methods
5.2.1. Genomic DNA samples ..... 97
5.2.2. Markers typed ..... 97
5.2.3. Allele-specific oligonucleotide typing ..... 97
5.2.3.1. Dot blotting ..... 98
5.2.3.2. Preparing dot blots ..... 98
5.2.3.3. Denaturing PCR products and generating the dot blots ..... 98
5.2.3.4. Oligonucleotide labelling ..... 99
5.2.3.5. Radioactive labelling of allele-specific oligonucleotides ..... 99
5.2.3.6. Preparation of hybridisation solutions ..... 99
5.2.3.7. Hybridisation protocol ..... 100
5.2.3.8. Removing probes and re-using membranes ..... 101
5.2.4. Sexing genomic DNA samples ..... 101
5.2.5. Y chromosome specific microsatellite analysis ..... 101
5.2.5.1. Microsatellite multiplex systems ..... 101
5.2.5.2. Microsatellite amplification protocol ..... 102
5.2.5.3. Polyacrylamide gel preparation ..... 103
5.2.5.4. Preparation of amplified DNA products ..... 104

### 5.3. Results

5.3.1. Biallelic and allele-specific oligonucleotide analyses ..... 106
5.3.2. The positioning new markers on the maximum parsimony tree ..... 106
5.3.3. Continental specificity of the $Y$-chromosomal haplogroup data ..... 108
5.3.4. Microsatellite analysis of YAP+ chromosomes ..... 111
5.3.4.1. Additional microsatellite data ..... 111
5.3.5. Distribution of microsatellite lengths ..... 112
5.3.5.1. The CTS multiplex ..... 112
5.3.5.2. The MS1 multiplex ..... 112
5.3.5.3. The EBF multiplex ..... 113
5.3.6. The examination of pairwise differences within haplogroups ..... 113
5.3.7. Examination of microsatellite diversity at each locus ..... 113
5.3.8. The dating of mutations defining haplogroups from microsatellite data ..... 114
5.3.9. Generation of phylogenetic networks ..... 120
5.3.10. Does increasing the number of loci increase the separation of samples within the networks? ..... 120
5.3.10.1. Generation of networks from seven microsatellite loci ..... 121
5.3.10.2. Generation of networks from seventeen microsatellite loci ..... 123
5.3.11. Assessing the power of assigning haplogroups from only microsatellite data ..... 125
5.4 Discussion
5.4.1. The inclusion of new markers within the tree ..... 126
5.4.2. Examination of diversity within biallelic haplogroups using microsatellite loci ..... 126
5.4.3. Problems dating lineages and polymorphisms ..... 127
5.4.4. The dating of $Y$-chromosomal mutations ..... 128
5.4.5. Deductions about the geographical distribution of YAP+ lineages ..... 129
5.4.6. Future work ..... 129
Chapter 6: General Discussion
6.1. The $Y$ chromosome ..... 130
6.2. Sampling ..... 130
6.3. Markers ..... 131
6.3.1. Ascertainment bias of markers ..... 131
6.3.2. A hierarchical approach to typing - some disadvantages ..... 131
6.3.3. A need for parallel methods of typing ..... 131
6.3.4. New Y-chromosomal markers now available ..... 132
6.4. A new study of European Y-chromosomal diversity ..... 132
6.5. Dating ..... 133
6.6. The Human Genome sequence - what will be its effect? ..... 133
6.7. Selection ..... 134
6.8. Surnames and the $Y$ chromosome ..... 134
References ..... 135

## Appendices

Appendix 1 - actual haplogroup data generated for all European samples analysed at the University of Leicester
Appendix 2 - complete biallelic and ASO data for YAP+ chromosomes
Appendix 3 - complete microsatellite for all YAP+ chromosomes
Appendix 4 - re-print of the Rosser et al., 2000 paper

## Chapter 1: The human $Y$ chromosome as a tool to examine diversity within modern populations

### 1.1. General introduction

This thesis uses the diversity of the haploid, paternally inherited $Y$ chromosome to examine aspects of the histories of human populations. The three different $Y$-chromosomal polymorphic systems used during this work (biallelic, microsatellites and the minisatellite MSY1), can be employed to examine diversity within different time scales. Within this chapter an introduction to what is known about human pre-history will be given, as well as the two hypotheses proposed to explain the evolution of modern humans from a number of possible different ancestral hominid species. Genetic tools for examining human histories will be considered in general terms, including a description of the different components of the human genome, and finally the genetic evidence taken to support these two hypotheses will be discussed. Introductions to more detailed aspects of the work can be found in each chapter.

### 1.2. Evolution of modern humans

Human evolution has always been of interest to numerous areas of research-not only genetics, but palaeontology, archaeology, and linguistics. Each of these will be discussed with respect to human evolution, but first a brief overview is given of what is known and what is proposed for the development of modern humans. Following this brief description, this area of human evolution will be mentioned infrequently during this thesis.

Hominids encompass all primates including modern humans, chimpanzees and gorillas (Jones et al., 1999). Distinct differences in teeth, skull, brain and body morphology between modern humans and African apes are clearly visible (Wood, 1996). It is proposed that a common ancestor of modern humans and chimpanzees lived between 5 and 8 million years ago (Horai et al., 1992), but chimpanzees show more nuclear genome diversity than humans (Crouau-Roy et al., 1996). This may be due to the modern human population expanding from a small population quite recently, or a major bottleneck reducing the effective population size (Jorde et al., 1998). Additional analysis argues against the latter explanation, because the sharing of alleles from the major histocompatibility complex (MHC), $\beta$-globin and apolipoprotein C-II has been
observed between chimpanzees and humans which would have been lost during a bottleneck event (Ayala, 1995b). Fossil evidence and other artefacts have led to a great deal of knowledge about ancestors of modern humans (Homo sapiens), however some of the suggested connections between the proposed ancestors (Figure 1.0) are contentious among anthropologists and archaeologists alike.

### 1.2.1. Australopithecines

Most of the species discussed belong to the genus Homo with the exception of the more ancestral Australopithecines, with the best known female skeletal example ('Lucy') being discovered during the 1970's. The majority of fossil evidence originates around Ethiopia, with some additional evidence coming from Tanzania. A number of species are thought to have co-existed between three and four million years ago including Australopithecus africanus and Australopithecus afarensis, i.e. 'Lucy'. Another species known as Australopithecus bahrelghazali discovered in northern central Africa (Tchad) has been dated to approximately 3.5 million years old (Wood, 1996). The bone structure suggests the ability to stand upright (Jones et al., 1999), and walk in a bipedal manner (Wood, 1996), but these creatures would have appeared smaller than modern humans (Jones et al., 1999).

### 1.2.2. Homo habilis

This species is considered by some as an early direct ancestor to modern humans, and the name means 'handy man' (Wood, 1996; Jones et al., 1999) or the 'maker of tools' (Wood, 1996). This species lived approximately 2.4 million years ago, in eastern and southern Africa, with fossils being found mainly around Tanzania; it has been proposed that this species may represent an intermediate between Australopithecines and Homo sapiens. Archaeological evidence suggests that this species was the first to manufacture and use stone tools in everyday life (Jones et al., 1999).

### 1.2.3. Homo erectus

Homo erectus ('erect man') is considered another direct ancestor to modern humans, and this species lived approximately 1.8 million years ago in Africa, and archaeological artefacts suggest geographic locations including Algeria, Morocco, Ethiopia, Kenya, South Africa and Asia. These appear more widespread than the previous ancestor (Homo habilis)


Figure 1.0. Hominid species
The broken line reveals the proposed links between the different species. Some links are only putative and indicated by a question mark. Figure adapted from Wood, 1996.
(Jones et al., 1999) and were the first hominid species detected outside Africa (Wood, 1996). Asian examples include Java man and Peking man, with a proposed later Java example, known as Ngandong that has been proposed to fit into the models suggested to explain the expansion of modern humans (see Figure 1.1). Archaeological evidence also suggests that stone and wood flaked tools as well as hand axes were used by Homo erectus (Jones et al., 1999).

### 1.2.4. 'Archaic Homo sapiens'

'Archaic Homo sapiens' lived approximately 400,000 years ago on the African, Asian and European continents. A number of species may have existed, known as Homo heidelbergensis, named after the jaw bone found in Heidelberg (Germany), and Homo rhodesiensis (not shown in Figure 1.0), after a skull discovered in (Rhodesia) Zambia but this is again controversial (Jones et al., 1999).

### 1.2.5. Homo neanderthalensis

Archaeological evidence indicates that Neanderthals lived between 120,000 and 30,000 years ago in western Europe, and evidence from the Iberian peninsula suggests that they survived for approximately 5,000 to 10,000 years after the arrival of modern humans (Mellars, 1998). There is no current evidence to demonstrate they lived in Africa or Asia (Jones et al., 1999). One of the most famous features of a Neanderthal was the marked brow ridge, and it is thought their body proportions (larger and more stocky than modern human ancestors) resemble modern Saami and Inuit populations, which may indicate they were adapted to cold climates (Mellars, 1998). Neanderthals are traditionally considered to be extremely primitive, but recent evidence suggests they were highly evolved, engaged in cultural rituals such as burying their dead, and using tools of varied designs (Wood, 1996). They made bone and teeth body ornaments (Mellars, 1998) which has been taken to demonstrate a level of communication (Bahn, 1998).

### 1.2.6. Homo sapiens

The emergence of modern humans (Homo sapiens) and the decline and disappearance of Neanderthals has been hotly debated, and one proposed connection between these phenomena may have been their
intense competition for resources. Homo sapiens differ from other early hominids in particular Australopithecines in a number of ways: the brain was larger; jaws and teeth were smaller; and the skeletal shape of Australopithecines indicates that they moved like modern baboons, but also were capable of bipedal walking. In comparison, modern humans walked continually upright, and it has been proposed that modern human infants rely on parents for longer periods of time than the early hominids (Wood, 1996), although there is no real evidence to support this. A famous archaeological find at Cro-Magnon (France) in 1868 indicated a less prominent brow ridge, a more rounded skull and a shorter lower jaw. Homo sapiens inhabited Asia and possibly Europe approximately 40,000 years ago. Artefacts discovered indicate they used materials such as bone, antler and ivory to make blade tools, and also engravings, sculptures, wall and cave paintings have been discovered, which suggest they were highly intelligent, organised, and socially orientated.

### 1.3. Two hypotheses proposed to explain the origins of modern humans

Two opposing hypotheses have been proposed and widely debated, to explain the origins of modern humans. These are the Multiregional and 'Out of Africa' hypotheses (Figure 1.1), and both agree that ancient human ancestors originated on the African continent, but disagree about the origins of our more recent ancestors.

The Multiregional model suggests that there was no single origin for all modern humans (Wolpoff, 1989; Wolpoff, 1992). Following the radiation of Homo erectus from Africa into Europe and Asia approximately 800,000 to 1.8 million years ago (Stringer, 1992), there was a continuous transition among regional populations from Homo erectus to Homo sapiens. Such 'parallel evolution' among geographically dispersed populations could have been achieved by considerable amounts of gene flow, between populations (Wolpoff, 1989; Wolpoff, 1992) as well as migration to other areas. The lack of transitional fossils detected in China and Australia linking archaic and modern humans, has been taken to indicate that little or no local evolution occurred, supporting the 'Out of Africa' hypothesis (Jones et al., 1999).

The 'Out of Africa' model suggests that all modern populations descended from an anatomically modern Homo sapiens ancestor that lived in Africa
approximately 100,000 to 200,000 years ago, and spread and diversified throughout the rest of the world, replacing any existing archaic Homo populations still present outside Africa (Wolpoff, 1989; Swisher III et al., 1994). Under this model, racial differences observed today developed post Homo erectus. The presence of transitional fossils in South Africa and Ethiopia, dated to between 100,000 and 70,000 years ago has been taken to support this 'Out of Africa' hypothesis (Jones et al., 1999).

### 1.4. Linguistics as a tool

Some culturally transmitted traits can be used to examine the relationship between human populations, and the most commonly used is language. The underlying assumption is that closely related populations will be more likely to have similar languages than distantly related ones. However, it is important to note that the time-depth of human linguistic relationships is not great - at most 10,000 years (Renfrew, 2000), and so language has nothing to say about the debate on the origins of anatomically modern humans. Some languages can be completely replaced, particularly after military invasions (Cavalli-Sforza et al., 1994) as when the Magyars conquered Hungary in approximately the 9th century AD, and imposed the Hungarian language spoken today (Ruhlen, 1991). Linguistics will be described in more detail during the introduction to Chapter 4. Similarities between languages have been noted for years, and comparisons of such properties allow hypothetical ancestral languages to be constructed known as proto-languages (Ruhlen, 1991). Unfortunately little is known about protolanguages for the majority of modern languages, because written documentation is often poor.

Two Russian linguists proposed probably the most famous macrofamily 'Nostratic', which encompasses Indo-European, Afro-Asiatic, Dravidian, Altaic and Uralic (Cavalli-Sforza et al., 1988; Renfrew, 1994). However, the construction of such macrofamilies remains controversial amongst linguists. It was also suggested that a proto-Nostratic language was spoken in the Middle East approximately 15,000 years ago (Renfrew, 1994), and this has been associated with the dispersal of agriculturalists from the Near East, called the 'Nostratic demic diffusion'. Autosomal gene frequencies generated suggested farming spread with at least three of the Nostratic languages (Barbujani and Pilastro, 1993). A more detailed discussion of the dispersal of agriculture will be described within Chapter 3.

### 1.5. Genetics

Within these next sections the contribution of genetics to human origin studies will be described, including a detailed description of the components of the human genome and in particular the Y chromosome, as it forms the basis for this work. The human genome comprises three thousand megabases ( Mb ) of DNA per haploid cell, and this is divided between the nuclear and mitochondrial genomes. The human diploid karyotype has 46 chromosomes, which are divided into 44 autosomes and two sex chromosomes (XY in males and XX in females). There are unusual regions that unlike the rest of the genome, do not undergo recombination at meiosis, the majority of the Y chromosome, and mitochondrial DNA (mtDNA) (Ayala, 1995a).

### 1.5.1. Autosomal DNA

Autosomal DNA undergoes the process of recombination, which re-shuffles genetic information each generation. Gene frequencies have been used to determine population relationships, and initial studies included the electrophoretic analysis of blood groups (ABO) and serum proteins (e.g. rhesus blood system) (Cavalli-Sforza et al., 1994). Molecular genetic evidence taken from additional autosomal loci has been used to indicate the relationships between populations and the origin of modern humans.

For example the minisatellite MS205 (Armour et al., 1993) has been used to make deductions about human evolution - supporting the Out of Africa hypothesis (Armour et al., 1996), and a similar picture emerges from diversity analysis at the CD4 gene locus (Tishkoff et al., 1996). Analysis of polymorphic Alu repeats within humans and non-human primates showed the ancestral state observed within the latter was the absence of Alu insertions (Batzer et al., 1994). Some data have been taken to support the opposing Multiregional hypothesis; for example, the analysis of polymorphisms within the $\beta$-globin gene revealed the most diversity within African populations, suggesting an African origin for this locus, although ancient Asian haplotypes were also observed (Harding et al., 1997). Detailed examination of the major histocompatibility complex (MHC) indicated that most variation was outside the African continent
(Ayala et al., 1994). The genetic evidence taken to support the different hypotheses will be discussed in more detail in a later section.

### 1.5.2. Mitochondrial DNA

The mitochondrion is an organelle involved in cellular energy production and is found in multiple copies in cells. Within each mitochondrion are contained 5 to 10 copies of a circular (mitochondrial) DNA molecule of 16.5 kb . Two important features of mtDNA are that it lacks recombination at meiosis and is maternally inherited and so can provide information about our female ancestors (Stoneking and Soodyall, 1996). It has been shown that paternal mtDNA from the sperm enters the egg during fertilisation (Ankel-Simons and Cummins, 1996), but apparently does not contribute any mtDNA to the zygote. Due to their haploid nature, and lack of recombination, all modern mtDNAs coalesce (or lead back) to a single ancestor, 'mitochondrial Eve', who originated in Africa (Cann et al., 1987). This conclusion was later supported by sequence data (Horai et al., 1995). The time taken to reach the most recent common ancestor (MRCA) is essentially dependent on the effective population size and the mutation rate (Jorde et al., 1998). Dates suggested from coalescence analysis (explained in a later section) include 120,000 to 150,000 years (Ward and Stringer, 1997), and 200,000 years (Cann et al., 1987; Cavalli-Sforza, 1991). This does not suggest that only one woman lived then, but rather the other mtDNA lineages went extinct (CavalliSforza, 1991), and this is a simple consequence of the variance in number of offspring. Mitochondrial DNA has a mutation rate about ten times higher than nuclear DNA, leading to high diversity at this locus that can be examined via RFLP analysis (Cann et al., 1987). Two hypervariable sequences (HVS-I and -II) of the non-coding D-loop have an even higher mutation rate (Parsons et al., 1997), and in other studies examination of these sequences allowed additional information to be added to the mtDNA haplotypes (Stoneking, 1993; Jorde et al., 1998).

The analysis of mtDNA polymorphisms has been widely adopted to examine evolutionary relationships, although there are problems with this system (Stoneking and Soodyall, 1996). The main types of polymorphism analysed are base substitutions, or small insertion/deletions, but the high mutation rate means the base
substitutions could have occurred more than once (recurrent). This can cause problems (seen as reticulations) when drawing trees (Bandelt et al., 1995). Another problem related to the high mutation rate is the determination of ancestral state from chimpanzee mtDNA sequence. Chimpanzee DNA is too diverged to be a suitable outgroup, and alternatives include the Neanderthal mtDNA sequence (Krings et al., 1997) and nuclear mtDNA insertions (Zischler et al., 1995) which have proved better for out-group rooting. Some individuals within pedigrees have more than one mitochondrial allele, which is called heteroplasmy (Ivanov et al., 1996), a state which can persist for many generations (Sykes, 1999). The occurrence of sequencing errors can be a serious problem when examining any portion of the human genome, but such a mistake led to the proposal that mtDNA did in fact recombine (Hagelberg et al., 1999), which was later retracted (Hagelberg et al., 2000). The analysis of mtDNA routinely involves the combined use of control regions and additional polymorphism assayed throughout the genome, (Sykes, 1999).

### 1.5.3. $X$ and $Y$ chromosomes

Comparative studies show that the two sex chromosomes, $X$ and $Y$ (Figure 1.2), were once homologous, but the Y now consists largely of repeated sequences and pseudogenes, and has lost material with respect to the $X$. Males are heterogametic (XY) compared to females that are homogametic ( XX ), but sex determination requires a dosage compensation mechanism or males would only express half the complement of X-linked genes as females (Graves, 1995). This is carried out by the inactivation of one of the X chromosomes in a female. X chromosome inactivation is a random process (Lyon, 1961), and the inactivated $X$ can be visualised as a structure called a Barr body. However, not all $X$ genes are inactivated, because some have active $Y$-homologues, for example RPS4X/Y. Certain genes on the $Y$ have evolved a function in male sex determination ( $S R Y$ ) and spermatogenesis - the AZF (azoospermia factor) genes (Graves, 1995). The absence of Y-linked diseases may be due to selection, because if they were deleterious and appeared before the reproductive age, it is unlikely they would be passed to subsequent generations (Jobling and Tyler-Smith, 2000).


Figure 1.2. Idiograms of G-banded human sex chromosomes ( $\mathbf{X}$ and $\mathbf{Y}$ ) The banding pattern of the Y chromosome is after Magenis et al., 1985, and the pseudoautosomal and non-recombining regions are also indicated.

### 1.6. Structure of the $Y$ chromosome

The Y chromosome comprises 60 Mb of linear DNA, and is divided into two sections known as the pseudoautosomal (PARs) and Y-specific regions (Ellis, 1991). It is haploid, paternally inherited, and the majority of the chromosome (excluding the PARs) does not undergo recombination at meiosis (Ellis, 1991), and so this can provide genetic information about male ancestors, without recombination re-shuffling the genetic information between generations. About half of the Y chromosome, comprising the heterochromatic block on distal Yq (see Figure 1.2), is composed of polymorphic sequences which are organised into large interspersed repeated arrays (Cooke, 1976). Two short homologous regions (PAR1 and 2) at the tips of the X and Y chromosomes (Pearson and Bobrow, 1970; Solari, 1980; Cooke et al., 1985; Simmler et al., 1985; Schmidt et al., 1994) undergo homologous recombination, and contained within these two regions are a number of genes (Ciccodicola et al., 2000), which will not be further discussed.

### 1.6.1. Human $Y$-chromosomal diversity

Assuming a 1:1 sex ratio, the number of $Y$ chromosomes in any population is one-quarter that of any one autosome, making Y-linked sequences more prone to genetic drift (Kayser et al., 1997) and leading to the potential loss of some the $Y$ chromosomes. The spread of an individual $Y$ chromosome can occur by chance, but is likely to be greatly influenced by factors such as selection, male migration and reproductive behaviour (Whitfield et al., 1995a). Mating practices such as polygyny enable a small number of males to father a disproportionately large number of offspring, and the frequency of their $Y$ chromosomes increases rapidly (Whitfield et al., 1995a). However, not all fathers produce male children, and because of this, some $Y$ chromosomes will die out over time, therefore if all modern lineages were traced back, they would eventually coalesce back to a single ancestor.

The dating of the MRCA (most recent common ancestor) of modern $Y$ chromosomes has been attempted by comparison between humans and apes, using the human-ape divergence times as reference points. Generally accepted dates for the MRCA have been similar to those for mtDNA - 188,000 years ( $95 \%$ CI 51,000 to 411,000 years) (Hammer, 1995), and 162,000 years ( $95 \%$ CI 69,000-316,000 years). An estimated age of 186,000 years ( $95 \%$ CI 77,000-372,00 years) ( Fu and $\mathrm{Li}, 1997$ ) has also been
suggested; and similar dates to Hammer (1995) were proposed by Underhill et al. (1997). A recent estimate is younger than these, at approximately 59,000 years ( $95 \%$ CI 40,000-140,000 years) (Thomson et al., 2000), but this could be a reflection of the population genetic model used (exponential population growth). When a constant population model was used, an older age of 84,000 years ( $95 \%$ CI $55,000-149,000$ years) was estimated (Thomson et al., 2000).

The Y chromosome is more geographically differentiated than are mtDNA or autosomal loci (Seielstad et al., 1998). This has been examined by comparing how genetic distances vary with geographical distance (Seielstad et al., 1998). Y-chromosomal genetic differentiation is affected by a number of factors including population history, the mutation rate and effective population size (Nei, 1987; Stoneking, 1998). The smaller the effective population size, the greater the effect drift can have on allele frequencies, leading to a more rapid divergence. It has been proposed that the mtDNA (Jorde et al., 1998) and Y chromosomes had similar effective population sizes (5000) (Hammer, 1995). The relatively high geographical differentiation of $Y$ chromosomes has been explained by a higher female migration rate, influenced by practices such as patrilocality (Seielstad et al., 1998). This is where females are more likely to move than males, from their birthplace upon marriage, and this may be because it is easier for women to cross cultural boundaries (Poloni et al., 1997). This suggests that the mtDNA has moved the greatest distances across the global landscape during successive generations, and this may have reduced the geographical differentiation at this locus. The geographical differentiation of the Y chromosome is visualised by population specificity of certain lineages (Jobling et al., 1997).

Selection is another factor that may to have a dramatic effect at nonrecombining loci, such as mtDNA and on the $Y$, which would obscure the true relationships between populations. It has been proposed that with no selection (neutral theory), and an effective population size of 5000, the expected MRCA would be approximately 200,000 years ago (Hammer, 1995), which fits well with initial estimates (Hammer, 1995; Underhill et al., 1997). However, newer estimates are much younger, which might reflect either selection, or differences in the population modelling. The
earlier proposed date has not been confirmed using two of the different polymorphic marker systems available on the $Y$ (see below), biallelic (Underhill et al., 1997; Hammer et al., 1998; Shen et al., 2000) and microsatellites (Goldstein et al., 1996). If positive selection, via a hitchhiking effect, allowed certain Y-chromosomal lineages to reach fixation within a population, this would result in the loss of all the other diversity contained within the extinct lineages. This would result in an early date for the MRCA (Whitfield et al., 1995b; Shen et al., 2000; Thomson et al., 2000). Positive selection has been briefly discussed; there is also the possibility of negative selection shown as mutations in certain Y chromosomal genes which may lead to varying degrees of sub-fertility. If certain haplotypes predispose to infertility, they may be underrepresented in populations (reviewed by Jobling and Tyler-Smith, 2000).

### 1.6.2. $Y$-chromosomal genes

A range of genes are contained on the Y chromosome (Figure 1.3), and these include the Testis Determining Factor (TDF) also known as sexdetermining region $Y(S R Y)$ which dictates which gonads develop, ovary or testis, from the indifferent gonad structure during early development. The SRY gene has a functionally diverged X-linked homologue, SOX3 (Lahn and Page, 1999). Additional genes required for spermatogenesis (Burgoyne, 1982; Weissenbach et al., 1987; Ellis and Goodfellow, 1989; Ellis, 1991; Graves, 1995; Whitfield et al., 1995a) also lie on the Y. The implication from the $X$ and $Y$ gene similarities is that the sex chromosomes must have evolved from a homologous pair of autosomes in a vertebrate ancestor. This may have occurred due to a large single rearrangement or gradual degradation of the $Y$ following a local suppression of recombination to protect the sex-determining genes (Graves, 1995; Lahn and Page, 1999).

### 1.6.3. $Y$-chromosomal biallelic polymorphic markers

The non-recombining region of the $Y$ chromosome has different polymorphic systems characterised by different mutation rates, and these include biallelic markers that can be considered 'unique', due to their extremely low mutation rate (approximately $10^{-8}$ per base per generation) (Shen et al., 2000; Thomson et al., 2000). In contrast, recurrent polymorphisms occur when the same allele arises independently more


Key: Y-specific; XY-homologous: $Y$-autosome-homologous
Expression: testis-specific; tooth-bud-specific; brain-specific; ubiquitous

Figure 1.3. Genes and phenotypes on the non-recombining region of the $Y$ chromosome The $Y$ chromosome is divided into its eight basic intervals (Page, 1986), and the positions of the two pseudoautosomal regions (PARs), in which recombination with the $X$ chromosome occurs, are indicated. Correspondence between intervals and G-bands on the idiogram is schematic only. Genes are colour coded according to whether they have homologues elsewhere, and copy number (' $m$ ' = multiple). $S R Y$ is indicated as XY homologous because of its homology to SOX3 (Lahn and Page, 1999). Gene families lacking open reading frames (TTY1, TTY2; Ref. (Lahn and Page, 1997) have been omitted. DFFRY is also known as USP9Y (Sun et al., 1999). *Copy number for RBMY refers to the entire Y chromosome, since some copies exist outside interval 6 (Chai et al., 1998). (Figure adapted from the Jobling and Tyler-Smith, 2000 review).
than once (Santos and Tyler-Smith, 1996). Biallelic markers have an ancestral and a derived allelic state, that can be determined for the majority of markers by sequencing homologous regions in the DNA of non-human primates (chimpanzees, gorillas and orangutans). The allelic states of the biallelic markers can be combined to generate a unique maximum parsimony tree (Mathias et al., 1994), comprising monophyletic compound haplotypes (haplogroups), and if all the ancestral states are known the tree can be rooted. (Figure 1.4 shows an example of such a tree).

Y-chromosomal haplogroups show geographical and population specificity (see Figure 1.4); for example, some haplogroups are only detected within Sub-Saharan (hg 7) (Hammer and Bonner, 1994; Karafet et al., 1999) or Japanese populations (hg 5) (Hammer and Horai, 1995). In a case such as hg 5, it is most likely that the defining mutation occurred in the location where its derived allele is most frequent, and spread only locally (Santos and Tyler-Smith, 1996).

Due to the small numbers of Y-chromosomal biallelic markers, initial analysis employing them as an evolutionary tool was limited, but this situation has changed with the publication of more than 400 biallelic markers on the Y (Underhill et al., 1997; Underhill et al., 2000). These major advances were aided by technology such as DHPLC (denaturing high performance liquid chromatography) developed by Peter Oefner and coworkers (Stanford University, USA). This process relies on the detection of a single nucleotide polymorphism (SNP) difference within the same region of two pieces of DNA (heteroduplex), which results in one strand of the DNA forming a 'bubble' under near-denaturing conditions, eluting from a HLPC column earlier than a homoduplex, and so generating a peak that is detectable.

### 1.6.4. Microsatellites

Microsatellites or STR (simple tandem repeat) loci, consist of units ranging from $2-5 b p$ in length, which can be tandemly repeated up to 30 times. Different allele lengths are generated by variation in the number of the repeats along the microsatellite array. Microsatellite variation is thought to be generated via stepwise mutation mechanisms possibly

$\bigcirc$ Haplogroups

Figure 1.4. An example of a Y-chromosomal maximum parsimony tree Individual haplogroups (hg) are indicated as yellow circles and each is defined by a separate polymorphism, with the exception of haplogroups 7 and 3, which are defined by a recurrent mutation. This tree is defined by Y-chromosomal markers and some haplogroups show geographical and population specificity, for example hg 7 and hg 5 .
involving replication slippage (Heyer et al., 1997; Kayser et al., 1997). Microsatellites are dispersed throughout the human genome, with one approximately every 10kb (Deitrich et al., 1992; Weissenbach et al., 1992), and have been characterised in many different species among mammals, birds, fish, and plants (Lagercrantz et al., 1993). In humans these loci are a useful tool for mapping disease-associated genes (Weber and Wong, 1993), and simple sequence regions have been implicated in triplet expansion disorders (Bates and Lehrach, 1994) which result from massive expansions of predominantly (CAG)n (Hancock, 1996).

Microsatellites also exist on the Y chromosome, and have some useful applications, including the ability to detect and discriminate male DNA in forensic casework, which makes them an important addition to the well-established autosomal PCR-based systems (Kayser et al., 1997). A set of seven, mostly tetranucleotide repeat loci have been widely adopted. When a number of different microsatellite loci are examined, the allelic lengths can be combined into microsatellite haplotypes, which can be used to construct networks (or unrooted trees). These are used for inferring relationships between human populations (Di Rienzo et al., 1994), and can be used to reconstruct human demographic histories (Shriver et al., 1997; Di Rienzo et al., 1998). It has been observed that alleles differ between population groups but in general, populations from the same continent appear to resemble each other at a microsatellite level (Barbujani et al., 1997). It has also been suggested that $Y$-specific microsatellites are equally as diverse as their autosomal counterparts (Roewer et al., 1992; Goldstein et al., 1996), which confirms the idea that mutation processes at these loci are intra-allelic (e.g. slippage), rather than recombination-based (Heyer et al., 1997; Kayser et al., 1997).

Within a haplogroup, microsatellite haplotype diversity is restricted compared to all of Y-chromosomal variation, since each haplogroup was founded by a single male, carrying zero microsatellite haplotype diversity. As time passes, diversity accumulates. The use of microsatellite loci to estimate the ages of certain Y-chromosomal lineages is now common; a number of different dating methods are available, and one important factor is the determination of a mutation rate. Different mutation rates for human autosomal tetranucleotide repeat loci have been estimated from
pedigree analysis, and these range from $0.015 \%$ (Jin et al., 1994), to $0.21 \%$ (Weber and Wong, 1993). A comparable mutation rate of $0.20 \%$ ( $95 \% \mathrm{CI}$ $0.06 \%$ to $0.49 \%$ ) for $Y$ chromosome tetranucleotide repeat loci has been estimated (Heyer et al., 1997), on the basis of haplotypes within deeprooting pedigrees. An average mutation rate of $3.17 \times 10^{-3}(95 \%$ CI 1.89$4.94 \times 10^{-3}$ ) was determined from direct analysis of father-son pairs, although additional rates were presented for eight tetranucleotide and fifteen loci respectively (Kayser et al., 2000). While identical microsatellite haplotypes tend to indicate membership of the same haplogroup, and often population, the high mutation rate of $Y$ microsatellites will lead to identical haplotypes sometimes being found across the world and between haplogroups due to recurrent mutations and not because of direct descent (Heyer et al., 1997).

### 1.6.5. Minisatellites

The human genome contains more than 1500 dispersed hypervariable minisatellites or VNTR (variable number tandem repeat) loci, which can show extreme levels of allelic variability (Jeffreys et al., 1985). Repeat unit lengths at these loci are between 9bp and 100bp, and repeat copy number can vary between approximately 10 and 1000 repeats. Repeat units vary in sequence, and this level of variation can be assessed using MVR-PCR (Jeffreys et al., 1991). The most unstable minisatellites are GC-rich (Armour and Jeffreys, 1992) and exhibit mutation rates of up to $13 \%$ in the male germline (Vergnaud et al., 1991), with the preferential gain of repeat units at one end of the tandem repeat array (Armour and Jeffreys, 1992; Jeffreys et al., 1994), which suggests that minisatellites could be recombination hot spots (Jeffreys et al., 1985; Jeffreys et al., 1991). This has been confirmed for at least one locus, by direct recombination analysis (Jeffreys et al., 1998).

GC-rich minisatellites are concentrated towards the telomeres of human chromosomes (Royle et al., 1988), and hence might be expected to be scarce in the non-recombining region of the $Y$; they are known to be frequent in the pseudoautosomal regions (Cooke et al., 1985; Fretwell, 1996). Furthermore, since they appear to be by-products of the process of recombination (Jeffreys et al., 1998), they may be absent from nonrecombining regions. Consistent with this, an extensive search for such
loci on the non-recombining region of the Y chromosome failed to find any (Fretwell, 1996). Two variable minisatellites have been detected on the $Y$ chromosome, and both are atypical - MSY1 (Jobling et al., 1994; Jobling et al., 1998a), and MSY2 (Bao et al., 2000).

MSY1 was the first minisatellite to be discovered and consists of 60 to 100 copies of a 25 bp AT-rich ( $75-80 \%$ ) unit that exists in at least ten forms (Jobling et al., 1998a) which have the potential to form nearly perfect hairpin structures (Jobling et al., 1998a). Highly diverged MSY1 structures have been detected in African-specific haplogroups suggesting they may represent deep-rooting branches in the $Y$ chromosome tree, and this lends some support for a recent African origin for modern humans (Jobling et al., 1998a). The overall level of virtual heterozygosity (i.e. the probability of two randomly selected alleles being different from each other) detected at the MSY1 locus was $99.9 \%$, and work from pedigree analysis suggests a mutation rate of 6 to 11\% (Jobling et al., 1998a). An MVR-PCR (minisatellite variant repeat mapping via PCR) system was developed to examine the internal diversity within MSY1, which involved the amplification of a flanking piece of DNA including the minisatellite, followed by repeat-specific PCR leading to the generation of a complete code for the minisatellite (Jobling et al., 1998a). This system will be described in detail within Chapter 4.

MSY2 does not exhibit such a large range of diversity as observed with MSY1, with only three or four copies of a repeat varying from 99 to 110 bp in length. The most common length of allele is four repeats ( $95 \%$ of chromosomes), but a two-repeat allele appears to be ancestral, as this has been detected in both chimpanzees and orang-utans. The three-repeat allele was identified in Africa and eastern Asia, with the highest frequency being detected in China (Bao et al., 2000).

One advantage of haploid Y-chromosomal minisatellites over autosomal ones is that the single allele can be internally mapped directly via MVR-PCR without having to separate it from another allele (see Chapter 4 - materials and methods). MSY1 alleles can be mapped completely, allowing the examination of allele length diversity and internal structure within specified haplogroups. The majority of repeats
are of three types, and occur as blocks within the minisatellite array. The order of these blocks is known as the modular structure, and correlates well with biallelic-defined haplogroups (Bouzekri et al., 1998; Hurles et al., 1998; Jobling et al., 1998a; Hurles et al., 1999). The MSY1 codes exhibit a limited allele length distribution (Jobling et al., 1998a), in comparison to the autosomal locus MS32 where a large range of allele lengths (repeats) has been observed (Jeffreys et al., 1990), and this limited distribution is thought to be due to restricted mutational processes. However, MSY1 exhibits an extremely high level of internal structural diversity (from 465 chromosomes, 386 have different codes) suggesting an extremely high mutation rate. Some polarity at one end of the allelic array has also been observed at MSY1 in diversity studies, which is similar to autosomal minisatellites.

### 1.7. A genealogical approach to $Y$-chromosomal diversity

The three polymorphic types of marker (biallelic, microsatellites and minisatellites) allow the Y chromosome to be extremely useful when examining evolutionary relationships between populations. These markers are usually typed in a hierarchical, or 'genealogical' way (Richards et al., 1997), in that biallelic markers are initially employed to define Y-chromosomal haplogroups, followed by subsequent intra-haplogroup diversity investigation using the faster mutating, microsatellite (de Knijff et al., 1997) and minisatellite loci (Hurles et al., 1999). This approach allows the maximum information to be extracted, whether it be for purposes such as individual identification, or to examine population diversity (Jobling and Tyler-Smith, 1995; Mitchell and Hammer, 1996; Santos and Tyler-Smith, 1996).

### 1.7.1. Trees of $Y$-chromosomal multiallelic haplotypes

Haplotypes generated from multiallelic marker information (Cooper et al., 1996; Roewer et al., 1996), can be used to examine relatively recent diversity within lineages (Jobling and Tyler-Smith, 1995; Malaspina et al., 1998) due to their higher mutation rates. This diversity can be displayed graphically by either summarising the data through methods such as principal components analysis, or by identifying important evolutionary links between the haplotypes. Recurrent mutations can be problematic (Cooper et al., 1996), and cause serious problems when trying to form these into networks (unrooted) or trees (rooted) (Rogers et al.,
1996). The structures of networks can indicate historical events such as population splits, and can be drawn by hand or by computer programs, and have become commonly employed (see Chapters 4 and 5). These programs cannot discriminate between recurrent mutation (identity by state) and non-recurrent mutation (identity by descent).

A single step mutation model is usually assumed in these studies (Cooper et al., 1996; Malaspina et al., 1998). Multi-step mutations are sometimes observed in pedigree studies (Kayser et al., 2000), but the single-step model is a good approximation. This model underlies various network-construction methods. The process behind minimum-spanning networks (MS) relies on a basic principle that links all haplotypes initially by a unit distance, which gradually increases until all haplotypes are included (Zerjal et al., 1997). All diversity should be contained within such a network, and the final network output is thought to contain all possible parsimonious trees. Minimum spanning networks contain only observed haplotypes. Finally, median-joining (MJ) networks (Bandelt et al., 1995; Bandelt et al., 1999) have been widely adopted in mtDNA studies (Forster et al., 1996; Richards et al., 1996) because they allow the reduction of reticulation within the data set, which mtDNA is prone to due to the high level of recurrent mutations (Bandelt et al., 1995). This process involves identifying unobserved nodes that shorten the overall length of branches within the tree, and is now also being used in Y studies (Forster et al., 2000; Helgason et al., 2000). A further process 'reduces' the MJ network to remove reticulations (reduced MJ network).

### 1.7.2. The dating of DNA lineages

The dating of tree branch points (lineages) is important when trying to interpret geographic distributions with respect to population history. The diversity within a lineage can be related to the time since the most recent common ancestor (MRCA) of the sampled chromosomes. A number of different approaches are available to assign ages to haplotypes, and these will be discussed in the following sections.

Coalescent analysis has been used to date the ages of biallelically defined alleles at many loci (Harding et al., 1997), including those on the $Y$ chromosome (Hammer et al., 1998), without any within-haplogroup
diversity data. The structure of the tree is important, and the observed frequency of alleles can be used to approximate their age. This method uses estimates of sequence divergence rates that depend on fossil record calibrations. Unfortunately, large confidence limits are assigned to coalescent dates, mainly due to the reliance upon the divergence time within the fossil record. However coalescence analysis can provide information about effective population sizes. For mtDNA the time in generations ( $t$ ) is related to the effective population size, $N_{e}$ (using the equation $t=2 N_{e}$ ); and dates proposed range from 100,000 to 200,000 years ago, very similar to the date proposed for the Y chromosome. This led to suggestion that the effective population size of humans has been relatively small, only a few thousand (reviewed by Jorde et al., 1998). If the effective population size was large, the date expected to reach the MRCA would be longer than that for a smaller population.

Alternatives to coalescent analysis employ multiallelic data to deduce the ages of haplogroups. Large confidence limits are also assigned here, and are mainly due to uncertainty about mutation rates (Thomas et al., 1998). Inaccurate mutation rates have led to problems when undertaking dating calculations, and drawing inferences from the results (Sykes, 1999). The three main approaches will be briefly described in turn.

The average (mean) number of mutants method makes the assumption that the average number of mutations from the root haplotype is equal to the mutation rate multiplied by the time since the first appearance of the ancestral haplotype. This method has been applied to microsatellite haplotype data incorporating a single-step mutational model, and was used initially to date the origin of the commonest cystic fibrosis mutation within Europeans (Bertranpetit and Calafell, 1996). This method requires a root haplotype be identified, and the average number of mutational steps from all haplotypes to the root is then averaged over all loci. This value known as $\rho$ (rho) (Forster et al., 1996) is divided by the mutation rate ( $\rho / \mu=t$ ) to give the age of the MRCA in generations, or in other words the date the lineage was founded (Bertranpetit and Calafell, 1996). This method was applied to determine the age of Y-chromosomal hg 3, a central Asian lineage, to approximately 3,800 years (1600-13,000 years) (Zerjal et al., 1999). Another lineage (hg 22) was examined and a date of

2,693 years ( $1,154-9,425$ years) was estimated for the $S R Y-2627$ mutation which defines this haplogroup (Hurles et al., 1999).

The average squared distance (ASD) method uses the equation ( $\mathrm{ASD}=\mu \mathrm{t}$ ), and calculates the squared mutational distance between a root haplotype and any other haplotypes within the lineage, averaged over all loci and all haplotypes (Goldstein et al., 1996). This value is divided by the mutation rate, leading to a date to the MRCA for that particular lineage in generations, and was used to date a Jewish priest-specific lineage (Thomas et al., 1998). A slightly adapted method has been used for dating population splits, and corrects for within-population diversity. The equation used is ( $\left.\delta \mu^{2}=2 \mu \mathrm{t}\right)$, and if the generation time to MRCA is to be determined this must be divided by two times the mutation rate (Goldstein et al., 1995). The basic principles involved are that the ASD incorporates the linear distance from the root to all haplotypes (unidirectionally) and the variance detected is due to age. However, ( $\delta \mu^{2}$ ) allows the distance from population A to the MRCA, and the distance from MRCA to population $B$, and so the distance is considered twice.

The third method, known as variance dating, uses a model to predict how the variance at microsatellite loci increases with time, which will depend on the effective population size, and mutation rate. This method does not need a root to be specified, and relates the variance of microsatellite allele lengths within a lineage to time (Goldstein et al., 1996). However, this method is extremely complex and not used very often. This method contains its own internal confidence intervals, which has the effect of broadening date ranges even further.

Generation time and mutation rates are important values in all dating methods discussed above. The proposed mutation rates for 7 Y chromosomal microsatellites been calculated from deep rooting pedigrees to be $2.1 \times 10^{-3}$ ( 95 CI 0.6 to $4.6 \times 10^{-3}$; Heyer et al., 1997), compared to newly estimated rates of $3.17 \times 10^{-3}\left(95 \%\right.$ CI $\left.1.89-4.94 \times 10^{-3}\right)$ generated from eight tetranucleotide loci, and $2.8 \times 10^{-3}\left(95 \%\right.$ CI 1.72-4.27×10 $0^{-3}$ ) from fifteen loci studied (Kayser et al., 2000). Errors on dates calculated have a number of sources, and while a major source is mutation rate, the generation time is another little discussed source of error. It is accepted that generation time
for males is longer than for females; the values used vary between 20 to 30 years, although 27 years (based on Weiss and von Haeseler, 1996) has been routinely adopted by some (Cavalli-Sforza et al., 1994; Underhill et al., 1996). Longer generation times ( 35 years) have been proposed from the analysis of well-documented genealogies (Tremblay and Vézina, 2000) and have been applied in other studies (Helgason et al., 2000), although the influence of these values and their relevance to prehistorical societies remains unknown.

Comparisons between the different dating methods are problematic because the confidence intervals are so large (Bosch et al., 1999). If the confidence intervals were ignored, it appears that ages from intra-allelic diversity are younger than those from coalescent dating. For example, microsatellite diversity within haplogroup 16 gives an age of about 4,000 years (Zerjal et al., 1997), while coalescent analysis suggests an age of 8,400 years (Karafet et al., 1999) (confidence intervals are $\pm 7,000$ years!). Possible reasons for such discrepancies between dating methods might relate to the differences in their underlying population and mutational models.

Following the determination of the age of a lineage, its evolutionary relevance can be considered. One important issue is the habit of comparing the age of a lineage with a population age which has been widely attacked as failing to appreciate the differences between a population and a lineage MRCA (Cavalli-Sforza and Minch, 1997). The age of the MRCA of a single allele is susceptible to stochastic effects and is a poor estimator of population age, and in general alleles are older than the populations in which they are found. A number of Y-chromosomal studies have used such lineage ages to exclude one of a number of competing hypotheses for a lineage's distribution (Hurles et al., 1998; Hurles et al., 1999). Some studies have recently applied all four methods on the same microsatellite data (Hurles et al., 1999), and the dates vary from 2,693 years $(1,154-9,425)$ for the mean mutations method, to 3,452 years ( $1,480-12,083$ ) for the ASD method, to 3,116 years $(1,166-16,001)$ for the variance method; and finally 1,650 years ( $1,044-8,248$ ) from coalescence dating (Hurles et al., 1999). As suggested previously, the coalescence method appears to suggest much younger dates compared to microsatellite dating methods.

### 1.7.3. Assigning a root

Non-variance-based dating methods rely on a root being assigned to a particular lineage which may or may not exist within the data. Sometimes the most highly represented haplotype is assigned as the root, which assumes the true root is the most frequently observed (Thomas et al., 1998), and this is not always the case. The root can be a combination of all the observed haplotypes into a common unobserved haplotype, and this is known as the modal haplotype. Another possibility is using diversity data from an ancestral (outgroup) lineage, to find the most parsimonious link between them, and so identify the root haplotype in the network of the derived lineage (Zerjal et al., 1997) although this is not favoured.

### 1.8. Genetic evidence for place of origin

The two hypotheses, 'Out of Africa' and Multiregional agree that human ancestors originated on the African continent, but disagree about the date of the MRCA. While most genetic evidence supports the Out of Africa hypothesis, there are still doubts - for example, the greater the diversity within Africa could reflect a larger long term effective population size, rather than origins (Templeton, 1997). African populations appear to be the most diverse, and nonAfrican diversity tends to show a subset of the African diversity (Armour et al., 1996; Templeton, 1997). This has been established from a range of genetic systems, including mtDNA evidence, that can be traced back to a single ancestor that lived in Africa between 100,000 and 300,000 years ago (Cann and Lum, 1991; Nei and Livshits, 1989). Autosomal minisatellites such as MS205 (Armour et al., 1993), also revealed that African variation is greatest, and non-African variation is a subset of that found in Africa (Armour et al., 1996). A similar picture emerges from diversity analysis at the CD4 gene locus (Tishkoff et al., 1996). Other loci such as the Y-chromosomal minisatellite (MSY1) (Jobling et al., 1998a), autosomal microsatellites, autosomal RFLPs, dinucleotide microsatellites (Bowcock et al., 1994; Jorde et al., 1997; Pérez-Lezaun et al., 1997a), blood groups and protein polymorphisms (Cavalli-Sforza et al., 1988) also support a recent African origin for modern humans (Shriver et al., 1997). Analysis of polymorphic Alu repeats within humans and non-human primates showed that non-human primates lacked Alu insertions proving this was the ancestral state, and most diversity is observed within Africans (Batzer et al., 1994). However, in all of these studies the
greater diversity may be due to a larger prehistoric long term population size (Templeton, 1997).

Examination of diversity at two other loci, the $\beta$-globin gene and the major histocompatibility complex (MHC), have been taken to support the opposing, Multiregional hypothesis. A 3 kb region of the $\beta$-globin gene was analysed and these data revealed that the greatest diversity was exhibited within African populations, suggesting an African origin for this locus, although ancient Asian haplotypes were also observed (Harding et al., 1997). However, differences between African and Asian samples could be due to the larger African effective population size. The MRCA was African in origin and dated to around 800,000 years ago (Harding et al., 1997). The majority of variation observed at several loci within the major histocompatibility complex (MHC) is outside the African continent (Ayala et al., 1994). These data appears more compatible with the Multiregional rather than the 'Out of Africa' hypothesis because an effective population size of more than 10,000 across the speciation event was proposed (Ayala et al., 1994). However, this difference has also been proposed to be due to the role of selection in preserving certain allelic lineages (Satta et al., 1994).

### 1.9. Ancient and modern DNA comparisons

Modern DNA samples can be examined to determine population histories but analysis of ancient DNA is also possible. This work has concentrated mainly on mtDNA extracted from bones, because there has been no reliable amplification of ancient nuclear DNA, including the Y chromosome. Extreme care must be taken when extracting mtDNA from either ancient or forensic samples (skeletal remains) because of the possibility of contamination from modern DNA sequences. Multiregionalists believe that Neanderthals were our ancestors, but mtDNA sequence evidence suggests there has been no maternal influence from Homo neanderthalensis (Krings et al., 1997; Ward and Stringer, 1997; Ovchinnikov et al., 2000).

### 1.10. The origin and relationships between modern populations

A brief summary of different types of DNA evidence that have been taken to support some population relationships, and specific examples will be discussed in later chapters. DNA variation can be used to determine regional population histories, because closely related populations will share some of the same polymorphisms, at a number of different loci. As mentioned previously,
mtDNA and the Y chromosome are extremely useful, due to their haploid nature and lack of recombination, with regards to determining information about ancestral populations. Evidence from mtDNA and nuclear DNA suggests that in general human populations are quite similar to each other, when compared to non-human primates (Nei and Graur, 1984; Stoneking et al., 1990; Takahata, 1991; Takahata, 1993) (Figure 1.5). Nuclear DNA-based studies were initially undertaken, to examine human prehistory (Cavalli-Sforza et al., 1994), although recently studies utilising mtDNA and Y chromosomes have become more popular (Hammer and Bonner, 1994; Richards et al., 1996; Bianchi et al., 1997; Hurles et al., 1998; Hurles et al., 1999).

### 1.11. Scope of this thesis

Three distinct but complementary studies examining Y-chromosomal diversity have been undertaken and will be described within this thesis. Initially, a large continental-wide study examining the distribution of biallelic haplogroup diversity within 48 European and circum-European populations will be discussed (Chapter 3). Secondly, a more detailed micro-geographic study examining the haplogroup distribution and diversity within a smaller subpopulation, a linguistic isolate from the Italian Alps will be presented (Chapter 4). Finally, a global study of a single lineage, defined by the biallelic marker YAP (Y Alu polymorphism) will be described employing all three types of Ychromosomal markers (biallelic, microsatellites and the minisatellite, MSY1) to try and explain the unusual global distribution of this lineage (Chapter 5).


Figure 1.5. Species tree drawn from mitochondrial DNA sequence comparisons
Indicating the close genetic relationship between human populations, with regards to mitochondrial DNA. The non-human primates are much more diverse in this respect, shown by the length and number of branches linking the different primate species (adapted from Gagneux et al., 1999).

## Chapter 2: Materials and Methods

### 2.1. Materials

### 2.1.1. Chemical and molecular biology reagents

Materials were purchased from the following companies: Advanced Biotechnologies (Epsom, UK), Amersham pharmacia biotech Ltd. (Buckinghamshire, UK), Boehringer Mannheim (Lewes, UK), Clontech (Basingstoke, UK), Fischer Scientific Limited (Loughborough, UK), Fisons Scientific equipment (Loughborough, UK), FMC Bioproducts (Rockland, USA), (Gibco BRL) Life Technologies Ltd (Renfrewshire, Scotland), HT Biotechnology Ltd. (Cambridge, UK), National Diagnostics, (Hull, UK), NEN Dupont, (Belgium), New England Biolabs (Hertfordshire, UK), Perkin-Elmer, and PE Biosystems (Beaconsfield, UK), Pharmacia (Milton Keynes, UK), Sigma-Aldrich Chemical company (St. Louis, USA), Stratagene (La Jolla, USA), Qiagen Ltd. (Crawley, UK).

### 2.1.2. Oligonucleotides

These were synthesised by PNACL (Protein and Nucleic Acid Laboratory, University of Leicester, UK). See the relevant chapters for the lists of oligonucleotide sequences used during this project.

### 2.1.2.1. Fluorescently labelled oligonucleotides

For the fluorescently labelled oligonucloetides used during Y chromosomal specific microsatellite analysis, see Chapter 5 (materials and methods).

### 2.1.3. Genomic DNA samples

A variety of collaborators provided some of the DNA samples analysed during all three studies, and for comprehensive lists, see Chapters 3, 4 and 5.

### 2.1.4. Commonly used solutions

 TE> 10mM Tris-HCl pH7.5, 1mM EDTA (pH8.0)

10X TBE buffer

10X TBE (sequencing)

5X Bromophenol blue loading buffer

5X Xylene-cyanol loading buffer

Jeffreys' PCR buffer (11.1X) (Jeffreys et al., 1990)

Advanced Biotechnologies
PCR buffer
(Tbr 1x) PCR buffer

MVR-PCR stop solution

1M Tris-borate, 2 mM Na 2 EDTA (pH 8.3)

5X TBE, $0.25 \%(w / v)$ bromophenol blue, $0.25 \%$ (w/v) xylene cyanol, $15 \%$ (w/v) Ficoll (type 400, Pharmacia) in $\mathrm{H}_{2} \mathrm{O}$
0.25\% (w/v) xylene cyanol (Sigma), 15\% (w/v) Ficoll (type 400, Pharmacia) in $\mathrm{H}_{2} \mathrm{O}$

50 mM Tris- $\mathrm{HCl}(\mathrm{pH} 8.8), 12 \mathrm{mM}\left(\mathrm{NH}_{4}\right)_{2} \mathrm{SO}_{4}$,
$5 \mathrm{mM} \mathrm{MgCl} 2^{\prime}, 7.4 \mathrm{mM} \beta$-mercaptoethanol, $5 \mu \mathrm{M}$ EDTA, 1 mM each of dATP, dCTP, dGTP, dTTP (Pharmacia), $126 \mu \mathrm{~g} / \mathrm{ml}$ BSA (DNAse-free, Pharmacia).

1 X reaction buffer IV [200mM $\left(\mathrm{NH}_{4}\right)_{2} \mathrm{SO}_{4}$,
750mM Tris- PCR buffer HCl ( pH 9.0 ), $0.1 \%$ ( $\mathrm{v} / \mathrm{v}$ ) Tween-20] (minus $\mathrm{MgCl}_{2}$ ), 0.2 mM each of dATP, dCTP, dGTP, dTTP (Pharmacia), $1.5 \mathrm{mM} \mathrm{MgCl}_{2}, 200 \mu \mathrm{~g} / \mathrm{ml}$ BSA (DNAse-free, Pharmacia).

10 mM Tris- HCl (pH8.8), $50 \mathrm{mM} \mathrm{KCl}, 1.5 \mathrm{mM}$ $\mathbf{M g C l}_{2}, \mathbf{0 . 1 \%}$ (v/v) Triton X-100 (Sigma).
$98 \%$ (v/v) formamide, $20 \mathrm{mM} \mathrm{Na}_{2}$ EDTA, $0.05 \%$ (w/v) bromophenol blue, $0.05 \%(w / v)$ xylene cyanol (Sigma)

### 2.2. Methods

### 2.2.1. Extraction of $D N A$ from blood

Two different methods were employed to extract DNA from blood samples; initially the 'silica method' was used, but this changed to the Qiagen Ltd. blood kit due to the speed and ease of method.

### 2.2.1.1. Guanidinium isothiocyanate (GuSCN) purification method using a silica suspension

This method is from Boom et al. (1990), and is listed in detail.

## Diatom/silica suspension

A 100 ml glass measuring cylinder was treated with 10 M HCl for 30 minutes, and rinsed with sterile distilled Elga water. 12 g silica particles (Sigma) were weighed out into the cylinder, and topped up with Elga water. They were completely mixed by inversion, and allowed to sediment for 24 hours at room temperature 86 ml of water were removed, and the water was topped up to 100 ml again, mixed, and allowed to sediment for 5 hours, and then 88 ml were removed. $120 \mu \mathrm{l} 10 \mathrm{M} \mathrm{HCl}$ was added, the silica completely resuspended, divided into $500 \mu \mathrm{l}$ aliquots, and stored in the dark.

## Buffer preparation

L6 $\quad 24 \mathrm{~g}$ GuSCN (Sigma) were weighed out in a 50 ml sterile Falcon tube in the fume hood. 20 ml 0.1 M Tris- HCl ( pH 7.4 ) were added, and the GuSCN dissolved in a $60^{\circ} \mathrm{C}$ water bath. 1.8 ml 0.5 M EDTA ( pH 8 ) and 0.5 ml Triton X100 were added, mixed by inverting, and stored wrapped in aluminium foil.

L2 As L6 minus EDTA and Triton X, and stored wrapped in aluminium foil.

New Wash solution $50 \%(\mathrm{v} / \mathrm{v})$ ethanol, $0.1 \mathrm{M} \mathrm{NaCl}, 10 \mathrm{mM}$ Tris ( pH 7.5 ), and stored at $-20^{\circ} \mathrm{C}$.

## Extraction procedure

A volume of $100 \mu \mathrm{l}$ of blood was extracted using 6 to 7 volumes ( $700 \mu \mathrm{l}$ ) of buffer L6, in a 1.5 ml microcentrifuge tube. This was incubated for 15 minutes at $50^{\circ} \mathrm{C}$. The tubes were allowed to cool at room temperature. A silica solution was vortexed to ensure full suspension, and a $20 \mu \mathrm{l}$ aliquot was added to the blood mixture. This was placed on a slow vertical rotor for at least 30 minutes, at a speed to ensure the silica remained in suspension. The tubes were centrifuged at 15000rpm for 2 minutes, and the supernatant was removed and placed into a 3 M NaOH solution (to prevent cyanide fumes being produced). Five volumes ( $500 \mu \mathrm{l}$ ) of L2 wash buffer were added, and the mixture vortexed. The tubes were centrifuged at 15000 rpm for 2 minutes, and the supernatant was removed and placed into a 3 M NaOH solution.

Five volumes ( $500 \mu \mathrm{l}$ ) of ice cold New Wash solution were added, the tubes centrifuged and the supernatant discarded into a beaker. This step was repeated an additional two times, and during the final wash step, all the residual New Wash was removed. The diatom pellets were allowed to air dry. $50 \mu \mathrm{l}$ of 1 X TE ( pH 8.0 ) were added to each tube, and then incubated at $50^{\circ} \mathrm{C}$ for between $15-20$ minutes, with occasional agitation. The tubes were centrifuged at 15000 rpm for 5 minutes, and the eluant was removed, labelled ' 1 st eluant' ( $\sim 50 \mathrm{ng} / \mu \mathrm{L}$ DNA) and stored at $-20^{\circ} \mathrm{C}$. This step was repeated, to generate a 'second eluant' ( $\sim 5 n g / \mu \mathrm{l}$ DNA). Before use in PCR reactions, the tubes were centrifuged to prevent residual silica inhibiting amplification.

### 2.2.1.2. QIA amp blood mini kit

Buffers and solutions were all provided in the blood extraction kit supplied by Qiagen Ltd.

## Extraction procedure for liquid blood samples

A volume of $200 \mu \mathrm{l}$ of blood was extracted, by added $25 \mu \mathrm{l}$ of Qiagen protease/proteinase K stock solution, and $200 \mu \mathrm{l}$ of buffer AL. This was vortexed for 15 seconds, to dissolve the pellet. Incubated at $70^{\circ} \mathrm{C}$ for 10 minutes, and then $210 \mu \mathrm{l}$ of ethanol ( $100 \%$ ) were added, and vortexed to mix. A QIA amp spin column was placed in a 2 ml collection tube, and the
mix was added to the spin column (without touching the rim), and centrifuged at 8000 rpm for 1 minute. The collection tube and filtrate were discarded into a beaker containing approximately $50 \%$ (v/v) Chloros. The spin column was placed into a new 2 ml collection tube. A $500 \mu \mathrm{l}$ volume of buffer AW were added, and centrifuged at 8000 rpm for 1 minute. The collection tube and filtrate were discarded, and replaced by another collection tube. Another $500 \mu \mathrm{l}$ of buffer AW was added, and centrifuged at 13000 rpm for 3 minutes. The spin column was placed into a clean 1.5 ml microcentrifuge tube, and the DNA was eluted using $200 \mu \mathrm{l}$ of buffer AE (heated to $70^{\circ} \mathrm{C}$ ). The mixture was incubated at room temperature for 1 minute, and then centrifuged at 8000 rpm for 1 minute.

## Extraction procedure for dried blood samples

The dried material was moistened with a drop of PBS (phosphate buffered saline), and then $180 \mu \mathrm{l}$ PBS were added to a 1.5 ml microcentrifuge tube, and the blood was scraped into the tube. Agitation aided the removal of the blood from the material, but in some cases a small spin column had to be used to allow the liquid to be removed from the material. The rest of the procedure is as described above.

### 2.2.2. Precipitating oligonucleotide primer stocks

Primers synthesised by PNACL (University of Leicester) were supplied as a crude stock in ammonium hydroxide solution. The aliquoting of primer stocks was carried out inside the laminar flow hood to try and reduce contamination. The oligonucleotides were ethanol-precipitated ( $1 / 10$ volume of 3 M sodium acetate ( pH 5.2 ) and $100 \%$ ethanol) and the dried pellet was resuspended in sterile distilled Elga water. A $2 \mu \mathrm{l}$ aliquot was taken and the primer concentration was calculated from optical density (OD) readings taken (LKB Biochrom Ultraspec plus UV spectrophotometer) at a wavelength of 260 nm using quartz cuvettes. The concentration of the primer stock in $\mu \mathrm{M}$ was calculated using the formula below:

### 2.2.2.1. Re-hydrating lyophilised oligonucleotide primer stocks

Lyophilised primers were routinely ordered after the service became available at PNACL (University of Leicester), due to a problem of poor yield we initially had when precipitating primers. The yield and reproducibility increased dramatically. The lyophilised pellet was resuspended in a volume of $125 \mu$ l sterile distilled water, and the concentration was calculated as described above.

### 2.2.3. DNA amplification

DNA amplification was achieved using the Polymerase Chain Reaction (PCR [Saiki et al., 1985; Mullis and Faloona, 1987]). Assays employed a final concentration of $1 \mu \mathrm{M}$ for each primer for most of the assays undertaken, $0.1 \mathrm{U} / \mu \mathrm{l}$ of Taq polymerase (Advanced Biotechnologies), between 5-10ng of DNA, either the 1X PCR buffer system previously described (Jeffreys et al., 1990) or 1X AB PCR buffer (Advanced Biotechnologies). PCR reactions were cycled using an MJ Research Peltier Thermal PTC-200 cycler employing a number of different cycling conditions (see Chapter 3 for the list of different conditions relating to each assay). To minimise contamination, PCR reactions were set up in a laminar flow hood using pre-PCR dedicated pipettes and consumables, and zero DNA controls were also included with each set of PCR reactions. Each PCR reaction was set up using an Advanced Biotechnologies' skirted Thermo-Fast 96-well microtitre plate and PCR Micro-Mat, using a pre-PCR dedicated 12-channel pipette.

### 2.2.4. Primer annealing temperature

Primer annealing temperature was calculated using the approximation, $\mathrm{A} / \mathrm{T}=2^{\circ} \mathrm{C}$ and $\mathrm{C} / \mathrm{G}=4^{\circ} \mathrm{C}$. This allows an estimation of the annealing temperature but if initial amplifications appear to generate unspecific amplicons, the annealing temperature was varied to determine which gave the best amplification.

### 2.2.5. Restriction enzyme digestion

A $2 X$ reaction mix was set up including one unit of the appropriate restriction enzyme (and BSA, required in certain assays) which was then directly aliquoted into the 96 -well microtitre plates using a post-PCR dedicated 12 channel pipette. Digests were carried out for 1 to 2 hours at either $37^{\circ} \mathrm{C}$ or $65^{\circ} \mathrm{C}$ depending on the particular restriction enzyme being used.

### 2.2.6. Agarose gel electrophoresis

Agarose gel electrophoresis was carried out to resolve PCR-RFLP products and to isolate template DNA for MVR-PCR and automated sequencing reactions. This technique was carried out using 96 -well horizontal submarine agarose gels containing ethidium bromide ( $0.5 \mu \mathrm{~g} / \mathrm{ml}$ ) (Sigma) in 1XTBE buffer using electrophoresis tanks manufactured at the University of Leicester and Bio-RAD electrophoresis power supplies. DNA was visualised on UV transilluminator, and photographs were taken with a video system (Ultra Violet products, Inc./Mitsubishi) or saved to disk. The agarose gel (Seakem, FMC Biologicals) concentrations varied ( $1-3 \% \mathrm{w} / \mathrm{v}$ ) depending on the size of the DNA fragments being resolved, and were run at $7.5 \mathrm{~V} / \mathrm{cm}$.

For the majority of gels a size marker of $\phi$ X174 phage DNA digested with HaeIII (Gibco) was used. This gives the following fragment sizes (in bp): 1353, 1078, 872, 603, 310, 281, 271, 234, 194, 118, 72. Occasionally, a higher molecular weight size marker ladder was required, and $\lambda$ phage DNA digested with HindIII was used, which gave the following fragment sizes (in bp): 23130, 9416, 6557, 4361, 2322, 2027, 564, 125.

### 2.2.7. Biallelic analysis

For the detailed information on all the biallelic assays used during the three studies, see Chapter 3 (materials and methods).

### 2.2.8. Allele-specific oligonucleotide analysis

For the detailed explanation of all materials and protocols used during this analysis, see Chapter 5 (materials and methods).

### 2.2.9. $Y$-chromosome specific microsatellite analysis

For a detailed description of experimental protocols including primer sequences, and amplification conditions see Chapter 5 (materials and methods).

### 2.2.10. Minisatellite variant repeat mapping via $P C R$

The detailed explanation of the complete MVR-PCR analysis undertaken during this project is described in Chapter 4 (materials and methods).

# Chapter 3: A low resolution, continental study, examining Y chromosome diversity within European and circumEuropean populations 

### 3.1. Introduction

In this chapter, Y -chromosomal diversity within a number of European and circum-European populations will be examined. The influence upon this diversity of pre-historic events such as the advent and spread of agriculture and the establishment of currently spoken languages will be assessed.

### 3.1.1. The development and dispersal of agriculture across Europe

The earliest accepted date for the occupation of Europe by anatomically modern humans is around 40,000 years before present (YBP) (Boyd and Silk, 1997). Population size during the Palaeolithic was probably stable and small, limited by the resources available from a hunting-gathering economy (Landers, 1992). The development of agriculture (the Neolithic transition) from the fertile crescent in the Near East (Piazza et al., 1995) began approximately 10,000 years ago, and was important because the abundance of food supplies allowed populations to expand (Hassan, 1973). Previously hunter-gathering was the main way food was obtained; it is known that food sharing was more common among humans than other primates, and it is thought that individuals adjusted their foraging behaviour due to their age, sex and reproductive status (Hawkes et al., 1997).

Agricultural industry spread quickly across the whole continent, in all directions from Turkey to Macedonia to Greece northwards to the Balkans, westwards to Italy, southern France, and eastern Spain (Cavalli-Sforza et al., 1994). Cereal cultivation and domesticated animal husbandry were adopted (Champion et al., 1987; Cunliffe, 1998), and secondary by-products such as milk, wool and bone tools were also developed. This dispersal of agriculture from the Middle East (Piazza et al., 1995) can be followed using archaeological artefacts (Jones, 1991) (Figure 3.0) such as the spread of pottery (Cavalli-Sforza et al., 1994) and remnants of cultivated cereals and the skeletons of domesticated animals (Renfrew, 1987). Primary agricultural settlements were chosen due to soil fertility and location near to water, and a range of living accommodation was


Figure 3.0. Dates for the advent and spread of agriculture, from archaeological evidence.
The dating of archaeological artefacts such as pottery across Europe suggests a dispersal along a south-east to north-west axis (taken from Champion et al., 1987). This supports the suggestion that agriculture arose close to the fertile crescent in the Near East, approximately 10,000 years ago, during the Neolithic.
known to have existed such as homesteads ( 10 people), hamlets ( 10 to 50 people) and villages ( 50 to more than 250 people). Farming intensification led to diversification and specialisation, for example in cereal crops and animals, and the production of fruit in Southern Europe (Champion et al., 1987).

The origins of agriculture have become the focus of attempts to interpret the genetic landscape of modern Europe. It is generally accepted that agriculture began in the Near East around 10,000 years ago; but the argument has arisen over the mechanism of its subsequent dispersal. Two opposing hypotheses have been developed to try and explain how this expansion occurred. The first, demic diffusion (Ammerman and Cavalli-Sforza, 1984) proposes a slow expansion of people into Europe. During this process displacement or absorption of the existing hunter-gatherer populations would have occurred (Richards et al., 1996), and this would therefore have substantially changed the genetic composition of European populations. The second hypothesis is cultural diffusion, which suggests that ideas and technologies were transferred without substantial population movement (Dennell, 1983; Zvelebil and Zvelebil, 1988) and thus that current patterns of genetic diversity should have their roots in the Palaeolithic. A few intermediate hypotheses (Zvelebil, 1986; Willis and Bennett, 1994; van Andel and Runnels, 1995) have also been proposed but are not as frequently discussed as demic and cultural diffusion. Unfortunately, the presence of archaeological artefacts does not suggest which method of agricultural expansion (demic or cultural diffusion) was more likely to have occurred.

These opposing hypotheses are undoubtedly over-simplistic, but have been widely adopted as models in genetic studies (Sokal et al., 1991; CavalliSforza et al., 1993; Barbujani et al., 1994; Piazza et al., 1995; Semino et al., 1996; Chikhi et al., 1998a; Chikhi et al., 1998b; Richards and Sykes, 1998; Simoni et al., 2000a), because they predict patterns of diversity which should be easily recognisable. In particular, demic diffusion is expected to result in 'clines' (gradients) with foci in the Near East. Principal components analysis of classical gene frequency data reveals clines within Europe, and the first principal component, which indeed has a Near Eastern focus, has been taken to support the demic diffusion hypothesis (Menozzi et al., 1978; Cavalli-Sforza et al., 1993). A similar pattern has been observed in spatial autocorrelation analysis of DNAbased polymorphisms including microsatellites, which identified geographical patterns compatible with a substantial directional demographic expansion
affecting much of the continent (Chikhi et al., 1998a). However, while these patterns in the genetic data are impressive, and suggest major east-west population movements, their time-depths are not known, and associating them with particular demographic events is usually speculative. They could equally well be due to the original peopling of Europe in the Upper Palaeolithic, as to the Neolithic transition. In this regard, some support for the latter does come from the finding of significant partial correlations between classical marker frequencies and the relative dates for the origin of agriculture in different locations (Sokal et al., 1991).

### 3.1.2. Physical geography and palaeoclimatology

Genetic diversity can be influenced by factors such as geography and in particular geographical barriers, such as large expanses of water and mountain ranges. This following section discusses the modern geographical landscape of Europe and the pre-historic landscape, which would have influenced the genetic diversity observed in modern European populations.

The varied geography of the modern European landscape has been developed over millions of years through important climatic influences such as weathering, erosion, and the formation and retraction of expansive ice sheets. European climatic history is complex, and involved reductions in sea levels, and intense cold periods, interspersed by warmer, drier periods (Adams, 2000). Europe covers approximately 10 million square kilometres (Cavalli-Sforza et al., 1994); and is the second smallest continent in the world after Australia, with a current approximate population of 678.2 million (Castello-Cortes, 1998). Europe has numerous mountain ranges mainly located in the Southern part of the continent running from east to west, with only one exception in Scandinavia (Champion et al., 1987). Additional physical barriers include the Norwegian, Barents, North and Baltic seas; the Mediterranean Sea which separates Southern Europe from North Africa; the Atlantic Ocean which borders the West coast of Ireland; and the Caspian and Black Seas that divide the landmass in the East. Europe is located in the world's northern temperate zone (Champion et al., 1987) and a variety of climates and landscapes exist ranging from the well-drained, fertile regions of the North European Plain (that spans from Germany through to western Russia), to the warm, dry hilly regions of the Mediterranean, and to the heavily forested areas of Russia and the Ukraine (Castello-Cortes, 1998).

Advances in prehistoric technology have been identified due to the tool artefacts (known as industries) that have been unearthed; for example, heavy duty/multipurpose tools from around 35,000 years before present (YBP) which were replaced by smaller more intricate tools between 35,000 to 10,000 YBP. Geographical patterns have been observed for a number of prehistoric industries, which suggest that the environment may have affected daily life. A major change in the European climate occurred during the last glacial maximum, which spanned from around 22,000 years up to approximately 13,000 years ago (Adams, 2000). This may have caused some central Europeans to abandon their homes due to the extreme cold conditions, because the ice sheets extended across most of Europe, down to the region around central France (Adams, 2000). The beginning of the post-glacial period and the end of the Palaeolithic within Europe is considered to be around 10,000 YBP (Champion et al., 1987), although the vegetation would have taken hundreds if not thousands of years to regenerate, and the massive ice sheets, particularly in the far north-west would probably have taken thousands of years to melt due to their size (Adams, 2000).

### 3.1.2.1. The definition of populations used within this study

Populations included within this European study span from Iceland in the north to Northern Africa in the south, and from Ireland in the west through to Turkey in the east, plus many populations in between. In this study, populations are defined by their belonging to a particular country. However, modern nations are in many cases recent inventions, and can contain much diversity. Self-defined ethnic groups, which include cultural elements such as language and religion, may be a better way to sub-divide samples. This information was not available for some of the samples, and a general definition by place of birth of paternal grandfather was used.

### 3.1.3. European linguistic comparison

Languages are another important factor that may influence the genetic composition of modern European populations, and can be used to provide additional evidence about past demography (Renfrew, 1989), although direct information about past languages from writing is limited to the last 5,000 years, and inferences beyond that time are controversial (Renfrew, 2000). A recent classification listed 5000 languages world wide (Ruhlen, 1991). Languages can allow the relationships of modern populations to be studied even though they
evolve quicker than genes, and can be completely replaced (Cavalli-Sforza et al., 1994). Europe is remarkable for its linguistic homogeneity, with the IndoEuropean (IE) language family being extremely widespread, and spoken by most populations from India to Ireland (Renfrew, 1989), although a few examples of other language families including Uralic and Altaic are also spoken within Europe (Ruhlen, 1991). Figure 3.1 indicates the Indo-European subfamilies spoken by populations within this study. The distribution of language families and Indo-European language sub-families spoken within Europe are indicated in Figures 3.2 and 3.3.

The Indo-European family contains 140 of the world's languages, and is spoken by two billion speakers constituting nearly half of the world's population (Ruhlen, 1991); the spread to other continents of the IE languages has been documented since 1500 (Ruhlen, 1991). The Uralic language family includes Finnish and Hungarian, and today is spoken mainly in northern Europe, but is also spoken to the east of the Ural mountain range in north-western Asia. The Caucasian languages (e.g. Georgian) are spoken around the Caucasus mountain range in south east Europe and Asia (Ruhlen, 1991).

The way in which languages are inherited is not always straightforward, and just because two different populations speak the same language this does not prove they share a common ancestry. For example, the Finns and Saami speak closely related Uralic languages (Zerjal et al., 1997), but recent analysis of autosomal and mitochondrial DNA (mtDNA) suggests they do not share a common ancestry despite this language similarity, which led to the proposal that the Finns had undergone replacement of an originally IE language (Sajantila and Pääbo, 1995). Some examples of non-Indo-European languages do not reflect persistence, but recent acquisition via a number of different processes. Two models have been proposed to try to explain language replacement, the first is a 'wave of advance' model (Ammerman and Cavalli-Sforza, 1984), which is when a language enters the landscape during a major demographic event such as the spread of agriculture. The second, is the 'elite-dominance model', where a small group of people invade a country and impose their language. An example of this is the Hungarian language, because it is recorded that the Magyars conquered Hungary around 1,100 YBP and imposed the Uralic language spoken today (Ruhlen, 1991; Cavalli-Sforza et al., 1994). Similarly the Altaic language of the


Figure 3.1. The subfamilies of Indo-European
Indicates all the sub-families contained within the Indo-European language family spoken by populations analysed during this European Y-chromosomal diversity study.


Figure 3.2. Language families spoken across Europe
The distribution of different language families spoken across continental Europe; each language family is represented by a different colour (see key). One linguistic isolate (Basque) is also indicated.



Uralic family


Altaic family


Afro-Asiatic family


Basque $\bigcirc$ Basque

Figure 3.3. Language subfamilies spoken across Europe
The different language families have been sub-divided into language sub-families spoken by populations being studied, to show their distribution across Europe.

Turks was acquired as a result of the Turkic invasions in the $11^{\text {th }}$ to $15^{\text {th }}$ centuries (Renfrew, 1989).

The Celtic language and people were once widespread in central and northern Europe (Renfrew, 1987), and archaeological evidence suggests that Celtic culture expanded until the Romans conquered parts of Europe in the first century B.C. Latin replaced Celtic and other languages (under Roman rule), with the exception of the extreme regions where the Romans did not reach (CavalliSforza et al., 1994). Language acquisition by elite dominance would not be expected to result in a high degree of genetic admixture, and if this is so, populations such as the Hungarians and Turks are unlikely to be separated from surrounding populations by genetic barriers. However, various methods for the detection of genetic barriers in autosomal gene frequencies within Europe (Barbujani, 1991) show that most of these barriers correlate with linguistic boundaries, and it may be that language and geographical proximity are equally good predictors of genetic affinity (Barbujani, 1997).

It has been proposed that Neolithic migrants from the Near East, who may have introduced agriculture to Europe spoke proto-Indo-European languages (Renfrew, 1987; Dolgopolvsky, 1988). If the farmers were greater in number than the hunter-gatherers it would be more likely they retained their language. Other ideas have been proposed, however: one which has been adopted by some geneticists, because of its apparent compatibility with the pattern seen in the third principal component of variation of classical gene frequencies (CavalliSforza et al., 1994), is that the Indo-European language was spread by the movement, from north of the Caspian Sea, of the Kurgan people, pastoral nomads who domesticated the horse (Gimbutas, 1970). An alternative view has it that the spread of Indo-European preceded the origins of agriculture, and was due to the re-expansion of hunter-gatherers after the end of the last glacial maximum (Adams and Otte, 2000). Despite the homogeneity of Indo-European, there is diversity within it, and some members of other language families also exist; for example, Basque, a linguistic isolate, is a good example of how people can retain their language from an earlier era (Renfrew, 1987).

### 3.1.4. Previous European genetic studies

Major demographic events such as population migrations, bottlenecks and expansions should leave detectable imprints on human genetic diversity (Jorde et
al., 1998). Even though subsequent migrations may have progressed across Europe, it is unlikely that they will have completely erased existing patterns that may have been generated during the Neolithic expansion of agriculture (Rendine et al., 1986). On the one hand, European populations are classified by country in which they live but are often heterogeneous; on the other, many countries have been unified politically or linguistically for enough time to remove any original differentiation. Central parts of Europe appear to be genetically homogeneous, and it is proposed this may be a direct result of a Neolithic diffusion, reducing the initial differences, leading to a smooth gradient of genetic differentiation (Cavalli-Sforza et al., 1994).

Classical polymorphic markers, such as protein polymorphisms, and mtDNA evidence suggest an overall lack of genetic diversity within European populations, when compared to the rest of the world (Cavalli-Sforza et al., 1993; Piazza, 1993; Cavalli-Sforza et al., 1994; Comas et al., 1997), with only a few exceptions including the Saami, Sardinians, Icelandics and Basques (CavalliSforza et al., 1993; Piazza, 1993; Cavalli-Sforza et al., 1994). A genetic tree generated from 'classical' data converted to genetic distance (Fst) values (Figure 3.4), shows that European populations form geographical clusters. The smaller the genetic distance between populations, the closer the relationship (CavalliSforza et al., 1994). The analysis of European mtDNA diversity reveals a relatively homogeneous landscape (Comas et al., 1997), with clines detectable only in the south (Simoni et al., 2000a). However, this is a contentious area, and conclusions may depend on the depth of analysis - for example, which sublineages are studied. An east-west gradient of pairwise differences has been discerned, and claimed to be compatible with expansion from the Middle East (Comas et al., 1997). However, attempts to identify and date founding lineages (Richards et al., 1996) have suggested that Palaeolithic lineages may persist in Europe to a degree which is inconsistent with the demic diffusion hypothesis, although an ancient origin of certain alleles or haplogroups is certainly compatible with a later spread of those alleles in a geographical region (Langaney et al., 1992; Templeton, 1993).

The combination of data from mtDNA, the Y-chromosome, and autosomal loci suggests a Palaeolithic population lived in the south-west region of Europe that significantly contributed to the gene pool of western and northern Europeans (Torroni et al., 1998). An almost complete European origin for the


Figure 3.4. A tree generated from genetic distance data encompassing 26 European populations The generation of trees allows the relationships between populations to be examined. This tree incorporates data on genetic distances from 88 genes within 26 European populations. Data taken from work by Cavalli-Sforza et al., 1994.
maternal component (mtDNA) of the Saami and Finns has been proposed (Lahermo et al., 1996), whereas in contrast, Y-chromosomal data suggest that Uralic speaking populations, including the Saami and Finns, share a large portion of their gene pool with western and central Asian populations (Zerjal et al., 1997). These data thus suggest different migration rates and mating practices for men and women.

A number of studies have detected a correlation between genetic and linguistic barriers across Europe (Sokal et al., 1988; Barbujani and Sokal, 1990; Calafell and Bertranpetit, 1994; Piazza et al., 1995; Poloni et al., 1997), but in some cases (Sokal et al., 1988; Barbujani and Sokal, 1990) these divisions also corresponded to physical barriers which may also induce the genetic and linguistic barriers (Piazza et al., 1995). In one study it was proposed that language rather than geography was the main predictor when analysing Y chromosomal genetic diversity, and in particular approximately $25 \%$ of the total diversity was due to differences between language families (Poloni et al., 1997).

### 3.1.5. Previous $Y$-chromosomal studies

Published data on European $Y$ chromosome diversity are not extensive; markers have been of limited informativeness, and the distribution of population samples has often been unsatisfactory. Previous work suggested that the $Y$ chromosome haplogroup frequencies vary across geographical regions (Hammer, 1994; Seielstad et al., 1994; Spurdle et al., 1994; Hammer et al., 1997; Zerjal et al., 1997). The YAP (Y Alu insertion polymorphism) positive chromosomes were detected at the highest frequencies within the African continent but also at lower levels in Europe, western and eastern Asia (Hammer et al., 1997). Another polymorphism, known as Tat (a T to C transition), defines haplogroup 16 (hg 16). The highest frequencies of the C allele are detected in Asia and northern Europe, and in particular amongst Altaic and Uralic speakers. Analysis of this polymorphism has suggested a large Asian paternal genetic contribution to northern European populations (Zerjal et al., 1997), and these Y-chromosomal data are in contrast to the mtDNA component of these same populations because that appears to be mainly European (Lahermo et al., 1996).

Using two 'classical' Y-chromosomal markers, the complex and highly polymorphic p49a,f/TaqI system (Ngo et al., 1986; Lucotte and Loirat, 1999) and the biallelic 12f2, patterns of diversity were demonstrated which were claimed to
be clinal, and to support the demic diffusion model (Semino et al., 1996). Subsequent analysis using Y-specific microsatellites (Quintana-Murci et al., 1999), and using a combination of microsatellites and two biallelic markers (Malaspina et al., 1998), showed similar east-west gradients. The 49f locus has been exploited more fully to analyse the correlation between $Y$ diversity, mtDNA diversity, and language in a global sample, and suggested that the Y shows a stronger correlation with language (Poloni et al., 1997).

One useful property of the Y chromosome is its high degree of geographic differentiation compared to other parts of the genome (Seielstad et al., 1994; Ruiz Linares et al., 1996; Underhill et al., 1996; Underhill et al., 1997), which has been explained by drift and a greater effective migration of females than males through the phenomenon of patrilocality (Seielstad et al., 1998), where females are more likely to move from their birthplace upon marriage than are males. In general it may have been easier for women to cross cultural boundaries (Poloni et al., 1997), as is demonstrated by the caste system in India (Bamshad et al., 1998). A recent study proposed that women have a higher migration rate than men (Seielstad et al., 1998) and even today many societies still practice patrilocality (Stoneking, 1998), although, some populations, such as in Northern Iberia, may have practised matrilocality (Collins, 1986). Other mating practices such as polygyny, which is when certain men father a disproportionately large number of children compared to other men will decrease the overall level of $Y$ chromosome variation because they reduce the effective population size (Seielstad et al., 1998). The Y may therefore be a sensitive system for detecting the population movements, which have shaped European genetic diversity; there again, it may be so susceptible to drift that ancient patterns have been obscured.

### 3.1.6. Questions to be addressed

The major aims of this project are to employ the Y chromosome to examine genetic diversity within 48 European and circum-European populations; to try and determine if important pre-historic events such as the Neolithic expansion of agriculture can influence $Y$-chromosomal genetic diversity; to examine the effect of geography and language on genetic diversity; and to determine if genetic barriers are detected, and if they correlate more with geography or language. The work described in this Chapter has been published (Rosser et al., 2000), and this paper is included in Appendix 4.

### 3.2. Materials and Methods

### 3.2.1. Genomic DNA samples

A total of 3677 males from 48 European and circum-European populations were analysed (Figure 3.5) and the majority were classified by birthplace of paternal grandfather. Ideally samples should be collected with detailed information regarding what ethnic group or nationality they consider themselves to be, their surname, and native language spoken and by their immediate ancestors, but this is not always possible. Many samples were obtained as genomic DNA, but some were prepared from blood using a silica extraction method, or a Qiagen kit, see Chapter 2 (materials and methods). DNA and blood samples were provided by a variety of collaborators, and obtained with informed consent. A total of 311 samples from the Baltic region are from the study of Zerjal et al (in preparation). The 257 Irish Y chromosomes included 221 previously studied chromosomes (Hill et al., 2000), which here were typed with three additional markers. The 129 North African samples were those previously studied by Bosch et al. (1999); chromosomes with the M9 G allele and 92R7 C allele were additionally typed with LLY22g (see below).

The actual geographic locations for all the sample sites were determined using a range of resources, such as maps, and the internet - a number of web sites are available that list the longitude and latitude values for a range of geographic locations, such as capital cities. In the majority of cases the actual location of each population was known, but in some cases, the samples were known to originate from a selection of places, and so the capital city of the relevant country was chosen as a geographic location.

### 3.2.1.1. DNA samples typed at Leicester

The following samples listed in alphabetical order were all analysed by myself at the University of Leicester: Bavarian (T. Meitinger, Munich, Germany), Belarusian (Y. Dubrova, Leicester, England), Bulgarian non-gypsies (L. Kalaydjieva, Perth, Australia), Chuvash (R. J. Mitchell, Victoria, Australia), Danish and Greenlandic Inuit (S. Nørby, Copenhagen, Denmark), Dutch (P. de Knijff, Leiden, Netherlands), East Anglian (G. Cooper, Cambridge, England), French (K. McElreavey, Paris, France), German (M. Kayser, Berlin, Germany), Greek Cypriot and Greek mainland (P. Patsalis, Nicosia, Cyprus), Polish (A.


| Number | Population | Number | Population |
| :---: | :---: | :---: | :---: |
| gld | Greenlandic | rom | Romanian |
| ice | Icelandic | yug | Yugoslavian |
| saa | Saami | sln | Slovenian |
| swe | N. Swedish | hun | Hungarian |
| got | Gotland | pol | Polish |
| nor | Norwegian | ita | ${ }_{\text {Italian }}$ |
| ${ }_{\text {fin }}$ | Danish | sar | ${ }_{\text {Sardinian }}$ |
| ${ }_{\text {est }}$ | Estonian | bav | ${ }_{\text {Gavarian }}$ |
| lat | Latvian | dut | Dutch |
| lit | Lithuanian | fra | French |
| rus | Russian | bgm | Belgian |
| brs | Belarussian | scw | Scottish (W. Isles) |
| ukr | Ukrainian | scm | Scottish |
| mar | Mari | enw | Cornish |
| chu | Chuvash | ene | East Anglian |
| geo | Georgian | irl | Irish |
| oss | Ossetian | bas | ${ }_{\text {Basque }}$ Spanish |
| arm | Armenian | spa | ${ }_{\text {Spanish }}$ |
| tur | Turkish | pos | S. Portuguese |
| gk | Greek | alg | Algerian |
| bul | Bulgarian | naf | N. African |
| cze | Czech |  |  |
| slk | Slovakian |  |  |

Figure 3.5. The names and actual locations of all populations examined This map and corresponding table shows the different European and circum-European populations being analysed within this Y-chromosomal study. Each population has been allocated a different number, that can be seen on the map indicating the actual position of each population.

Jeziorowska, Lødz, Poland), Sardinian, (T. Zerjal, Oxford, England), Scottish (Western Isle) (B. Sykes, Oxford, England), Slovenian (B. Peterlin, Ljubljana, Slovenia), Spanish (E. Arroyo, Madrid, Spain), Turkish (A. Tolun, Istanbul, Turkey).

### 3.2.1.2. DNA samples typed elsewhere

The following samples listed alphabetically in relation to the typer, were typed in a variety of European laboratories: Estonian, Finnish, Gotland, Hungarian, Icelandic, Latvian, Lithuanian, North Swedish, Norwegian, Mari, Russian, Saami (T. Zerjal, Oxford, England); Basque, Algerian (A. Pandya, Oxford, England); Italian (C. Previderé, Pavia, Italy); Cornish and Scottish Western Isles (J. Nicholson, Oxford, England); Armenian, Estonian, Georgian, Ossetian, Polish, Russia, Slovakian, Slovenian (R. Villems, Tartu, Estonia); Romanian (P. Malaspina, Rome, Italy); Danish, Irish, Turkish (E. Hill, Dublin, Ireland); Yugoslavian (D. Alavantic, Belgrade, Yugoslavia); Belgian and Ukrainian (R. Decorte, Leuven, Belgium); North African (E. Bosch, Barcelona, Spain); North Portuguese (H. Côrte-Real, Lisbon, Portugal); South Portuguese (M.-J. Prata, Oporto, Portugal).

### 3.2.2. Experimental procedure for biallelic typing

The following experimental procedures were based on those described in (Hurles et al., 1998). The cycling programs differ for the different assays, but the composition of the reaction mix remains virtually the same. PCR amplification (Saiki $e t$ al., 1985) was achieved by adding 5-20ng of genomic DNA to each PCR reaction mix (11.1X) (Jeffreys et al., 1990) or 1X AB PCR buffer (Advanced Biotechnologies), including $1 \mu \mathrm{M}$ of each primer and 0.5 U of Taq polymerase (Advanced Biotechnologies, Surrey, England). The PCR reactions were done in 96 -well Thermowell microtitre plates (Advanced Biotechnologies) in a MJR PTC200 PCR machine. PCR reactions were prepared in a $10 \mu \mathrm{l}$ reaction volume, and a zero DNA control was included per set of reactions.

All but one of the polymorphisms assayed can be typed using a PCR-RFLP approach. RFLP analysis was carried out post-PCR with the appropriate restriction enzyme and 1 volume of 2 X restriction mix being added directly to the 96 -well microtitre plate, and the samples were digested for more than 2 hours at the appropriate temperature. The digested samples were loaded using a postPCR 12-channel pipette and electrophoresed on agarose gels (Seakem agarose,

FMC Bioproducts) ranging from 1\% to 3\%, in 1X TBE (Tris-Borate) buffer, in electrophoresis tanks designed in the laboratory (University of Leicester). The above procedure was carried out at the University of Leicester, and slightly different methods may have been employed by the different laboratories during this collaboration, and so to verify typing methodologies, a set of 12 quality control DNA samples were satisfactorily typed 'blind' by all participating laboratories.

### 3.2.3. Biallelic markers

In this study a total eleven biallelic markers were employed to assay the diversity of Y-chromosomal lineages in a large sample of males distributed over most of Europe. These markers were chosen on the basis of previous work (Santos and Tyler-Smith, 1996; Semino et al., 1996; Underhill et al., 1997; Zerjal et al., 1997; Hammer et al., 1998; Hurles et al., 1999) indicating that the haplogroups (hg) they define are likely to be found within European populations (Figure 3.6). Clearly, this could bias the results and will be discussed later. Haplogroup 7 is specific to sub-Saharan African populations (Karafet et al., 1999), but is typed here by default, since it is defined by the ancestral state of the recurrent SRY-1532 polymorphism. Maximum parsimony analysis of haplotypes defined by these markers generates a unique tree, in which DYS257 (Hammer et al., 1998) and 92R7 (Mathias et al., 1994) are phylogenetically equivalent (Jobling et al., 1998b; and unpublished data). For this part of the phylogeny, 92R7 was typed routinely, and DYS257 to confirm results where necessary.

Nine of the markers have been described previously: The Y Alu polymorphism YAP (Hammer and Bonner, 1994) was typed according to Hammer and Horai (1995) (Figure 3.7), SRY-1532 (SRY 10,831 of Whitfield et al., 1995a) according to Kwok et al. (1996), SRY-2627 (originally known as SRY-2628) according to Veitia et al. (1997), 92R7 (Mathias et al., 1994) according to Hurles et al. (1999), DYS257 according to Hammer et al. (1998), M9 according to Underhill et al. (1997) (Figure 3.8) as described by Hurles et al. (1998), sY81 according to Seielstad et al. (1994), Tat as described by Zerjal et al. (1997), and SRY-8299 (Whitfield et al., 1995a) according to Santos et al. (1999). The LLY22g HindIII polymorphism was typed by an unpublished PCR-RFLP assay (E. Righetti and C.T.-S. unpublished).


Figure 3.6. A unique tree showing Y-chromosomal haplogroups
Maximum parsimony tree of all the Y-chromosomal haplogroups, and their geographic affiliations within global populations. The haplogroups assayed during this European study are shown as yellow circles. If the direction of the mutation is known, then this is indicated by an arrow, with the digested allele also being indicated.


## Plate organisation of the DNA samples

Wells A3 to C10 contain Bulgarian non-gypsy DNA
Wells C11 to D9 contain Bulgarian gypsy DNA
Wells D10 to G3 contain Turkish DNA
Well G4 is empty
Wells G5 to H12 contain Polish DNA

Figure 3.7. The YAP (Y Alu polymorphism) assay
Ethidium bromide-stained $2.5 \%$ agarose gel showing 96 PCR products from an assay of the YAP polymorphism on Bulgarian gypsy, non-gypsy, Turkish and Polish DNA samples. This assay allows the detection of an Alu insertion element (+455bp compared to -150bp) which defines the YAP branch of the tree. Negative and positive controls are included to try and reduce the possibility of mis-scoring samples. A weak band is visible in the positive control lane. The organisation of the samples within the plate is also indicated, with each row being allocated a letter and each individual well a separate number, e.g. A3, C12, H5.

The 12f2 (Casanova et al., 1985) marker was typed using a newly developed PCR assay (Blanco et al., 2000; Rosser et al., 2000). This polymorphism was originally suggested to be a $\sim 2 \mathrm{~kb}$ insertion/deletion, but recent analysis suggests its molecular basis is more complex than this. The PCR assay generates a 500bp product from chromosomes carrying the TaqI/10kb allele, but this product is absent from TaqI/8kb allele chromosomes (hg 9). An 820bp amplicon from the $S R Y$ region, present in all chromosomes, is amplified as a control. Analysis of the 12 f 2 region gives no information about ancestral state, but it is assumed that presence of the 500bp amplicon is ancestral. Primer sequences for the 12 f 2 amplicon are 12 f 2 D ( $5^{\circ}$-CTG ACT GAT CAA AAT GCT TAC AGA TC$3^{\prime}$ ) and 12 f 2 F ( $5^{\circ}-\mathrm{TCT}$ TCT AGA ATT TCT TCA CAG AAT TG-3'), and for the SRY control amplicon are $3^{\circ} S R Y 15$ ( $5^{\prime}$-CTT GAT TTT CTG CTA GAA CAA G-3') and $3^{\prime} S R Y 16$ ( $5^{\prime}$-TGT CGT TAC ATA AAT GGG CAC-3'). PCR conditions were $33-35$ cycles of $94^{\circ} \mathrm{C}$ for $30 \mathrm{~s}, 59^{\circ} \mathrm{C}$ for $30 \mathrm{~s}, 72^{\circ} \mathrm{C}$ for 45 s . An alternative assay, generating shorter amplicons, was used with degraded DNAs. The primers 12f2D (see above) and 12f2G ( $5^{\prime}-G G A$ TCC CTT CCT TAC ACC TTA TAC-3') produce an 88bp product from TaqI/10-kb allele chromosomes (and no product from TaqI/8kb allele chromosomes), which is coamplified with the Tat 112-bp amplicon (Zerjal et al., 1997) as a control, under the following conditions: 33-36 cycles of $94^{\circ} \mathrm{C}$ for 30 s, $59^{\circ} \mathrm{C}$ for $30 \mathrm{~s}, 72^{\circ} \mathrm{C}$ for 30 s . All chromosomes known from previous hybridisation analysis to carry TaqI/8kb alleles lacked the 12 f 2 test amplicons in both of these assays. However, some YAP+ chromosomes belonging to hg 4 also lack the 12 f 2 amplicons, suggesting that the polymorphism may be recurrent (Blanco et al., 2000) (see Chapter 5).

The deep-rooting markers $S R Y$-1532, M9, YAP and 92R7 were typed on all samples. For many samples, all other markers were also typed. However, in some cases remaining markers were typed hierarchically - for instance, SRY-8299 and sY81 were in some cases only typed on chromosomes classified as YAP+.

### 3.2.4. Statistical analysis

A number of statistical approaches can be utilised during the analysis of $Y$ chromosomal haplogroup frequency data, and each will be detailed in turn.

### 3.2.4.1. Principal components analysis

Principal components analysis (PCA) was proposed by Hotelling (1933) and allows large sets of data to be summarised with the minimum
amount of loss of information (Cavalli-Sforza et al., 1994), and was carried out here according to Harpending and Jenkins (1973). Principal components (PCs) are generated by calculating the standard correlation via pairwise analysis, between all pairs of characters such as gene frequencies. For example, data from 100 gene frequencies can be summarised by a single value, known as the first principal component (PC1). This value can incorporate upto $80 \%$ of the total variation, and a second PC can be calculated to incorporate the majority of remaining variance, and subsequent PCs can be calculated until nearly all of the variance can be accounted for. Generally, the first two PCs are plotted against each other, and this is a graphical way to visualise more than half of the genetic variance (Cavalli-Sforza et al., 1994).

Work carried out employing PCA, but superimposing the data onto a geographic landscape, leading to the generation of synthetic maps (Figure 3.9), changed how many people carried out their genetic analysis. Some problems exist with this type of analysis; for example, it is preferable to use measures of the relationship that are proportional to evolutionary time, such as genetic distances. Genetic distances are very sensitive to geographic distances between populations, and increase with them (Cavalli-Sforza et al., 1994). PC synthetic maps can supplement information such as phylogenetic trees, and because they incorporate more data than trees, it can be more informative to compare two PC maps than two trees. Data from PC1 have been taken to support the demic diffusion method of dispersal for agriculture from the Near East (Cavalli-Sforza et al., 1994). However, the direction of gradients observed in synthetic maps cannot be determined and no actual dates can be assigned to these patterns, and correlating them with significant human events such as migrations can be speculative, and must be done with caution.

### 3.2.4.2. Spatial autocorrelation analysis

Spatial autocorrelation analysis was carried out using AIDA (Bertorelle and Barbujani, 1995) for the entire data set, and SAAP (Sokal and Oden, 1978) for individual haplogroups. The SAAP analysis was carried out by Guido Barbujani (University of Ferrara, Italy), and used distance classes ranging from 400 km up to $4391 \mathrm{~km}(400,800,1200,1600$, $2000,2500,3000,3500,4000,4931 \mathrm{~km}$ ). The AIDA analysis was undertaken


Figure 3.9. Two synthetic maps generated from principal components analysis Synthetic maps generated from principal components analysis (PCA) on classical markers superimposed onto a European and western Asian landscape. Map (A) uses the values generated for principal component three (PC3), and map (B) uses values generated for PC4. Such maps have been taken to support a number of pre-historic migrations and expansions(Figures taken from Cavalli-Sforza et al., 1994).
using the Bertorelle software (http://www.unife.it/genetica/Giorgio /giorgio_soft.html); the output can also be visualised in a graphical form as correlograms and allows clinal variation, reflecting population movement or natural selection, to be distinguished from isolation by distance, reflecting short-range dispersal and drift, and from nonsignificance. One problem with SAAP analysis is that it cannot indicate the axis of a cline. These methods (like PCA) give a measure of genetic similarity between populations within particular geographical distance classes (Bertorelle and Barbujani, 1995).

### 3.2.4.3. Mantel analysis

Mantel tests (Mantel, 1967), using Arlequin version 2.0 (Schneider et al., 2000), were used to determine if language or geography has the stronger impact upon genetic differentiation. Three types of input data were required for this Mantel program: the first was the frequency haplogroup data used to calculate genetic distances, in the form of pairwise Fst comparison values, between all populations analysed. The second was actual geographical distances between all sample locations derived from actual longitudinal and latitudinal positions using great circle distances in a program written in Interactive Data Language 5.1 (Research Systems Inc.) by Matt Hurles (see Table 3.0). The third involves distances between language families and between language subfamilies (see Table 3.1). Figure 3.10 shows the proposed relationship between all language families spoken within this study.

Within Indo-European (IE), linguistic distances were adapted from Dyen et al. (1992), who used the lexicostatistical method of Swadesh, (1952), based on comparisons of 200 -word lists (including commonly used words in all languages, such as 'mother', 'father', 'head' and 'tree'. Percentage similarities were first converted to dissimilarities (e.g. 80\% similar equals $20 \%$ dissimilar), and these numbers then assigned as nonpercentage distances between languages (ranging from 9 [Czech to Slovak] to 88 [Armenian to Irish]). All IE languages within the data set were represented with the exception of Scottish, which was assigned a distance of 10 from Irish, due to their sharing a common Celtic language; the effects of other values in the range 5-20 were also tested. The Belgian sample was divided into its two linguistic groups, because information was




Figure 3.10. Relationship between all the language families spoken by populations during this project These language trees have been generated based on work by Ruhlen, 1991, and they indicate the relationship between sub-family members, within each family. Languages are not inherited like genes and because of this, a bifurcating tree may not be the best way to display them. The numbers superimposed onto some of the tree branches are the lexicosatistical distances, adapted from Dyen et al., 1992, or estimated from work by Ruhlen, 1991. For the complete list of lexicostatistical values between all language sub-families analysed see Table 3. For the abbreviations see Figure 3.5.
known about which language were spoken by the individuals sampled, the majority speaking French (56 individuals) and the remainder Dutch (36).

An arbitrary and conservative larger value of 200 was then assigned as a distance between language families. As was done by Poloni et al. (1997), Mantel tests were also performed using different inter-languagefamily distances, of 400 and 1000. Two of the non-IE language families, Altaic and Uralic, are represented by more than one language within the data set. To date no known lexicostatistical analyses for language families other than Indo-European have been undertaken (Paul Black, personal communication). Based on a consideration of the classification of Ruhlen (1991), and on the inter-IE distances of Dyen et al., (1992), plausible distances were assigned within these families, and the effect of altering these values over a range was tested. Within Uralic, values were as follows: Finnish to Estonian, distance of 25 (altered value range 10-30); Finnish-Estonian to Saami, 30 (20-40); Finnish-Estonian-Saami to Mari, 40 (30-70); and Hungarian to all other Uralic languages, 80 (40-90). Values for Chuvash and Turkish (Altaic) were, 40 (20-60).

### 3.2.4.4. Genetic Barrier Analysis

To locate zones of abrupt genetic change, or genetic boundaries within these data and to assess their significance, the Orinoco program written in Interactive Data Language 5.1 (Research Systems Inc.) by Matt Hurles (Hurles, 1999) was used. This process adapts a method proposed by Womble (Womble, 1951) known as 'wombling' (Barbujani et al., 1989), and initially developed for the analysis of allele frequencies, and identifies genetic boundaries as regions of rapid genetic change. The genefrequency values are used as the raw data, and boundaries that are detected between two regions can be visualised by a line. A sharp boundary is likely to indicate reduced gene flow in a defined region. The Algerian and North African samples were excluded from the barrier analysis, since their high degree of difference from all other samples (as shown in principal components analysis, for example) represents a strong genetic barrier which would bias the detection of barriers elsewhere.

During this method, an inverse-distance squared weighted algorithm was used to interpolate the frequencies for each of the eight observed haplogroups at each grid point within a 100 by 100 array, taking into account the curvature of the earth, and corresponding to a grid point every 0.36 degrees of latitude and 0.72 degrees of longitude. The derivatives (slopes determined from differentiation) of these eight interpolated surfaces were then calculated at every node of the grid, and the magnitudes of the derivatives summed. This identified the positions of steep gradients of genetic change, and measured the slope of the combined surfaces, i.e. the overall rate of $Y$-chromosomal genetic change in 10,000 rectangles covering Europe.

The significance of these gradients was considered in two ways, taking into account isolation by distance within the landscape (Barbujani et al., 1989). First, a simple significance threshold was applied, taking only the top five per cent of values. Second, a Monte-Carlo algorithm was used to permute (randomise) the haplogroup data 1000 times, and summed derivatives calculated for each permutation. This algorithm maintains the observed sample sizes and positions, and therefore controls for the conflated effects of sampling and heterogeneity in distances between sample sites in generating false positives. Grid points obtained with the original haplogroup data were then retained only if the values of their summed derivatives exceeded $95 \%$ of the values obtained from the permuted data. Values at the retained grid points can then be plotted on a map to show the position of significant barriers, which were colour coded to indicate the strength of the barrier. The positions of significant barriers were displayed on Delaunay triangulation connections (Brassel and Reif, 1979) between sample sites.

Previous analysis (Barbujani and Sokal, 1990) examining European populations, using autosomal markers identified 33 significant barriers based on 63 allele frequencies at 19 genetic loci, and 31 of these were also linguistic barriers, which can be broken down into 26 language and 5 dialect differences. In addition, 22 of 31 barriers also corresponded to physical barriers. From this analysis it was proposed that "language barriers may oppose the process of population admixture" (Barbujani and Sokal, 1990). Subsequent analysis of some of these data have suggested
possible reasons for genetic boundaries in addition to linguistic reasons; for example, boundaries were detected between the Netherlands and Germany, and northern and southern Germany, where no linguistic differences are found. A number of these populations appear to have long established and differing religions, customs and politics (Cavalli-Sforza et al., 1994). Linguistic differences would probably enhance genetic separation, because it is less likely that people would admix if they do not speak the same language (Cavalli-Sforza et al., 1994; Poloni et al., 1997).

### 3.3. Results

### 3.3.1. General observations

A total of 3677 Y chromosomes belonging to 48 European and circumEuropean populations (Figure 3.11) were successfully haplotyped using biallelic markers, and these were classified into haplogroups - for a summary of the population data see Table 3.2. Data on 129 North African Y chromosomes (Bosch et al., 1999) were also included (see materials and methods), giving a total of 3548. For the complete biallelic states for each sample analysed at the University of Leicester, see Appendix 1. The resulting frequency data have been summarised graphically for all populations, and superimposed onto a skeleton version of the maximum parsimony tree, see Figure 3.11A.

Two haplogroups (hg 4 and hg 7) are completely absent from the data set, which is consistent with published information. Previous work has shown that hg 4 (Figure 3.12) is generally restricted to eastern and central Asia, whereas hg 7 appears to be a sub-Saharan African lineage (Karafet et al., 1999). Another observation taken from these Y-chromosomal haplogroup data, is that certain haplogroups (hg 4, hg 12, and hg 26) are extremely rare or even absent from the majority of these populations sampled. These three Y-chromosomal lineages are known to be common within Asian populations (Underhill et al., 1997; Zerjal et al., 1997; Hammer et al., 1998; Karafet et al., 1999; Zerjal et al., 1999). All these haplogroups are ancestral and hold central positions within the maximum parsimony tree (Figure 3.6) whereas, the derived haplogroups (hg 21, hg 16, hg 1 and subgroups) that stem from these ancestral (hgs 4, 12 and 26) ones are frequent within these European populations (Table 3.2).

This pattern of haplogroup diversity may be a signature of a historical population movement from Asia, known as a range expansion (Templeton et al., 1995). This is when a small founding Asian population that contained all of these lineages, migrated towards Europe, and during this movement a population expansion occurred. Following the expansion, only a small number of these migrants continued their movement towards Europe, thus, leading to the low frequency of ancestral haplogroups. Another possibility is that these haplogroups have been 'diluted out' due to the large volume of derived chromosomes. However, even if these chromosomes were present it may be


Figure 3.11. Frequency distributions of Y-chromosomal haplogroups within Europe
(A) A skeleton version of the maximum parsimony tree with all the relevant colours for each haplogroup detected within this survey. The size of the circle is proportional to the sample analysed. A total of 3677 chromosomes were analysed. (B) All of the haplogroup data for each population location have been superimposed onto the European landscape. (C) The key showing the appropriate colour for each haplogroup. The frequency of the haplogroup is indicated by the size of the coloured sector contained within each pie chart, for each relevant sample.

Table 3.2. Biallelic haplogroup frequency data observed within 48 European and circum-European populations.

|  | Population | LF | LSF | n | hg1 | hg2 | hg3 | hg4 | hg7 | hg8 | hg9 | hg12 | hg16 | hg21 | hg22 | hg26 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| gdd | Greenlandic | EA | Esk | 61 | 42(69) | 17(27) | 0 | 0 | 0 | 0 | 1(2) | 0 | 0 | 1(2) | 0 | 0 |
| ice | Icelandic | IE | Ger | 28 | 13(46) | 9(32) | 6(21) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| saa | Saami | U | FU | 48 | 3(6) | 15(31) | 10(21) | 0 | 0 | 0 | 0 | 0 | 20(42) | 0 | 0 | 0 |
| swe | N.Swedish | IE | Ger | 48 | 11(23) | 22(48) | 9(19) | 0 | 0 | 0 | 1(2) | 0 | 4(8) | 1(2) | 0 | 0 |
| got | Gotlander | IE | Ger | 64 | 11(17) | 38(59) | 10(16) | 0 | 0 | 0 | 0 | 0 | 4(6) | 0 | 0 | 1(2) |
| nor | Norwegian | IE | Ger | 52 | 15(29) | 17(33) | 16(31) | 0 | 0 | 0 | 1(2) | 0 | 2(4) | 1(2) | 0 | 0 |
| dk | Danish | IE | Ger | 56 | 28(50) | 18(32) | 4(7) | 0 | 0 | 0 | 4(7) | 0 | 1(2) | 1(2) | 0 | 0 |
| fin | Finnish | U | FU | 57 | 1(2) | 13(23) | 6(10) | 0 | 0 | 0 | 0 | 1(2) | 35(61) | 1(2) | 0 | 0 |
| est | Estonian | U | FU | 207 | 18(9) | 30(14) | 56(27) | 0 | 0 | 0 | 2(1) | 8(4) | 76(37) | 6(3) | 0 | 11(5) |
| lat | Latvian | IE | BS | 34 | 5(15) | 4(12) | 14(41) | 0 | 0 | 0 | 0 | 0 | 11(32) | 0 | 0 | 0 |
| lit | Lithuanian | IE | BS | 38 | 2(5) | 5(13) | 13(34) | 0 | 0 | 0 | 0 | 0 | 18(47) | 0 | 0 | 0 |
| rus | Russian | IE | BS | 122 | 8(7) | 21(17) | 57(47) | 0 | 0 | 0 | 5(4) | 5(4) | 17(14) | 8(7) | 0 | 1(1) |
| brs | Belarusian | IE | BS | 41 | 4(10) | 14(34) | 16(39) | 0 | 0 | 0 | 1(2) | 0 | 1(2) | 4(10) | 0 | 1(2) |
| ukr | Ukrainian | IE | BS | 27 | 1(4) | 13(48) | 8(30) | 0 | 0 | 0 | 0 | 0 | 4(11) | 1(4) | 0 | 0 |
| mar | Mari | U | FU | 48 | 5(10) | 2(4) | 14(29) | 0 | 0 | 0 | 3(6) | 8(17) | 16(33) | 0 | 0 | 0 |
| chu | Chuvash | A | Tu | 17 | 2(12) | 4(24) | 3(18) | 0 | 0 | 0 | 1(6) | 0 | 3(18) | 1(6) | 0 | 3(18) |
| geo | Georgian | Cau | S.Ca | 64 | 12(19) | 31(48) | 4(6) | 0 | 0 | 0 | 15(23) | 0 | 0 | 1(2) | 0 | 1(2) |
| oss | Ossetian | Cau | S.Ca | 47 | 20(43) | 5(11) | 1(2) | 0 | 0 | 0 | 16(34) | 0 | 0 | 3(6) | 0 | 2(4) |
| arm | Armenian | IE | Arm | 89 | 22(25) | 28(31) | 5(6) | 0 | 0 | 0 | 26(29) | 0 | 3(3) | 3(3) | 0 | 2(2) |
| tur | Turkish | A | Tu | 167 | 34(20) | 41(25) | 8(5) | 0 | 0 | 0 | 55(33) | 2(1) | 2(1) | 17(10) | 0 | 8(5) |
| cyp | Cypriot | IE | Gk | 45 | 4(9) | 10(22) | 1(2) | 0 | 0 | 0 | 15(33) | 1(2) | 0 | 12(27) | 0 | 2(4) |
| gk | Greek | IE | Gk | 36 | 4(11) | 8(22) | 3(8) | 0 | 0 | 0 | 10(28) | 0 | 0 | 10(28) | 0 | 1 (3) |
| bul | Bulgarian | IE | BS | 24 | 4(17) | 10(42) | 3(12) | 0 | 0 | 0 | 3(12) | 0 | 0 | 4(17) | 0 | 0 |
| cze | Czech | IE | BS | 53 | 10(19) | 10(19) | 20(38) | 0 | 0 | 0 | 6(11) | 3(6) | 0 | 4(8) | 0 | 0 |
| slk | Slovakian | IE | BS | 70 | 12(17) | 12(17) | 33(47) | 0 | 0 | 0 | 2(3) | 1(1) | 2(3) | 7(10) | 0 | 1(1) |
| rom | Romanian | IE | It | 45 | 8(18) | 12(27) | 9(20) | 0 | 0 | 0 | 11(24) | 0 | 0 | 3(7) | 1(2) | 1(2) |


| yug | Yugoslavian | IE | BS | 100 | $11(11)$ | $49(49)$ | $16(16)$ | 0 | 0 | 0 | $8(8)$ | $2(2)$ | 0 | $13(13)$ | 0 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| sln | Slovenian | IE | BS | 70 | $15(21)$ | $19(27)$ | $26(37)$ | 0 | 0 | 0 | $4(6)$ | 0 | 0 | $5(7)$ | $1(1)$ |
| hun | Hungarian | U | FU | 36 | $11(30)$ | $10(28)$ | $8(22)$ | 0 | 0 | 0 | $1(3)$ | 0 | 0 | $6(17)$ | 0 |
| pol | Polish | IE | BS | 112 | $20(18)$ | $19(17)$ | $61(54)$ | 0 | 0 | 0 | $4(4)$ | $1(1)$ | $5(4)$ | $2(2)$ | 0 |
| ita | Italian | IE | It | 99 | $44(44)$ | $14(14)$ | $2(2)$ | 0 | 0 | 0 | $20(20)$ | 0 | 0 | $13(13)$ | 0 |
| sar | Sardinian | IE | It | 10 | $3(30)$ | $4(40)$ | 0 | 0 | 0 | $1(10)$ | 0 | 0 | 0 | $2(20)$ | 0 |
| bav | Bavarian | IE | Ger | 80 | $38(48)$ | $18(23)$ | $12(15)$ | 0 | 0 | 0 | $4(5)$ | 0 | 0 | $6(8)$ | $2(3)$ |
| ger | German | IE | Ger | 30 | $12(40)$ | $6(20)$ | $9(30)$ | 0 | 0 | 0 | $1(3)$ | 0 | $1(3)$ | 0 | 0 |
| dut | Dutch | IE | Ger | 84 | $36(43)$ | $27(32)$ | $11(13)$ | 0 | 0 | 0 | $6(7)$ | 0 | 0 | $3(8)$ | $1(1)$ |
| fra | French | IE | It | 40 | $20(50)$ | $10(25)$ | $2(5)$ | 0 | 0 | $1(3)$ | $2(5)$ | 0 | 0 | $3(8)$ | $2(5)$ |
| bgm | Belgian | IE | Ger | 92 | $58(63)$ | $21(23)$ | $4(4)$ | 0 | 0 | 0 | $5(5)$ | 0 | 0 | $2(2)$ | $1(1)$ |
| scw | W.Scottish | IE | Cel | 120 | $87(72)$ | $23(19)$ | $8(7)$ | 0 | 0 | 0 | 0 | 0 | 0 | $2(2)$ | 0 |
| scm | Scottish | IE | Cel | 43 | $34(79)$ | $5(12)$ | $3(7)$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| enw | Cornish | IE | Cel | 51 | $42(82)$ | $9(18)$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ene | E.Anglian | IE | Ger | 172 | $97(56)$ | $52(30)$ | $15(9)$ | 0 | 0 | 0 | $1(1)$ | 0 | 0 | $5(3)$ | $1(1)$ |
| irl | Irish | IE | Cel | 257 | $207(81)$ | $39(15)$ | $2(1)$ | 0 | 0 | 0 | $2(1)$ | 0 | $1(0.5)$ | $6(2)$ | 0 |
| bas | Basque | Bsq | Bsq | 26 | $19(73)$ | $2(8)$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $5(19)$ |
| apa | Spanish | IE | It | 126 | $86(68)$ | $17(13)$ | $3(2)$ | 0 | 0 | 0 | $4(3)$ | 0 | 0 | $12(10)$ | $3(2)$ |
| spa | $1(1)$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| pos | S.Portuguese | IE | It | 57 | $32(56)$ | $8(14)$ | $1(2)$ | 0 | 0 | 0 | $5(9)$ | 0 | 0 | $10(17)$ | 0 |
| pon | N.Portuguese | IE | It | 328 | $203(62)$ | $54(16)$ | 0 | 0 | 0 | 0 | $21(6)$ | 0 | 0 | $3(2)$ |  |
| alg | Algerian | AA | Sem | 27 | 0 | $1(4)$ | 0 | 0 | 0 | $1(4)$ | $11(41)$ | 0 | 0 | $14(52)$ | 0 |
| naf | N.African | AA | mix | 129 | $5(4)$ | $4(3)$ | 0 | 0 | 0 | $6(5)$ | $15(12)$ | 0 | 0 | $99(77)$ | 0 |
|  |  |  | Total | 3677 | $1379(38)$ | $820(22)$ | $512(14)$ | 0 | 0 | $9(0.2)$ | $292(8)$ | $32(0.8)$ | $226(6)$ | $327(9)$ | $23(0.6)$ |
|  | $57(1.5)$ |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Abbreviations for language families (LF) are as follows: IE (Indo-European), EA (Eskimo-Aleut), U (Uralic), A (Altaic), Cau (Caucasian), AA (Afro-Asiatic), and Bsq (Basque). Language sub-families (LSF): Esk (Eskimo), Ger (Germanic), FU (Finno-Ugric), BS (Balto-Slavic), Tu (Turkic), S.Cau (South Caucasian), Arm (Armenian), Gk (Greek), It (Italic), Cel (Celtic), Bsq (Basque), Sem (Semitic), and mix includes (Berber and Semitic).
Numbers in the haplogroup columns are the actual number of chromosomes detected followed by the frequency in brackets.

## Haplogroup 4



B


Figure 3.12. Frequency distributions of haplogroup 4 across Europe
(A) Indicates the position of the haplogroup within the maximum parsimony tree. (B) The Y chromosome hg 4 data have been superimposed onto the European landscape, and each pie chart has been positioned on the relevant population location. No examples of this lineage were detected in any of the 3677 chromosomes analysed. (C) The world wide distribution of the haplogroup is also shown.
nearly impossible to detect them even with additional markers, due to their extremely low frequency. The large number of derived chromosomes within Europe, may suggest they have a greater chance of being passed on to further generations, than a few ancestral chromosomes (that originated in Asia), unless events such as genetic bottlenecks occurred causing the local frequency of commonly detected chromosomes to fall dramatically. Expansion, such as the one discussed can be statistically assessed, using an approach called nested cladistic analysis (Templeton et al., 1995), and these approaches have been used in previous studies (Hammer et al., 1997), but unfortunately, the lack of Asian samples precludes this type of analysis during this project. However, this pattern could be a result of an ascertainment bias (which samples were analysed) or the use of a small number of markers, and this effect may be lost if the number of markers were increased, which is one reason why more markers are required for future analyses.

### 3.3.2. Clinal distribution of $Y$-chromosomal lineages within Europe

The frequency distribution observed in Figure 3.11A, is a combination of all the populations, and no single population analysed has a frequency distribution resembling that of the overall sample, emphasising the strong geographical differentiation of Y -chromosomal variation in Europe. This is illustrated in Figure 3.11B, where the frequency data for all haplogroups can be observed superimposed onto a geographic landscape. The distributions of haplogroups are clinal and highly non-random, with for example, a concentration of hg 1 chromosomes in the west, hg 9 chromosomes in the southeast, hg 16 chromosomes in the north-east, and hg 3 chromosomes in central and eastern Europe. Even though clinal variation had been assessed qualitatively, a more rigorous statistical approach was required to determine the clines quantitatively.

Statistical analyses known as AIDA (Bertorelle and Barbujani, 1995) and spatial autocorrelation (SAAP; Sokal and Oden, 1978) were undertaken. AIDA was initially used (Bertorelle and Barbujani, 1995), which takes into account molecular distances between haplogroups and provides autocorrelation indices (Moran's II) for the entire data set, including the rare haplogroups. The pattern (Figure 3.13) is strongly clinal, recognised as a change from positive to negative autocorrelation indices with increasing distance class.


Figure 3.13. AIDA spatial autocorrelation analysis on the European data This pattern is strongly clinal, shown by a change from positive to negative autocorrelation indices with increasing distance class.

The SAAP analysis (Table 3.3), omitting low frequency haplogroups (hgs $8,12,22$ and 26), confirms this clinal pattern, and reveals information about individual lineages (Figure 3.14A-F).

Table 3.3. SAAP analysis for 47 European and circum-European populations, and the figures represent Moran's I.

| Km | 400 | 800 | 1200 | 1600 | 2000 | 2500 | 3000 | 3500 | 4000 | 4931 | p values | Result |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Pairs | 34 | 110 | 140 | 179 | 165 | 174 | 113 | 68 | 30 | 22 |  |  |
| hg1 | $0.79^{* *}$ | $0.65^{* *}$ | $0.33^{* *}$ | $0.23^{* *}$ | -0.12 | $-0.34^{* *}$ | $-0.47^{* *}$ | $-0.69^{* *}$ | $-0.63^{* *}$ | $-0.56^{* *}$ | $\mathrm{p}=0.000$ | Cline |
| hg2 | 0.11 | -0.03 | -0.04 | 0.06 | -0.01 | $-0.14^{*}$ | -0.08 | 0.10 | -0.00 | -0.00 | $\mathrm{p}=0.357$ | N. C. |
| hg3 | $0.69^{* *}$ | $0.44^{* *}$ | $0.18^{* *}$ | $-0 / 02$ | $-0.19^{* *}$ | $-0.41^{* *}$ | -0.15 | -0.16 | $0.33^{*}$ | 0.15 | $\mathrm{p}=0.000$ | Cline |
| hg9 | $0.58^{* *}$ | $0.32^{* *}$ | $0.25^{* *}$ | $0.16^{* *}$ | 0.006 | $-0.26^{* *}$ | $-0.28^{* *}$ | $-0.42^{* *}$ | $-1.07^{* *}$ | $-0.68^{* *}$ | $\mathrm{p}=0.000$ | Cline |
| hg16 | $0.81^{* *}$ | $0.28^{* *}$ | $0.14^{*}$ | $0.12^{* *}$ | $-0.18^{* *}$ | $-0.26^{* *}$ | $-0.22^{* *}$ | $-0.32^{* *}$ | -0.01 | -0.02 | $\mathrm{p}=0.000$ | Cline |
| hg21 | $0.30^{* *}$ | $0.31^{* *}$ | $0.22^{* *}$ | 0.05 | -0.09 | $-0.16^{*}$ | $-0.31^{* *}$ | $-0.33^{*}$ | 0.07 | $-0.44^{*}$ | $\mathrm{p}=0.001$ | LDD |

The distance values (km) are shown on the top row. * statistical significance to $1 \%$ level ( $p=0.01$ ), and ${ }^{* *}$ is to $5 \%$ level ( $p=0.05$ ). N.C. - not clinal, LDD (long-distance differentiation)

The distributions of all of the haplogroups examined, with the exception of hg 2, are strongly clinal (Figure 3.14A-F), confirming the visual impression given by Figure 3.11. In two cases (hg 3 and hg 16), values become positive or zero in the longest distance class (a 'depression') and indicate a regional influence of these clines, rather than being continental wide.

Haplogroup 2 is the most ancestral lineage observed within Europe (Figure 3.15) and lies at a 'star-like' node within the tree. Chromosomes within this haplogroup are therefore essentially undefined, and are likely to consist of a set of discrete sublineages which themselves probably have greater geographical coherence that cannot be currently be separated using the present range of markers. New markers have just been published that will allow the separation of these chromosomes (Underhill et al., 2000; and see Chapter 6). Consistent with this, hg 2 chromosomes are widely distributed across the whole landscape, and comprise the only high-frequency lineage which does not show clinal variation (Figure 3.14B). This is not surprising considering the above remarks regarding its phylogenetic location within the tree (Figure 3.15) and the presence of multiple undefined lineages. Due to this uninformativeness, hg 2 will not be further considered here. Haplogroup 26 is detected at low frequency; again like hg 2, due to the phylogenetic position within the tree (Figure 3.16) it probably contains unidentified coherent sublineages.


Figure 3.14. SAAP analysis correlograms for six frequent haplogroups SAAP correlograms for the six most frequent haplogroups (A-F) within 47 different populations. The significance of each point is indicated by its symbol, and the overall significance of each correlogram is also given. 'LDD; - long-distance differentiation. In all correlograms the x -axes show distance classes in km .


Figure 3.15. Frequency distribution of haplogroup 2 across Europe
(A) Indicates the position of the haplogroup on a reduced version of the maximum parsimony tree. (B) Y-chromosomal hg 2 data have been superimposed onto the European landscape. Each pie chart has been positioned on the appropriate population location, with the frequency of the hg 2 chromosomes being indicated as a grey sector of the pie chart. The worldwide distribution of the haplogroup is also indicated in part (C).

Haplogroup 26



Figure 3.16. Frequency distributions of haplogroup 26 across Europe
(A) Indicates the position of the haplogroup within the reduced version of the tree. (B) The Y-chromosomal hg 26 data have been superimposed onto the European landscape. Each pie chart has been positioned on the appropriate population location, and the frequency of the hg 26 lineage detected within each population is indicated by the size of the turquoise segment within each pie chart. The world wide distribution of the haplogroup is also shown in part (C)

Haplogroup 8 is common in sub-Saharan African populations (Karafet et al., 1999), and present in the North African samples at approximately 5\% (Figure 3.17). Only two European examples were detected, in Sardinia and France, which may represent recent African admixture. Haplogroup 22 chromosomes are on the whole rare and reach appreciable frequencies only in the French (5\%) and Basques (19\%) (Figure 3.18). This haplogroup has been analysed in detail in a previous study (Hurles et al., 1999), which suggested that it has a recent Iberian origin, and that non-Iberian examples represent migrants, and the distribution here is consistent with this analysis.

### 3.3.3. A major cline consistent with the demic diffusion hypothesis

Haplogroups 1 and 9 show complementary clines on the continental scale (Figure 3.19), from the south east of Europe to the north west. The clines are represented quantitatively as correlograms, see Figure 3.14A and D. The hg 1 lineage appears to have a strong focus in the west, which was emphasised in an additional study: when Irish samples were sub-divided on the basis of geographical information contained within surnames (Hill et al., 2000), hg 1 reached near-fixation (98.5\%) in western Ireland. Haplogroup 9 reaches its highest frequencies ( $\sim 33 \%$ ) in the south east, around the Caucasus and in Anatolia, where it is proposed that agriculture originated (Cavalli-Sforza et al., 1994; Piazza et al., 1995). The strong clinal pattern of these two haplogroups, which together account for almost half of the chromosomes in this study (45\%), resembles the synthetic map of the first principal component (PC1) of genetic variation of 95 classical loci (Figure 3.20). The PC data have been taken to support the demic diffusion hypothesis (Cavalli-Sforza et al., 1994) proposed to explain the expansion of farming from the Near East (Ammerman and CavalliSforza, 1984).

It is known from previous analysis that approximately $80 \%$ of hg 1 chromosomes also possess the 49f/TaqI haplotype 15 (Jobling et al., 1994), and that the 12f2/TaqI 8 kb allele exactly corresponds to hg 9 ; both of these have been claimed to also support demic diffusion (Semino et al., 1996). Both of these lineages (hg 1 and hg 9) are old enough to represent existing pre-agricultural European chromosomes, although some may represent the migrating farmers. It has been proposed that hg 1 chromosomes could be Palaeolithic in origin (i.e. pre-date the event and spread of agriculture (Richards et al., 1996; Sykes et al.,

## Haplogroup 8




Figure 3.17. Frequency distribution of haplogroup 8 across Europe
(A) Indicates the position of the haplogroup within the reduced version of the tree. (B) The Y -chromosomal hg 8 data have been superimposed onto the European landscape, and each pie chart has been positioned on the appropriate population location. The world wide distribution of the haplogroup is also shown in part (C) and this haplogroup is mainly confined to Sub-Saharan African populations. The few examples of this lineage detected within Europe probably represent recent African admixture.

## Haplogroup 22




Figure 3.18. Frequency distribution of haplogroup 22 across Europe
(A) Indicates the position of the haplogroup within the reduced version of the tree. (B) The Y-chromosomal hg 22 data have been superimposed onto the European landscape, and each pie chart has been positioned on the relevant population location. The frequency of this lineage within the samples analysed is indicated by the pale blue sector of the pie chart. This haplogroup is not detected in all European populations. The world wide distribution of this haplogroup is also shown in part (C), and appears mainly isolated to south-west Europe and in part of South America.

## Haplogroup 1


Haplogroup 9



Figure 3.19. Frequency distributions of haplogroups 1 and 9 across Europe The combination of haplogroups 1 and 9 data on 48 European and circum-European populations, show clinal patterns that appear to be consistent with the demic diffusion model, reflecting agricultural dispersal from the Near East. (A+D) indicates the location of the haplogroup within the maximum parsimony tree. ( $\mathrm{B}+\mathrm{E}$ ) the area of the circles is proportional to the sample size, the dark coloured segments indicate the frequency of that particular haplogroup detected, and the position indicates the sample site location. $(\mathrm{C}+\mathrm{F})$ indicates the global distribution of each haplogroup.


Figure 3.20. Synthetic maps generated for PC1 and PC2 from 'classical' marker data These synthetic maps were generated using data from classical markers superimposed onto a European and western Asian landscape. Map (A) uses the values generated for the first principal component (PC1), and map (B) uses values generated for PC2. These clinal variations have been assigned to a number of different migration events during pre-history (Cavalli-Sforza et al., 1994), although this must be done with caution because no dates can be assigned to these patterns. PC1 has been taken to support demic diffusion, and PC2 has been under interpreted but has been taken by some to show a climatic influence from the south.
1996), and coalescence analysis enabled an approximate date of 23,000 years before present (YBP) (Karafet et al., 1999) to be determined. Haplogroup 9 has been approximately dated to 14,800 YBP ( $\pm 9,700$ years) (Hammer et al., 2000).

However, distributions of the remaining haplogroups are very different to these, and cannot be interpreted as a simple reflection of population movement from the Near East.

### 3.3.4. A north east to south west cline may represent a signal of an expansion from north of the Black Sea

The distribution of hg 3 chromosomes is strongly clinal (Figure 3.14C) but with a focus centred in the north-east of Europe, spreading to the south-west, that is more on a regional scale, and is likely to reflect population-historical events distinct from those responsible for the distributions of hgs 1 and 9. It reaches its highest frequencies in central and eastern Europe comprising around half of chromosomes in the Russian, Polish and Slovakian samples; frequencies in the south-east and south-west are low. Previous Y chromosome microsatellite analysis has shown that this hg is relatively young when compared to hg 1 or hg 9, with an age of 3,800 YBP ( $95 \%$ confidence limits of 1,600 to $13,000 \mathrm{YBP}$ ) (Zerjal et al., 1999). This hg is commonly found within Europe and Asia, but appears absent from Africa and the American continent (Zerjal et al., 1999), and this evidence appears to support a recent expansion from Eurasia.

This clinal distribution (Figure 3.21A) resembles the synthetic map of Europe based on the third principal component (PC3) of classical gene frequency data. (Figure 3.21B), which has been interpreted by some geneticists (CavalliSforza et al., 1994) as marking the movement, around 7,000 YBP, from north of the Caspian Sea along the Russian Steppes, of the Kurgan people, pastoral nomads who are thought to have domesticated the horse. It has also been proposed they may have been responsible for the spread of the Indo-European language into Europe (Gimbutas, 1970). Although some people believe that the migrating farmers may have brought Indo-European with them (Renfrew, 2000), there is no direct evidence that Indo-European was spoken by farmers (Gimbutas, 1970). Another discrepancy is observed with relation to the timing of these historical events, because the spread of agriculture is proposed to have occurred approximately 10,000 years ago, and the Kurgan expansion occurred more recently (approximately 7,000 years ago). This is a problem with the patterns


Figure 3.21. Frequency distribution of haplogroup 3 and a synthetic map generated from PC3 of 'classical' marker data
Shows the Y-chromosomal hg 3 data, its position within the tree, and its global distribution (A). This is compared to the synthetic map (B) generated from the third principal component (PC3) by Cavalli-Sforza et al., 1994. These two patterns appear similar and the PC3 distribution has been taken by Cavalli-Sforza and co-workers to support the Kurgan expansion of pastoral nomads approximately 7,000 years ago.
generated from PCA data, or even most genetic data, because unless alleles can be dated accurately, then the linking of genetic patterns to archaeological/historical events is always going to be dubious, and often frowned upon by archaeologists. Patterns generated from PCA analysis have been readily accepted and included in a number of genetic studies in order to prove what the authors want them to.

### 3.3.5. A north south cline may represent influence from North African populations

Within Europe, hg 21 chromosomes are concentrated in the south (Figure 3.22). A north south-cline is detectable with their frequency in the two Northern African samples being very high ( 52 and $77 \%$ ). The frequencies in the Greek and Cypriot samples also high ( $\sim 27 \%$ ), which might reflect a barrier to gene flow between Africa and Europe, as also shown by the analysis of autosomal protein markers (Simoni et al., 1999) and microsatellites (Bosch et al., 2000). In other southern European populations, such as those in Spain, Portugal, Sardinia, Italy, Turkey and Yugoslavia, frequencies are in the range of $10-20 \%$ and this may indicate a significant Northern African influence. The decline in frequencies to the north is rather uniform. This regional cline (Figure 3.14F) appears similar to that detected in the second principal component (PC2) of classical gene frequencies (Figure 3.22) (Cavalli-Sforza et al., 1994), which has been interpreted on a climatic basis. It is known from previous analysis examining the MSY1 diversity within this haplogroup, that diversity is high, and contains population specific lineages (Bouzekri et al., 1998; Jobling et al., 1998a). For a more detailed examination of this haplogroup, see Chapter 5.

### 3.3.6. A haplogroup localised in the north east of Europe may reflect the influence of Uralic speakers

Haplogroup 16 (Figure 3.23B) is at high frequency in the north, east of the Baltic Sea, a distribution consistent with that noticed previously in a global survey (Zerjal et al., 1997). Its pattern is again clinal, but regional (Figure 3.14E). Haplogroup 12 (Figure 3.23A), the ancestral haplogroup to hg 16 (Figure 3.6), is at low frequency in the sample overall. However, its distribution overlaps that of hg 16, with no examples in the western half of the continent, and concentrated more in the south. It is most frequent (17\%) in the Mari, who may be the population of origin of the Tat mutation, which defines hg 16 (Tatiana Zerjal and Chris Tyler-Smith, unpublished data).
Haplogroup 21

C




Figure 3.22. Frequency distribution of haplogroup 21 and a synthetic map from PC2 of 'classical' marker data
(A) Indicates the position of this haplogroup within the tree. (B) hg 21 data superimposed onto the European landscape, and each pie chart has been positioned on the relevant population location. The frequency of this haplogroup is indicated by the orange segments within each pie chart. (C) The world wide distribution of the haplogroup is also shown. (D) shows the synthetic map from PC2 data that appears to show similar patterns to the hg 21 data.
Haplogroup 12





Figure 3.23. Frequency distribution of haplogroups 12 and 16 across Europe Their respective positions within the tree are indicated (A+D). Haplogroups 12 and hg 16 data have been superimposed onto a European landscape ( $B+E$ ); each pie chart shows the frequency of each haplogroup detected as a coloured sector, and these pies have been positioned on the appropriate population location. The global distribution of each haplogroup is indicated ( $\mathrm{C}+\mathrm{F}$ ).

All except one of the Uralic speaking populations analysed (Finnish, Estonians, Saami and the Mari) show high frequencies of hg 16 chromosomes. The exception is the Hungarians, who acquired their Uralic language via elite dominance from the Magyars approximately 1,100 YBP (Cavalli-Sforza et al., 1994), and so the Hungarians do not share a common genetic origin with the other Baltic populations. However, two nearby Indo-European speaking populations, the Lithuanians and Latvians, also show hg 16 at high frequency for this lineage at least, the association appears to be geographical, rather than linguistic. In the following section methods to summarise variation among all lineages are used to examine this issue in more detail.

### 3.3.7. Which has the stronger affect on European Y-chromosomal genetic differentiation,

 geography or language?The above suggestion that both language and geography can influence the genetic makeup of modern European populations is now discussed. It has been previously observed that a correlation between linguistic and genetic boundaries occurs within Europe (Barbujani and Sokal, 1990). In 22 out of 33 cases, these boundaries also correlate with physical barriers, and it is thought they may be the cause of both the genetic and linguistic boundaries (Piazza et al., 1995). The aim of this statistical analysis on the Y-chromosomal haplogroup data generated for the European and circum-European populations is to determine if either language or geography has more of an influence on genetic barriers. A number of statistical approaches have been employed, including principal components analysis (PCA), Mantel tests, and genetic barrier analysis. Each of these will be discussed in turn. The Greenlandic Inuit were excluded from these analyses due to their geographic outlying position with respect to the European continent. The inclusion of a linguistic isolate, the Ladin-speakers, within this data-set will be examined in Chapter 4.

### 3.3.7.1. Population comparisons through principal components analysis

Yuri Dubrova (University of Leicester) carried out the principal components analysis (PCA) using the Statistica package. This type of statistical analysis enables a large amount of data to be summarised into a few values that when combined encompasses nearly all the observed diversity. Unfortunately, a degree of resolution is lost during this analysis but this is quite small and does not affect the overall picture of the genetic diversity. This type of analysis was extremely useful during this project,
due to the sheer quantity of data. Figure 3.24 shows the results of a PC analysis of the $Y$ haplogroup data, in which populations are labelled according to linguistic affiliation. PC1 and PC2 are plotted against each other (Figure 3.24A) and together summarise 71.4\% of the variance. PC2 and PC3 are plotted against each other (Figure 3.24B) and PC3 incorporates an additional $11.5 \%$ of variance (total $82.9 \%$ ).

The major division is between the two populations from Northern Africa and the others. This is unsurprising given their high frequencies of hg 21 and hg 9 , and near absence of hg 1 , and indicates that the Mediterranean, even at its narrowest point, has represented a barrier to gene flow, as has been suggested previously from autosomal DNA analysis. The Mediterranean populations of Greece and Cyprus occupy an intermediate position between the Northern Africans and the rest.

Basques speak a non-IE language unrelated to any other language (Ruhlen, 1991), and thus represent the most striking example of a linguistic isolate in Europe. This isolation seems to be reflected in the PC analysis, where they are separated from other populations (Figure 3.24B); however, this may be due to their high frequency of a young lineage ( hg 22; Hurles et al., 1999), rare elsewhere, rather than persistence of ancient ones. Their closest neighbours in the PC analysis are not the geographically close Iberian populations, but those from along the western coast of Europe, 'the Atlantic fringe', most of which are Celtic-IE speaking. This may suggest they are not as genetically distinct as they would like to believe, as been proposed in other Y-chromosomal studies (Hurles et al., 1999). In this context, the Cornish sample ('enw') is grouped not with the eastern English sample ('ene'), but with the Scottish and Irish - a reflection of geography, or of the original Celtic language of this region (Ruhlen, 1991), or both.

Among Uralic-speaking populations, this analysis confirms the impression given by Figure 3.11: with the exception of the Hungarians, who lie close to IE speakers, these populations are grouped together, with the Finns separated from the rest in PC3. Also within this group are the Lithuanians and Latvians, supporting the idea that this is primarily a geographic association.


Figure 3.24. PCA plots encompassing 47 European populations
Two examples of scatterplots generated from principal component analysis data on 47 populations (excluding the Greenlandic sample). (A) shows PC1 and PC2, and (B) shows PC2 and PC3. The linguistic affiliations are indicated in the key for all languages spoken by the different populations.

The overall impression from Figure 3.24 is that populations form groups, and that geographical proximity may be a better predictor of Ychromosomal genetic affinity than language. In addition to the examples discussed above, the Italic-IE speaking Romanians are distant from other Italic speakers, and the Altaic speaking Turkish lie between the geographically neighbouring but linguistically distant Indo-European speaking, Armenians and Greeks. However, there are exceptions who neither share close geographical proximity or a common language. Italians cluster with Ossetians; however the Ossetians speak Caucasian languages and the Italians Italic-Indo-European.

### 3.3.7.2. Correlating geography, language and genetics through Mantel testing

Mantel tests (Mantel, 1967) provide an objective way of assessing the relative importance of different factors in shaping genetic diversity. Using this method, correlation coefficients between pairs of factors (from genetics, geography and language) can be calculated, together with significance values. Partial correlation coefficients are then calculated between genetics and geography, and genetics and language, keeping the third factor constant, and thus controlling for the strength of the correlation between geography and language. The populations from North Africa are linguistically remote and geographically peripheral, and the PC analysis has shown their genetic differentiation. They were therefore excluded from the Mantel analysis, in order to just examine effects within Europe itself. The values generated during this analysis can be seen in Table 3.4.

Table 3.4. Correlation and partial correlation coefficients between genetic, geographic, and linguistic distance

| Distances considered | Correlation coefficient | p value |
| :--- | :---: | :---: |
| Genetics and geography | 0.387 | $<0.001$ |
| Genetics and language | 0.198 | $<0.01$ |
| Genetics and geography, 0.349 $<0.001$ <br> with language held constant <br> Genetics and language, with <br> geography held constant 0.088 not significant |  |  |

The correlation between genetics and geography is highly statistically significant ( $\mathrm{p}<0.001$ ), while the correlation between genetics and language is less strong, but still significant ( $p=0.014$ ). The partial correlation of genetics and geography, keeping language constant, is again strong and significant ( $\mathrm{p}<0.001$ ); in contrast, the partial correlation of genetics and language is low and non-significant ( $p=0.095$ ). The effect of changing the values assigned to distances within Uralic and Altaic, and between Irish and Scottish (see materials and methods), had a negligible influence upon these results. Increasing the distance assigned between language families had the effect of reducing still further the partial correlation between genetics and language, and its significance.

Correlations involving geography and language explain only 16.8\% of the total genetic variance exhibited by the haplogroup data, and so it is likely that other forces such as population movements, drift, and mating practices must also have importance influences on genetic affiliation.

### 3.3.7.3. Locating Y-chromosomal genetic barriers within Europe

While the analysis above indicates a lack of large-scale correlation between language and genetics, it does not address local genetic differentiation, which may reflect local effects of language. Genetic barrier analysis, which locates the zones of sharpest genetic change within a landscape, provides a way to do this. The genetic barriers can be superimposed onto a European geographic landscape, but the interpretation of such barriers can be subjective; for example, in previous studies a linguistic barrier has been considered as dialect differences (Barbujani and Sokal, 1990).

Figure 3.25A shows the results of a genetic barrier analysis on the $Y$ haplogroup data for 45 populations, taking the top $5 \%$ of barriers, and then using a $95 \%$ significance filter (see materials and methods). Within western Europe minor barriers separate the Basques from some neighbouring populations, the western from the eastern English, and the Dutch from the Belgians. In the east, there are two strong barriers, which appear to correlate with linguistic barriers, one between the Uralicspeaking Mari and Altaic-speaking Chuvash, and one between the Georgians and Ossetians, who speak languages belonging to different


Figure 3.25. Significant Y-chromosomal genetic barriers within Europe (A) output from the Orinoco program. Positions of genetic barriers showing 95\% significance after randomisation. (B) barriers (thick black lines) have been positioned on Delaunay triangles (thin line) between sample sites.
families, and are also separated by the Caucasus mountain range. Most of the major barriers lie in the middle of the European landscape, running from Italy in the south to the Baltic in the north, including one around the island population of Gotland.

To what extent are linguistic differences contributing to Y chromosomal barriers within Europe? Since 37 different languages are spoken among the 45 sample sites, it is expected that most genetic barriers should fall between populations speaking different languages. However, if language differences do constitute barriers to gene flow, then it might be expected that the degree of linguistic difference between a pair of populations should correlate with the chance of a genetic barrier occurring. For example, the greatest proportion of genetic barriers should fall between populations speaking languages from different families, a lesser proportion between those speaking languages from different subfamilies, and the least between those speaking languages within a subfamily.

There were a total of 122 Delaunay connections and (Figure 3.25B) 48 of these were crossed by a genetic barrier. The proportion of connections crossed by a genetic barrier in each of the three classes were counted, between language families, between sub-families, and within sub-families; (see Table 3.5) these values are $46.2 \%$ (18/39), 40.5\% (15/37) and $32.6 \%(15 / 46)$ respectively.

Table 3.5. Genetic barriers analysed by level of linguistic difference

|  | Between <br> families | Between sub- <br> families | Within sub- <br> families |
| :--- | :---: | :---: | :---: |
| Number of connections <br> crossed by genetic barriers | 18 | 15 | 15 |
| Number of connections not <br> crossed by genetic barriers | 21 | 22 | 31 |
| Total |  |  |  |

While the ranking of these three values is that expected under the hypothesis, differences between them are not significant ( $p>0.1$, three-way chi-square test). This suggests that language may not be the primary force contributing to genetic barriers here. However, this analysis does not take into account the fact that two non-Indo-European languages, Hungarian and Turkish, have been acquired recently: the PC analysis, and the relative absence of Y -chromosomal genetic barriers around these populations supports the idea that elite dominance was not accompanied by extensive genetic admixture. If two of these populations are removed and the above analysis repeated, differences between the proportions increase (to 50.0\% [13/26], 43.2\% [19/44] and 31.9\% [15/47] respectively), but remain nonsignificant ( $\mathrm{p}>0.1$ ).

### 3.4. Discussion

### 3.4.1. General discussion

This work describes a detailed survey of human Y-chromosomal diversity within Europe. Samples were distributed over most of the continent, including its western and eastern fringes; inclusion of these regions, omitted from some other studies, has allowed the detection of influences from the east, and clines extending to the extreme west, for example. However, some regions remain poorly sampled, and if the possible effects of local differentiation are to be studied, more extensive sampling is needed. At the eastern edge of Europe lie the steppes, which stretch uninterrupted to China. Analogous studies of Asian Y-chromosomal diversity are needed to place the European data in a broader context.

Eleven biallelic markers were used in this study, but there is still a need for more. For instance hg 2, constituting $22 \%$ of the total sample, and as much as $49 \%$ in the sample from Yugoslavia, is poorly defined, and therefore constitutes a potential source of error in these analyses, since equal weight is given both to this, as to well-defined haplogroups. The pace of new marker discovery is increasing (Underhill et al., 1997; Shen et al., 2000), and soon the resources will be available to adequately define all major European lineages. (See Semino et al., (2000) and Chapter 6).

The methods used to type the polymorphisms in this Chapter are timeconsuming and non-parallel. Because of this limitation, not all markers within the tree were typed, and this may represent a bias in the study. However, two haplogroups (hgs 4 and 7) were typed 'by default' which previous studies had suggested would not be prevalent in Europe. They were completely absent, suggesting that the bias may not be severe. New methods are required to type markers more efficiently (see Chapter 5).

Consistent with global surveys (Underhill et al., 1997; Karafet et al., 1999), this continental study confirms the high degree of geographical differentiation of Y-chromosomal lineages. This differentiation makes the Y chromosome a sensitive indicator of admixture, as demonstrated in studies of Polynesia (Hurles et al., 1998), South America (Bianchi et al., 1997) and Uruguay (Bravi et al., 1997),
for example, or of an absence of admixture, as shown in Jewish populations in Europe and North Africa (Hammer et al., 2000). Knowledge about admixture is of particular importance in the choice of populations for studies which use linkage disequilibrium analysis (McKeigue, 1997) in both simple and complex disorders.

### 3.4.2. Clines of $Y$-chromosomal haplogroups

The effects of drift on human $Y$ chromosome diversity are likely to be great. It is striking, therefore, to observe clear clinal variation in five of the six major lineages within Europe, and this suggests that drift has not erased the patterns of variation established by past population movement. Natural selection upon Y chromosomes (Jobling and Tyler-Smith, 2000) provides an alternative explanation for such clines; possible effects of geographically variable factors (such as temperature) on fertility within specific lineages have yet to be investigated, but in the absence of evidence to the contrary, it is assumed that the variation assayed is selectively neutral, and can therefore be interpreted in terms of population history.

The contrast between the clinal variation of Y -chromosomal lineages and the lack of clines in mtDNA data (Simoni et al., 2000a) is marked, although the latter is still a matter of debate (Simoni et al., 2000b; Torroni et al., 2000). It seems consistent with studies of global genetic diversity (Seielstad et al., 1998) which have ascribed such differences to patrilocality. However, direct evidence about mating practices in European prehistory is lacking - indeed, populations in some regions, such as Northern Iberia, may have practised matrilocality (Collins, 1986).

Clines for hg 1 and hg 9, encompassing 45\% of the chromosomes, and on a continental scale, show a similar pattern to that seen in the first principal component of classical gene frequency data, and also in the autocorrelation analysis of six Y-chromosomal microsatellites (Casalotti et al., 1999). A simplistic interpretation is that hg 9 chromosomes were carried in a major demographic expansion of agricultural migrants from the Near East, and that hg 1 chromosomes were a pre-existing predominant European lineage. Estimates of the ages of these lineages from coalescent analysis are not inconsistent with this scenario: the mutation defining hg 1 has been dated at approximately 23,000 YBP (Karafet et al., 1999), and that defining hg 9 at 14,800 $\pm 9,700$ YBP (Hammer et al., 2000).

Demic diffusion, and indeed any major directional gene-flow process, is generally expected to generate clines for only a fraction of the alleles at one locus (Sokal et al., 1989; Sokal et al., 1997). While two haplogroups show clines compatible with expansion from the Near East, three further lineages show different clinal patterns, indicating distinct population movements, southwards and westwards from north of the Black Sea (hg 3), from eastern Europe or Northern Asia westwards to the Baltic Sea (hg 16), and from south to north (hg 21). These clines are more regionally localised than those for hg 1 and hg 9 , pointing to phenomena affecting only part of the continent. It is tempting to assign known or surmised population-historical movements to these genetic gradients, but this should be done with caution.

The distribution of hg 3 chromosomes resembles the third principal component of variation of classical gene frequencies. There are several possible interpretations of this pattern: one (Cavalli-Sforza et al., 1994) is that it marks the Kurgan expansion from north of the Caspian Sea, dated to around 7,000 YBP. However, alternative explanations, such as the spread of pastoralism, or east to west movements of people such as the Scythians, Mongols and Huns, seem equally likely (Renfrew, 2000). Globally, hg 3 chromosomes are absent from Africa and the Americas, but their distribution is wide within Asia as well as Europe (Zerjal et al., 1999), consistent with their association with a recent and major expansion within Eurasia. An Asian origin has been proposed for hg 3 chromosomes because they are derived from hg 1 chromosomes (Figure 3.6), and therefore must have arisen in an area where hg 1 chromosomes were concentrated, which leads to a suggestion of either Europe or Asia. Additional microsatellite analysis of hg 1 chromosomes, within European populations showed low levels of diversity within European hg 1 samples, and hg 3 haplotypes clustered with Asian hg 1 haplotypes, suggesting an Asian origin (Zerjal et al., 1999).

Using 5 Y-chromosomal microsatellites (Kayser et al., 1997), the diversity within a panel of hg 3 chromosomes was determined. It was proposed the way in which the microsatellite haplotypes orientate within a median-joining network (Bandelt et al., 1999); which included a high frequency putative ancestral haplotype near the centre of the network, linked to multiple additional lower frequency haplotypes towards the outside of the network. This suggested a
relatively young haplogroup, because a significant proportion of the chromosomes have not yet had time to accumulate mutations at any of the loci examined (Zerjal et al., 1999). Microsatellite diversity analysis (Zerjal et al., 1999) used the mutation rate estimates of Heyer et al. (1997) to date the most recent common ancestor of a set of European and Asian hg 3 chromosomes to 3,800 YBP ( $95 \%$ confidence intervals: $1,600-13,000 \mathrm{YBP}$ ); the use of more recent mutation rate estimates (Kayser et al., 2000) would yield a date of 2,550 YBP (95\% CI: 1,650$4,260 \mathrm{YBP}$ ). Coalescent analysis has dated the $S R Y-1532$ mutation defining hg 3 to $\sim 7,500$ YBP (Karafet et al., 1999). If these dates are to be relied upon, they seem to suggest that the expansion of hg 3 chromosomes was due to later population movements than those of the Kurgan people.

Currently, dates cannot be attached to the clines, and the modern distributions of lineages are the outcome of many millennia of population movement. Assigning plausible dates to demographic movements is important, and here the Y chromosome can potentially contribute. Further, finer scale definitions of monophyletic lineages within Europe using new markers, and the analysis of these using microsatellites, offer the possibility of suggesting timescales for the major demographic events.

### 3.4.3. Language, geography and $Y$-chromosomal diversity

The Mantel tests demonstrate that patterns of Y-chromosomal genetic variation do not correlate as well with language as with geography. However, it should be borne in mind that geography and language together explain only $16.8 \%$ of the genetic variance (data not shown), and therefore other forces, such as founder effects and genetic drift, have also been important in determining current patterns of spatial variation. These findings seem at odds with those of Poloni et al. (1997), who showed that most of the population differentiation of $Y$ haplotypes was due to language. However, there are important differences between the two studies. The samples of Poloni et al. (1997) were global, rather than from a single continent, and showed a correspondingly greater linguistic and genetic diversity. The populations studied are located within a single continent, and most speak languages belonging to one language family, IndoEuropean; indeed, much of the genetic patterning observed may have its roots in the spread of that language family (Renfrew, 1987). The effect of increasing genetic, geographical and linguistic diversity in the input to the Mantel tests can be seen by including the North African samples (data not shown), which are both
geographically and linguistically distant from most other populations. This increases the partial correlations between genetics and geography, and between genetics and language, and also increases the significance of the latter to $p=0.024$, which, however, is still lower than the significance of the genetics-geography partial correlation ( $\mathrm{p}<0.001$ ).

The results of genetic barrier analysis (Figure 3.25) need to be interpreted with caution when, as in this case, sample distribution is uneven; the method is likely to be sensitive to the introduction of new populations, especially between existing sample sites which are far apart. In the next chapter, the robustness of these and other methods will be tested by the inclusion of an additional, and atypical population, the Ladin speakers. However, the analysis suggested that there is little correlation between genetic barriers and levels of linguistic separation, even when elite dominance is taken into account, by removing Hungarians and Turks from the analysis. One issue which has not been addressed, is the influence of the domination of Latin across Europe. This was widespread and may make deductions about relationships based purely on linguistic affinity difficult. Addressing this issue adequately would require a knowledge of the languages spoken prior to Romanisation, which we do not possess at the moment. While cultural factors other than language (such as politics and religion) might also be associated with genetic barriers, language was examined because it has the greatest time-depth. However, this is still likely to be less than the age of geographical barriers, the relative importance of which cannot easily be analysed. Twenty-five of 48 Delaunay connections crossed by genetic barriers also coincide with geographical barriers (taking a conservative definition, and considering only large stretches of water, and the two major mountain ranges, the Alps and the Caucasus), which seems to emphasise the greater importance of geographical factors in subdividing populations, resulting in large differences in Y-chromosomal haplogroup frequencies.

In summary, it seems that many kinds of barriers are probably recent, on an evolutionary time-scale (see Renfrew, 1987); after they were established, fluctuations of allele frequencies have become partly or largely independent in the populations separated by those barriers. Therefore, it is perhaps not surprising to find little correlation between the degree of language differentiation at a language boundary and the amount of genetic change observed across that boundary. As has been shown in the analysis of protein polymorphisms (Sokal et
al., 1990), linguistic differences tend to cause some degree of population subdivision, regardless of whether such differences are between language families, languages of the same family, or even dialects of the same language.

### 3.4.4. Conclusion

During this work the forces of geography and language have been separated, in reality they work together; spatially coincident weak geographical and linguistic barriers may together form strong barriers to gene flow. Some of the strongest genetic barriers observed, in central Europe, coincide with neither strong linguistic or geographical barriers. Linguistic and geographical heterogeneities and the effects of drift, on a background retaining a strong signal of expansion from the Near East and of other migrations, combine to shape the genetic landscape of Europe.

This analysis confirms the primacy of geography, rather than language, in shaping Y-chromosomal genetic diversity within Europe, but this may not be the case for mtDNA or autosomal loci.

# Chapter 4: A micro-geographic study examining the genetic diversity of linguistic isolates of the Italian Alps 

### 4.1. Introduction

Populations can be considered isolated for a variety of reasons; they can be genetically, geographically or culturally (linguistically and religiously) isolated. One particular Italian Alpine linguistic isolate, the 'Ladin speakers', will be discussed in detail within this chapter.

Languages are classified into groups known as language families and subfamilies, but there are exceptions and these are classified as linguistic isolates, unclassified, or pidgins and creoles (Ruhlen, 1991). Linguistic isolates and unclassified languages differ because a linguistic isolate is a language with a certain level of documentation that has been evaluated by linguists to determine if the language is closely related to any other language or group. A language referred to as unclassified is typically spoken by a newly discovered ethnic group, with little or nothing being known about the actual language. Pidgins and creoles are unusual languages; they refer to simplified languages that may be generated when people who speak different dialects wish to communicate. They can be a mixture of a number of different languages, and they are normally spoken as a second language, whereas creoles tend to be the native (first) languages (Ruhlen, 1991).

### 4.1.1. Linguistic isolates

There are a number of global linguistic isolates including Basque, Burushaski, Ket, Gilyak, Nahali, Sumerian, Etruscan, Hurrian, Meroitic, although the last four languages are now all extinct (Ruhlen, 1991).

The Basque language known as Euskera (Wardhaugh, 1987) has approximately 2 million speakers in the Basque homeland (Euzkadi), concentrated mainly in Spain, and 200,000 in France (Wardhaugh, 1987; Ruhlen, 1991) and has no obvious relationship to any other language (Ruhlen, 1991). In 1980, the Basque people were granted regional autonomy; but some sections seek more independence, the most infamous supporters being the Basque separatists, ETA (Ezukadi ta Askatasuna) which means 'Basque homeland and freedom' (Wardhaugh, 1987).

Threatened languages are referred to as minority languages, and the speakers as a linguistic minority, although some populations may consider the term 'minority' derogatory. Languages emerge and disappear as part of historical processes, such as migrations and military invasions when languages can be imposed upon others. Distinctions need to be made between languages that are classed as a minority in one country, such as French in Switzerland, but are major languages elsewhere, and languages that are a minority everywhere, such as Welsh, Scottish Gaelic and Sardinian (Wardhaugh, 1987). Minority languages have a number of characteristics, such as living in the shadow of a culturally dominant language, and this may be due to political, educational, social or religious factors. Speakers of the language may borrow from the dominant language, and are normally bilingual and may show a reluctance to pass it on to subsequent generations (Wardhaugh, 1987), which appears to perpetuate the status of the minority language until it may actually become extinct.

A new European charter, introduced in 1998, protects and promotes minority and regional languages throughout Europe, including non-territorial languages such as Yiddish and Romany. Certain legislation previously protected individuals and not languages: for example, the protection of human rights and fundamental freedoms states that any form of discrimination on the grounds of language or association with a national minority is prohibited. It has been suggested that approximately half of the world's 6000 languages and dialects are threatened with potential extinction, due to the dominance of certain languages such as English and French, and increased global communications that force people to conform to enable mass communication (Europe, 1998).

The charter stated that regional and minority languages should be permitted in education, media, administration, business and cultural life and a variety of European countries signed. A number of reasons were given in support of it, including that it may possibly help the stability of Europe. For example, the Good Friday agreement signed during the Northern Ireland Multiparty talks, acknowledged the need for understanding and tolerance of linguistic diversity. Both parties wanted equal recognition for their language, the Irish language spoken by the Nationalist/Catholic people, and Ulster-Scots spoken by the Unionist/Protestant people (Europe, 1998).

### 4.1.2. The Ladin-speaking population

The region of interest being studied here lies within the eastern Italian Alps (Southern Tyrol and Trentino) where three languages are spoken - German, Italian and Ladin (Stenico et al., 1996). These three languages belong to the IndoEuropean family, but German is from the Germanic sub-family, and both Italian and Ladin belong to the Romance sub-family (Ruhlen, 1991). The Ladin language is a linguistic isolate, and has only two known close linguistic relatives, Romansch which is spoken in eastern Switzerland, and Friulian which is spoken in north-eastern Italy (Stenico et al., 1998). The Ladin-speaking population mainly lives in the Dolomites in three Alpine valleys, Val Gardena, Val Badia, and Val di Fassa, but Ladin is also spoken in another location, Cortina d'Ampezzo (Figure 4.0), a town situated within a valley. Most Ladin speakers are bi- or trilingual, with German being the second most commonly spoken language (Stenico et al., 1998). The Ladin-speaking people are extremely proud of their heritage and culture (Figure 4.1).

Legal directives from the regional Trentino government allow the Ladin speakers certain linguistic rights, but in practice Ladin does not have linguistic equality with either Italian or German. However, the statute states that Ladins have the right to use their language for public official bodies within the Ladinspeaking valleys, and when dealing with regional offices, or in court (http://www.aber.ac.uk/~merwww/ladover.htm). Ladin is taught in both nursery school and primary school, and a recent survey suggested that $53.7 \%$ of Ladin speakers can write the language easily, and a further $23.1 \%$ can write it but with difficulty. The Ladin language is poorly represented within the media, but this is slowly improving and in 1975, the Ladin cultural centre was opened, whose function is to promote and safeguard the Ladin culture and language. It works with schools to develop teaching at the primary school level, and the teaching of Ladin to adults (http://www.istladin.net/presenta_en.htm).

Another linguistic isolate also resides in the Alps, the Mocheni, but they speak a dialect related to ancient Bavarian (German), inhabit a different valley east of Trento (Stenico et al., 1996) and will not be discussed in any detail within this chapter.


Figure 4.0. A map showing where the Ladin-speakers live in the north-eastern Italian Alps The black hatched area indicates land over 2000 m in height.

## Dolomites



Val Gardena


Figure 4.1. The Alpine landscape and a Ladin speaker in traditional dress
The Italian Alpine landscape including the Dolomites where the Ladin-speakers live, and a picture of a Ladin-speaker in traditional dress which indicates the desire to maintain links with the unusual Alpine history.

### 4.1.3. Mitochondrial DNA analysis

Much research employing mitochondrial DNA (mtDNA) has been undertaken within Ladin-speaking populations from these three valleys (Val Gardena, Badia and di Fassa), and quite unexpected results have been obtained. The hypervariable D-loop control region I was examined, and a high level of diversity has been observed within the Ladin groups and their geographical neighbours (Stenico et al., 1998). A sample of 20 Ladin speakers (from Val Gardena/ Wolkenstein and Colle Santa Lucia) differ from other Alpine groups by more than seven substitutions. No evidence was detected to support a close genetic relationship between the Ladins and their nearest linguistic relatives, the Romansch speakers of Switzerland (Stenico et al., 1996).

Mitochondrial DNA analysis of segment I of the hypervariable D-loop control region within 19 European populations led to the generation of a neighbour-joining tree (Figure 4.2), and two clear outliers were identified: the Ladins and Mocheni (Stenico et al., 1996). The closest association of the Ladin speakers with the Italian-speaking German sample has suggested they are probably the closest related of all 19 European populations examined (Stenico et al., 1996). Mitochondrial DNA variation within 43 Ladin speakers was summarised using principal component analysis, and these data comprising the first two principal components included $67 \%$ of the total observed genetic variance (Stenico et al., 1998). Genetic distances between all the samples were calculated, and negative distances were observed when within-sample diversity was high. Three Ladin sites (Val Gardena, Val Badia, and Val di Fassa) as well as the Near Eastern samples produced negative values for the first principal component. The level of diversity within the samples is also high, and this may suggest that the Ladin and Near Eastern samples are more similar than the Ladin and surrounding populations. These Ladin samples appear quite distinct from other European samples tested at the same time. Since small isolates are unlikely to maintain high genetic diversity, these results suggest that additional factors may have influenced the evolution of the Ladins, which may include different contributions from female migrations (Stenico et al., 1998).

### 4.1.3.1. Models proposed to explain the high level of genetic diversity within the Ladin speakers

Four models have been proposed from the mtDNA analysis to explain the level of genetic variation observed within Ladins (Stenico et al.,


Figure 4.2. Neighbour-joining tree produced from mtDNA sequence data A neighbour-joining (NJ) tree generated from genetic distances based on sequence data from segment I of the control region of mtDNA, within 19 European populations. The letter (F - French, I - Italian, G-German, R - Romansch) after the Swiss and Italian samples indicates the languages spoken. The Italian-G samples are GH (from Hefling) and GJ (from Jenesien), the Italian-I1 is IN (north of Trento) and IS (south of Trento), and the Italian-I2 sample is from Tuscany (From Stenico et al., 1996).
1996). Model A, 'local differentiation' suggests it is possible that the diversity has developed within this region, while the Ladins have remained isolated from their neighbours, and so drift will also have an influence. This is unlikely due to the historical regional climatic effects within the Alps, discussed later. Model B, a 'founder effect' proposes that the Ladins derive from small groups that may have included a proportion of the existing lineages, and this is a distinct possibility. Model C , 'ancient population sub-division' suggests that the Ladins may have originated from a older population that had already undergone a level of structuring, to produce the high degree of differentiation observed within the Ladins when compared to other Alpine groups. It is also possible that this ancestral population had very little influence on the remainder of the European maternal gene pool. Model D , 'selection' is considered unlikely because the control region of mtDNA is non-coding and it is unlikely to exhibit positive selection pressures with regards to adjacent sequences. If selection was active the level of diversity would be reduced, and this is not the case here (Stenico et al., 1996).

Five mitochondrial haplogroups are common in European populations (Richards et al., 1996), the oldest being haplogroup 2, with a most recent common ancestor (MRCA) living approximately 16,500 years ago. This haplogroup is also prevalent in Near Eastern populations and has also been observed at a frequency of less than $20 \%$ within European populations (Torroni et al., 1996), but at more than $60 \%$ within Ladinspeaking populations (Stenico et al., 1996). Two hypotheses have been proposed for how haplogroup 2 entered Western and Central Europe - a Palaeolithic spread of Homo sapiens, or a later spread of Neolithic farmers associated with the dispersal of agriculture (Stenico et al., 1998). This is supported by evidence that two expansions occurred both originating in the Near East (Richards et al., 1997) which contributed to the peopling of Europe (Richards et al., 1996).

A number of possibilities have been suggested to explain the high percentage of this rare European mtDNA haplogroup in the Ladins; haplogroup 2 was common in the founding Europeans and has been diluted out everywhere else due to drift. Another possibility was that haplogroup 2 was common in Neolithic farming communities which may
have contributed a significant proportion to the founding Ladin gene pool; or that haplogroup 2 has only recently entered the Alps and contained a substantial Middle Eastern component (Stenico et al., 1998). Two subgroups of haplogroup 2 have been identified (Sykes et al., 1996) and it has been suggested they appeared first in the Middle East approximately 12,000 years ago in Europe; this supports the proposed Neolithic spread of this haplogroup. An additional subgroup (Sykes et al., 1996) is thought to be greater than 30,000 years old and is found both in European and the Middle Eastern populations; this is taken as evidence to support the Palaeolithic spread of haplogroup 2.

Due to the climatic history of Europe including the development and retraction of expansive ice sheets, the Alps were uninhabitable before approximately 15,000 years ago, and archaeological evidence for human habitation dates back to only 12,000 YBP (Broglio, 1993). Because of the relatively low mutation rate of mtDNA, these dates exclude the possibility that the Ladin-speakers have generated such a high degree of mtDNA diversity in situ (model A), and so suggest that the Ladins may have originated from an ancestral gene pool not in the Alps, but perhaps in the Middle East (Stenico et al., 1998).

### 4.1.4. Linguistic comparisons

It has been proposed that Romance, Germanic, Baltic and Slavic languages may have originally been local dialects that spread and expanded by a variety of mechanisms, such as military conquests. It is thought that Romance languages spread mainly by military conquest by the Roman Empire, who imposed their social organisation, political control and the Latin language (Cavalli-Sforza et al., 1994).

A linguistic comparison between Ladin, Romansch and Friulian speaking populations all referred to as 'Alpine Romance' languages (Figure 4.3) has been undertaken (Forster et al., 1998b) and compares the relatedness of languages by using a network system, in which more related languages should cluster closer together. Similar network methods are routinely used in molecular biology to examine phylogenetic relationships of DNA sequences (Bandelt et al., 1995) and such like (see this Chapter and Chapter 5). A number of possible models have been proposed for how languages can evolve: the 'tree model' proposes that new


Figure 4.3. The position of the Ladin language within Indo-European subgroups The linguistic relationship between Ladin, Romansch and Friulian is indicated with respect to their overall position within the Italic sub-family and the Indo-European language family (Information taken from Ruhlen, 1991).
languages arise from an ancestral one over time; the 'wave model' suggests that languages evolve by borrowing words from other languages; and finally, languages can evolve independently by chance. Both the 'tree' and 'wave' models appear to be complementary (Forster et al., 1998b), because it is unlikely that languages evolve simply by one method alone.

The general procedure used during this statistical analysis of 'Alpine Romance' languages involved using a 100 word list (Swadesh, 1955) developed to include a basic vocabulary that nearly all languages contain. Commonly used words such as 'bird', 'ear', 'fish', 'nose' and 'water' are included in this list. An algorithm was developed that transformed the 100 word list into mathematical values depending upon how closely related they were. There were 60 identical words in terms of etyma (form of a word, usually the earliest form) and these were not included in the network generation. Reticulations observed within the network has been proposed to indicate the borrowing of words between languages, and extensive reticulations were observed.

Based on only 5 informative words ('dry', 'to lie', 'to sit', 'stone' and 'white'), the network separates into 2 distinct groups, one containing the Romansch speakers of Switzerland, and the other group the Ladin and Friulian speakers of Italy. These 2 clusters further sub-divide into western and eastern Romansch on the basis of 6 words ('liver', 'not', 'to eat', 'yellow', 'feather' and 'woman'), and this is claimed to be supported geographically by the division of the Graubünden Alps. The second cluster is divided, again by 6 words ('red', 'black', 'woman', 'man', 'to hear' and 'not'), into the northern and southern slopes of the Dolomite mountain range (Forster et al., 1998b). Even though this linguistic analysis produced some interesting results, its reliability, given the small number of informative words, seems questionable.

### 4.1.5. Questions to be addressed

Four models, local differentiation, founder effect, ancient population subdivision, and selection have been proposed to explain the high level of diversity observed within the mtDNA component of Ladin speakers. This work involves examining the Y-chromosomal diversity within Ladin speakers, and the testing of these four hypotheses, to try and determine if a similar pattern to that observed within the maternal component could be detected within the paternal lineages. If the patterns are not consistent, this may throw light on the different levels of
female and male differentiation (Seielstad et al., 1994; Ruiz Linares et al., 1996; Underhill et al., 1996; Underhill et al., 1997), and differing mating practices such as polygyny and patrilocality (Seielstad et al., 1998).

The initial part of the project employed biallelic analysis, examining ten markers routinely used during the larger European study. Once haplogroups had been determined for these samples, subsequent minisatellite (MSY1) analysis (Jobling et al., 1998a) was carried out. Additional analysis was undertaken by collaborators in Italy examining 5 Y -chromosomal microsatellite loci within these samples, and the internal diversity of microsatellites and the minisatellite within these linguistically isolated samples were determined. The Ladin speakers are an atypical European population, and the data generated here allows the robustness of previously used statistical methods employed in Chapter 3, for European populations as a whole, to be examined.

### 4.2. Materials and methods

### 4.2.1. Genomic DNA samples

The Ladin DNA samples were provided by Guido Barbujani, Michele Stenico, and Giulietta di Benedetto (University of Ferrara, Italy). They were to be assayed using both biallelic markers, and the minisatellite MSY1; this work has been carried out at the University of Leicester. The analysis of Y-chromosomal specific microsatellites was undertaken by Michele Stenico (Ferrara, Italy). Four Ladin-speaking populations were studied: Val Gardena (6 individuals), Val Badia (7), Val di Fassa (10), and Cortina d'Ampezzo (52). For their relative geographic locations see Figure 4.0.

### 4.2.2. Biallelic analysis

The complete method employed for analysing biallelic markers is described in detail in Chapter 3 (materials and methods). Ten commonly detected biallelic polymorphisms within Europe were used on all the Ladin DNA samples, which led to the determination of haplogroups.

### 4.2.3. MSY1 flanking DNA amplification via $P C R$

MSY1 flanking DNA amplification by PCR was carried out as described in detail by Jobling et al. (1998a), but briefly involves using two primers (Y1A+ and Y1B+) that flank the MSY1 minisatellite, on 50ng of genomic DNA. The PCR products were separated via agarose gel (1\%) electrophoresis, and the variant MSY1 band $(\sim 2-3 \mathrm{~kb})$ excised and incubated overnight at $4^{\circ} \mathrm{C}$ in $500 \mu \mathrm{l}$ of sterile distilled water. The buffer for the flanking PCR reaction was that supplied with the Thermus brockianus (Tbr) polymerase (NBL), and was supplemented with $200 \mu \mathrm{M}$ of each dNTP and $200 \mu \mathrm{~g} / \mathrm{ml}$ BSA (Boehringer). The Y1A+ and Y1B+ primers were used at 100 nM concentration and 0.02 U per reaction of Pfu DNA polymerase (Stratagene) was used in addition to the 0.66 U of Tbr DNA polymerase. The reaction volume was made up to $10 \mu \mathrm{l}$ with water.

### 4.2.4. MSY1 flanking PCR conditions

The following PCR conditions were employed in order to generate amplicons incorporating the MSY1 minisatellite.

Flanking PCR: $\quad$| $1-96^{\circ} \mathrm{C}$ for 40 seconds |
| :--- |
| $2-94^{\circ} \mathrm{C}$ for 8 seconds |
| $3-68^{\circ} \mathrm{C}$ for 1 minute $-0.5^{\circ} \mathrm{C}$ each cycle |
| $4-68^{\circ} \mathrm{C}$ for 3 minutes |
| $5-\mathrm{Go}$ to $2(\times 10$ times) |
| $6-94^{\circ} \mathrm{C}$ for 8 seconds |
| $7-63^{\circ} \mathrm{C}$ for 1 min |
| $8-68^{\circ} \mathrm{C}$ for 3 minutes +4 seconds each cycle |
| $9-\mathrm{Go}$ to step $6(\times 25$ times) |
| $10-5^{\circ} \mathrm{C}$ for ever |

Primer sequences ( $5^{\prime}$ to $3^{\prime}$ ):
Y1A+ ACA GAG GTA GAT GCT GAA GCG GTA TAG C
Y1B+ GCA ACT CAA GCT AGG ACA AAG GGA AAG G

### 4.2.5. Minisatellite variant repeat mapping by $P C R$

Five variant MSY1 repeats were previously identified by sequencing (Jobling et al., 1998a), and the three most commonly detected can be seen in Table 4.0 below.

Table 4.0. Three MSY1 repeat types


### 4.2.6. Radioactive end labelling of PCR primers ( $\left.{ }^{33} P \gamma-A T P\right)$

Four variant-repeat-specific primers were end-labelled using [ $\left.{ }^{33} \mathrm{P}\right] \gamma$-ATP (NEN Dupont) in the following reaction mix to provide sufficient primer for 24 MVR-PCR amplifications in a 96-well plate ( 24 reactions $\times 4$ primers $=96$ wells).

| ${ }^{33} \mathrm{P} \gamma$-ATP | $2.8 \mu \mathrm{l}$ |
| :--- | :--- |
| sterile distilled water | $15.4 \mu \mathrm{l}$ |

This mixture was incubated at $37^{\circ} \mathrm{C}$ for a minimum of 30 minutes, and the unincorporated ATP was not removed prior to PCR.

### 4.2.7. Three-state Minisatellite Variant Repeat PCR (MVR-PCR)

A flanking PCR was performed as described above (Jobling et al., 1998a). The gel slice (containing the MSY1 flanking amplicon) was left overnight at $4^{\circ} \mathrm{C}$ in $500 \mu \mathrm{l}$ of water. A $2 \mu \mathrm{l}$ aliquot of this eluant was used as template in each of four $10 \mu \mathrm{l}$ MVR-PCR reactions which contain one of the flanking primers and one ${ }^{33}$ P-end-labelled (see above protocol) repeat-specific discriminator primer that recognises only one of the three repeat types (see Table 4.1, and Figure 4.4).

Table 4.1. The combinations of primers used to completely map the MSY1 repeat types.

| Primers used | Repeat type detected | Direction |
| :--- | :---: | :--- |
| Y1A+ and TAG1D | 1 | Forward |
| Y1A+ and TAG3C | 3 | Forward |
| Y1B+ and TAG3R3 | 3 | Reverse |
| Y1B+ and TAG4R3 | 4 | Reverse |

The buffer for the MVR-PCR was also that supplied with the Tbr polymerase and was again supplemented with $200 \mu \mathrm{M}$ of each dNTP and $200 \mu \mathrm{~g} / \mathrm{ml} \mathrm{BSA}$ (Boehringer). 0.5 U of Tbr DNA polymerase alone was used per reaction. The flanking primer was used at a concentration of 100 nM and the labelled discriminator primers at concentrations of 100nM (TAG1D and TAG3C), 50 nM (TAG3R3) and 200nM (TAG4R3).

### 4.2.8. MVR-PCR conditions

The PCR conditions employed to generate MVR-PCR amplicons from excised flanking PCR reactions are listed below.

MVR-PCR: $\quad 1-96^{\circ} \mathrm{C}$ for 40 seconds
$2-94^{\circ} \mathrm{C}$ for 8 seconds
$3-62^{\circ} \mathrm{C}$ for 1 min


Figure 4.4. A schematic diagram of the MVR-PCR system on MSY1 (A) shows an allele, containing type 1,3, and 4 repeats which are coded in forward directions (type1 and 3), and reverse directions (type 3 and 4). (B) shows the predicted hairpin structure of MSY1 repeats, and $S$ indicates either a $G$ or $C$ base. (C) five variant MSY1 repeats, previously identified by sequencing.

4-68 ${ }^{\circ} \mathrm{C}$ for 3 minutes
5 - Go to step 2 ( $\times 2$ times)
$6-94^{\circ} \mathrm{C}$ for 8 seconds
$7-68^{\circ} \mathrm{C}$ for 30 seconds +4 seconds each cycle
$8-68^{\circ} \mathrm{C}$ for 3 minutes 30 seconds
$9-$ Go to step 6 (x20 times)
$10-5^{\circ} \mathrm{C}$ for ever

Primer sequences ( $5^{\prime}$ to $3^{\prime}$ ):
TAG1D tea tge gec cat ggt cog gat gtg tat ant ata cat cat gta tat tg
TAG3C tea tge gle cat ggt ccg gat gtg tat ant ata cat gat gta tat tg
TAG3R3 tca tge gtc cat ggt ccg gac atc atg tat att ata cac ant ata cat c TAG4R3 tea tge gtc cat ggt ccg gac atc atg tat att ata cat ant ata cat c

Lower case bases are the 'tag' sequence (Jeffreys et al., 1991) used to decouple discrimination from amplification (Jobling et al., 1998a).

Both the flanking and three step MVR-PCR reactions were carried out using Thermowell M plates (Advanced Biotechnologies, Surrey, England) in an MJR PTC-200 thermal cycler. When DNA quantities were limiting or quality poor, only $5-10 \mathrm{ng}$ of template were added to the flanking reaction and step 9 was extended by two cycles.

### 4.2.9. Polyacrylamide gel preparation

The structure of these minisatellites is visualised via $2.5 \%$ denaturing polyacrylamide gel (Ultrapure Sequagel sequencing system, National Diagnostics) electrophoresis. The polyacrylamide gels are prepared as a separate $6 \%$ plug at the bottom, and the remainder of the gel is $2.5 \%$. The $6 \%$ plug stops the short amplicons running off the bottom of the gel. The gel was allowed to set for between 2 hours and over night.

| 6\% plug | concentrate | 1.2 ml |
| :--- | :--- | :--- |
| buffer | 0.5 ml |  |
|  | diluent | 3.3 ml |
|  | APS $(10 \% \mathrm{w} / \mathrm{v})$ | $40 \mu \mathrm{l}$ |
|  | Temed | $6 \mu \mathrm{l}$ |


| $2.5 \%$ gel | concentrate | 10 ml |
| :--- | :--- | :--- |
| buffer | 10 ml |  |
|  | diluent | 80 ml |
|  | APS $(10 \% \mathrm{w} / \mathrm{v})$ | $600 \mu \mathrm{l}$ |
|  | Temed | $90 \mu \mathrm{l}$ |

### 4.2.10. Polyacrylamide gel running conditions

Standard 10X TBE is upgraded to sequencing grade 10X TBE, by the addition of 1 g Tris (Trizma Base) and 2 g Boric acid per 200 ml of 10X TBE. The sequencing grade buffer is used during the electrophoresis procedure. This buffer is heated in a microwave, and then added to the electrophoresis apparatus, before the pre-heating gel stage. The polyacrylamide gels are pre-heated at 85 mA for approximately 1 hour, or until the temperature of the gel reaches approximately $55^{\circ} \mathrm{C}$.

Post three-state MVR-PCR amplification $12 \mu \mathrm{l}$ of stop solution were added to each well. These samples were heat denatured at $94^{\circ} \mathrm{C}$ for 4 minutes in a PCR machine. A $6.5 \mu \mathrm{l}$ volume of each reaction was run out on a $50 \mathrm{~cm} 2.5 \%$ denaturing polyacrylamide gel (Sequagel, National Diagnostics). The gel was electrophoresed at 55 mA for approximately 2 hours, until the xylene-cyanol reached 29 cm from the top of the glass plate.

The gels were fixed with a solution of $10 \% ~(v / v)$ acetic acid and $12 \% ~(v / v)$ methanol (Sigma), and then dried for approximately 1 to 2 hours on a gel dryer (BioRad) linked to a vacuum pump. The dried gels were subsequently exposed to X-ray film (Fuji or Kodak) in a light-proof cassette for 4 hours to 7 days at room temperature. An example of a MSY1 code can be seen in Figure 4.5. These maps can be graphically represented, as shown in Figure 4.6.


Figure 4.5. An MSY1 MVR code
Example of a $2.5 \%$ polyacrylamide gel of MVR-PCR products displaying an entire MSY1 code: (3)5(1)8(3)2(1)2(3)26(4)4(0)4(4)12.

| A | Co10 | 134 |
| :--- | :--- | :--- |
|  | Co25 | 1343434 |
|  | Vf10 | 313404 |

B Co10 (1)15.(3)41.(4)18

Co25 (1)16.(3)39.(4)1.(3)3.(4)2.(3)2.(4)12

Vf10 (3)5.(1)7.(3)31.(4)3.(0)4.(4)19

## C Co10 1111111111111113333333333333333333333333333333333333444444444444444444 <br> Co25 111111111111111133333333333333333333333333333333333333343334433444444444444 <br> Vf10 333331111111333333333333333333333333333333344400004444444444444444444



Figure 4.6. Four different ways to represent MSY1 data
(A) shows the general order of the repeat types known as the modular structure. (B) shows the repeat type in brackest followed by the actual number of repeats. (C) shows the full structure in numbers only. (D) is a schematic representation of the structure, with each circle being equivalent to one repeat unit, and each repeat type having a different colour, type 1 (red), type 3 (yellow), type 4 (blue) and null repeats (green).

### 4.3. Results

### 4.3.1. Biallelic analysis

Preliminary analysis employing biallelic markers on a total of 72 Ladin chromosomes allowed haplogroups to be determined. These haplogroup frequency data from the four locations were compared with each other, using the Arlequin 2.00 program (Schneider et al., 2000) Fst analysis was undertaken (Table 4.2).

Table 4.2. Fst values between all four Ladin speaking populations.

|  | Val Gardena | Val Badia | Val di Fassa | Cortina |
| :--- | :--- | :--- | :--- | :--- |
| Val Gardena | 0.000 |  |  |  |
| Val Badia | $-0.15966(p=0.991)$ | 0.000 |  |  |
| Val di Fassa | $-0.07391(p=0.802)$ | $-0.03834(p=0.578)$ | 0.000 |  |
| Cortina | $-0.09545(p=0.838)$ | $-0.06263(p=0.703)$ | $0.01820(p=0.261)$ | 0.000 |

The name Cortina d'Ampezzo was abbreviated to Cortina to fit into the above table. All of the above values are non-significant, as indicated by the $p$ values given in brackets.

The non-significant values could be due to the small sample sizes analysed, and because of this all the Ladin samples were pooled, with the summed haplogroup data being, hgs 1 (61\%), 2 (13\%), 21 (19\%), and 9 (6\%). All of these four haplogroups are commonly detected within European populations. These Ladin data were compared to data from 7 geographically close populations, taken from Chapter 3 (see Table 4.3).

Table 4.3. Haplogroup frequency data for all four Ladin sample sites and for seven geographically close European populations.

| Pop | n | hg 1 | hg 2 | hg 3 | hg 8 | hg 9 | hg 12 | hg 16 | hg 21 | hg 22 | hg 26 | h | SE |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| VG | 6 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | nc | nc |
| VB | 5 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | nc | nc |
| VF | 9 | 6 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | nc | nc |
| Cor | 49 | 30 | 7 | 0 | 0 | 1 | 0 | 0 | 11 | 0 | 0 | nc | nc |
| Ladin | 69 | 44 | 9 | 0 | 0 | 3 | 0 | 0 | 13 | 0 | 0 | 0.539 | $\pm 0.0029$ |
| Fren | 40 | 20 | 10 | 2 | 1 | 2 | 0 | 0 | 3 | 2 | 0 | 0.674 | $\pm 0.0040$ |
| Bav | 80 | 38 | 18 | 12 | 0 | 4 | 0 | 0 | 6 | 2 | 0 | 0.693 | $\pm 0.0104$ |
| Ital | 99 | 44 | 14 | 2 | 0 | 20 | 0 | 0 | 13 | 0 | 6 | 0.722 | $\pm 0.0014$ |
| Greek | 36 | 4 | 8 | 3 | 0 | 10 | 0 | 0 | 10 | 0 | 1 | 0.776 | $\pm 0.0031$ |
| Yugo | 100 | 11 | 49 | 16 | 0 | 8 | 2 | 0 | 13 | 0 | 1 | 0.698 | $\pm 0.0015$ |
| Slov | 70 | 15 | 19 | 26 | 0 | 4 | 0 | 0 | 5 | 1 | 0 | 0.734 | $\pm 0.0019$ |
| Hung | 36 | 11 | 10 | 8 | 0 | 1 | 0 | 0 | 6 | 0 | 0 | 0.751 | $\pm 0.0035$ |

Abbreviations within the table: pop - population, n - total number of samples, nc - not calculated. Population abbreviations include: VG - Val Gardena, VB - Val Badia, VF - Val di Fassa, Cor - Cortina d'Ampezzo, Fren - French, Bav -Bavarian, Ital -Italian, Yugo -Yugoslavian, Slov - Slovenian, Hung - Hungarian. The Ladin samples are pooled data for comparison purposes. h - Nei's estimator of diversity, and SE - standard error values for h .

Nei's unbiased estimator of diversity was calculated for all eight populations using the following equation $\left[h=1-\Sigma\left(x^{2}\right)\right]$. How the value for diversity is generated is given in an example below.

## Ladin sample, $\mathbf{n = 6 9}$

$$
\begin{gathered}
{[44 / 69]^{2}+[9 / 69]^{2}+[3 / 69]^{2}+[13 / 69]^{2}=(0.4066+0.0170+0.0019+0.0355)=0.461} \\
1-0.461=0.539, \text { so } \mathbf{h}=0.539
\end{gathered}
$$

The standard errors were also calculated because the samples sizes were different; the equation used was as follows: $1 / n(2 n-1) x\left[\Sigma x^{2}-\left(\sum x^{2}\right)^{2}+4(n-1)\left\{\left(x^{3}-\right.\right.\right.$ $\left.\left.\left(\Sigma x^{2}\right)^{2}\right\}\right]$ (Nei, 1978). Where $n=$ sample size, and $x^{2}$ was as before. An example is shown below:

## Ladin sample, $\mathbf{n = 6 9}$

$$
\begin{gathered}
1 / 69(138-1) \times\left[0.461-(0.461)^{2}+4(69-1)\left\{0.313-(0.461)^{2}\right\}\right]=(1 / 9453) \times[0.248+272(0.1)] \\
=1 / 9453(27.448), \text { so } S E= \pm 0.0029
\end{gathered}
$$

These haplogroup data were transformed into a graphical representation superimposed onto the genetic landscape (Figure 4.7). A visual comparison appears to indicate that there are differences between all populations examined, for example the Ladin-speakers are less diverse with only four haplogroups being observed. The frequency of hg 1 chromosomes is much higher within the Ladins than any of the other populations examined. This can be tested statistically by undertaking a pairwise Fst comparison of eight populations (see Table 4.4). The Ladin speakers are significantly different from all populations with the only exception being the French.

### 4.3.2. Examination of haplotype diversity

The degree of genetic diversity (h) can be examined statistically using Nei's equation (see section 4.3.1). Are the Ladin-speakers as diverse when examining Y-chromosomal data, as other European populations? The observation from mtDNA was that they were more diverse (Stenico et al., 1998). The values generated can be seen in Table 4.3, and they suggest which of the four proposed models (local differentiation, founder effect, ancient population sub-


Figure 4.7. Comparison of biallelic haplogroup frequency data between Ladins and neighbouring populations A comparison of the Y -chromosomal biallelic data for 69 Ladin speakers and seven European populations, chosen for their close geographical proximity. The haplotypes analysed are highlighted in yellow in the small tree at the bottom of this figure

Table 4.4. Conventional pairwise Fst values generated from haplogroup frequency data for the Ladin speakers and seven surrounding populations.

|  | Greek | Yugo | Slov | Hung | Ladin | Ital | Bav |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Greek | 0 |  |  |  |  |  |  |  |
| Yugo | $0.07109(\mathrm{p}=0.090)$ | 0 |  |  |  |  |  |  |
| Slov | $0.0929(\mathrm{p}=0.000)$ | $0.05901(\mathrm{p}=0.000)$ | 0 |  |  |  |  |  |
| Hung | $0.05539(\mathrm{p}=0.027)$ | $0.04264(\mathrm{p}=0.018)$ | $0.00596(\mathrm{p}=0.252)$ | 0 |  |  |  |  |
| Ladin | $0.18521(\mathrm{p}=0.000)$ | $0.23448(\mathrm{p}=0.000)$ | $0.19207(\mathrm{p}=0.000)$ | $0.09895(\mathrm{p}=0.000)$ | 0 |  |  |  |
| Ital | $0.07535(\mathrm{p}=0.000)$ | $0.15175(\mathrm{p}=0.000)$ | $0.12180(\mathrm{p}=0.000)$ | $0.05492(\mathrm{p}=0.000)$ | $0.02930(\mathrm{p}=0.000)$ | 0 |  |  |
| Bav | $0.12129(\mathrm{p}=0.000)$ | $0.12048(\mathrm{p}=0.000)$ | $0.06482(\mathrm{p}=0.000)$ | $0.01168(\mathrm{p}=0.162)$ | $0.03684(\mathrm{p}=0.000)$ | $0.02663(\mathrm{p}=0.027)$ | 0 |  |
| Fren | $0.12444(\mathrm{p}=0.000)$ | $0.12740(\mathrm{p}=0.000)$ | $0.09873(\mathrm{p}=0.000)$ | $0.02780(\mathrm{p}=0.126)$ | $0.01909(\mathrm{p}=0.153)$ | $0.01623(\mathrm{p}=0.135)$ | $-0.00989(\mathrm{p}=0.757)$ | 0 |

The $p$ values are given in a bracket after each genetic distance value and the overall significance level is $95 \%$. The values that are not significantly different are indicated in yellow.
division, and selection) may be correct. The Ladin populations have the lowest value for $h(0.539)$, and so have the lowest level of diversity within the populations examined. The highest observed diversity is detected within the Greek population ( $\mathrm{h}=0.776$ ). The detection of four diverged biallelic haplogroups (hgs 1, 2, 9 and 21) suggest it is unlikely to be model A - local differentiation, and selection can be assumed to be absent (model D). However, the lack of diversity observed within the Ladin samples may be due to model B (founder effect) or model C (ancient population sub-division). This issue can be examined further using multiallelic markers, such as microsatellites and MSY1 (see sections 4.3.4 onwards).

### 4.3.3. Statistical analysis of the biallelic data

A number of statistical analyses carried out in Chapter 3, and can be used within this chapter to determine how similar or different the results are with the inclusion of the Ladin data. The statistical analyses include principal components analysis (PCA), and Fst analysis (not undertaken in Chapter 3). The robustness of other tests such as spatial autocorrelation (SAAP) and genetic barrier analysis, can be examined by the addition of new (Ladin biallelic) data - to determine if the clines alter in the SAAP analysis, and to see if the number of barriers and their structure is affected. Fst analysis has been previous described but the other analyses will be briefly outlined below.

### 4.3.3.1. Principal components analysis

Principal components analysis (PCA) was again repeated on the complete European data set but with the inclusion of the Ladin data (Figure 4.8). The major division is again between the North African and all the other populations. The Ladin samples do not appear as obvious outliers, which might have suggested genetic isolation, but they do not cluster with their closest geographical neighbours either. They appear to cluster most closely with both North and South Portuguese samples (Figure 4.8A), but this may reflect the high frequency of hg 21 chromosomes when compared with neighbouring populations. The Ladin speakers share no particular linguistic similarity with the Portuguese, although they both belong to the Romance section of the Italic group of Indo-European.


Figure 4.8. Two examples of scatterplots generated from principal component analysis data (A) encompasses 48 populations, and (B) - includes 47 populations (excluding Ladins). Both analyses exclude the Greenlandic sample, but include both North African populations. The linguistic affiliations are indicated in the key for all languages spoken by the different populations. The abbreviated population names correspond to Table 3.3.1 within Chapter 3.

### 4.3.3.2. Spatial autocorrelation analysis

As with the previous Chapter, SAAP analysis (Sokal and Oden, 1978) was undertaken, to ask whether the inclusion of the Ladins changes the clinal patterns observed previously. For a comparison of the European data set ( 47 populations, without the Ladin samples) and the complete data set with the Ladins, see Figure 4.9.

Overall patterns are unchanged when the Ladins are included, but the significance ( $p$-value) of certain points alters (from 0.001 to 0.003 , for example in hg 21). Within hg 1 the cline appears to be more continentwide without the inclusion of the Ladins, since their high frequency of hg 1 is atypical of central or southern Europe.

### 4.3.3.3. Genetic barrier analysis

Barrier analysis (Hurles, 1999) was carried out using the European biallelic haplogroup data, but with the inclusion of the Ladin-speaking populations, to see if the overall number and distribution of genetic barriers varied over the European landscape. A total of 48 European populations were examined, and the barriers detected were superimposed onto Delaunay triangles on a map of Europe (Figure 4.10).

An increased number of barriers are detected when the Ladin samples are included, and this seems to support the Fst analysis; they are different to their close geographic neighbours. The genetic barriers also seem to correspond well with local geographic barriers, such as the Alps, and this isolation may be an important factor, with a strong influence on their paternal genetic component. The correlation between genetic and geographic barriers can be visualised by projecting the genetic barriers onto Delaunay triangles that have been superimposed onto a European landscape (Figure 4.10). A change in the other genetic barriers with the inclusion of the Ladin samples is also observed, for example the barrier around the Polish increases in size, and this confirms that this analytical approach is not very robust to the inclusion of additional data.

## With Ladins



Without Ladins


Figure 4.9. Two sets of correlograms for the Y-chromosomal European haplogroup data
(A) includes 48 populations, whereas (B) 47 populations (excluding the Ladin samples). The significance of each point is indicated by the different coloured circles, ranging from $p<0.01$, to $p<0.05$, to non-significant values. Five of the 6 haplogroups are clinal, the exception being haplogroup 2.


Figure 4.10. The inclusion of Ladin speakers and how this influences the location of genetic barriers on the European landscape.
Delaunay triangles have been positioned onto the European landscape, and genetic barriers (thick black lines) detected from haplogroup data have been superimposed onto the maps. A - comprises of 46 populations, and $B$ - comprises 45 populations and excludes the Ladin samples.

Table 4.5. Complete microsatellite data for the Ladin-speaking populations

| Sample | Valley | Hg | DYS390 | DYS391 | DYS392 | DYS393 | DYS19 | Haplotype | Hap Frequency |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Co51 | Cortina | 1 | 24 | 11 | 13 | 13 | 14 | Hap 1 | 9 |
| Co52 | Cortina | 1 | 24 | 11 | 13 | 13 | 14 | Hap 1 |  |
| Co25 | Cortina | 1 | 24 | 11 | 13 | 13 | 14 | Hap 1 |  |
| Vf1 | Val di Fassa | 1 | 24 | 11 | 13 | 13 | 14 | Hap 1 |  |
| Vf5 | Val di Fassa | 1 | 24 | 11 | 13 | 13 | 14 | Hap 1 |  |
| Vf6 | Val di Fassa | 1 | 24 | 11 | 13 | 13 | 14 | Hap 1 |  |
| Vf4 | Val di Fassa | 1 | 24 | 11 | 13 | 13 | 14 | Hap 1 |  |
| Co16 | Cortina | 1 | 24 | 11 | 13 | 13 | 14 | Hap 1 |  |
| Co42 | Cortina | 1 | 24 | 11 | 13 | 13 | 14 | Hap 1 |  |
| Co53 | Cortina | 1 | 25 | 11 | 13 | 12 | 14 | Hap 2 | 3 |
| Co48 | Cortina | 1 | 25 | 11 | 13 | 12 | 14 | Hap 2 |  |
| Vg4 | Val Gardena | 1 | 24 | 10 | 13 | 13 | 14 | Hap 2 |  |
| Co11 | Cortina | 1 | 25 | 11 | 13 | 14 | 14 | Hap 3 | 1 |
| Co19 | Cortina | 1 | 23 | 10 | 13 | 13 | 14 | Hap 4 | 11 |
| Co22 | Cortina | 1 | 23 | 10 | 13 | 13 | 14 | Hap 4 |  |
| Co24 | Cortina | 1 | 23 | 10 | 13 | 13 | 14 | Hap 4 |  |
| Co36 | Cortina | 1 | 23 | 10 | 13 | 13 | 14 | Hap 4 |  |
| Co40 | Cortina | 1 | 23 | 10 | 13 | 13 | 14 | Hap 4 |  |
| Co17 | Cortina | 1 | 23 | 10 | 13 | 13 | 14 | Hap 4 |  |
| Co47 | Cortina | 1 | 23 | 10 | 13 | 13 | 14 | Hap 4 |  |
| Co30 | Cortina | 1 | 23 | 10 | 13 | 13 | 14 | Hap 4 |  |
| Co35 | Cortina | 1 | 23 | 10 | 13 | 13 | 14 | Hap 4 |  |
| Co49 | Cortina | 1 | 23 | 10 | 13 | 13 | 14 | Hap 4 |  |
| Co55 | Cortina | 1 | 23 | 10 | 13 | 13 | 14 | Hap 4 |  |
| Co10 | Cortina | 1 | 23 | 11 | 13 | 13 | 14 | Hap 5 | 4 |
| Co43 | Cortina | 1 | 23 | 11 | 13 | 13 | 14 | Hap 5 |  |
| Co21 | Cortina | 1 | 23 | 11 | 13 | 13 | 14 | Hap 5 |  |
| Co13 | Cortina | 1 | 23 | 11 | 13 | 13 | 14 | Hap 5 |  |
| Co14 | Cortina | 1 | 24 | 11 | 13 | 14 | 14 | Hap 6 | 1 |
| Vg1 | Val Gardena | 1 | 24 | 11 | 13 | 12 | 14 | Hap 7 | 2 |
| Vg5 | Val Gardena | 1 | 24 | 11 | 13 | 12 | 14 | Hap 7 |  |
| Co18 | Cortina | 1 | 25 | 10 | 13 | 13 | 14 | Hap 8 | 1 |
| Co20 | Cortina | 1 | 25 | 11 | 13 | 13 | 14 | Hap 9 | 4 |
| Co29 | Cortina | 1 | 25 | 11 | 13 | 13 | 14 | Hap 9 |  |
| Co32 | Cortina | 1 | 25 | 11 | 13 | 13 | 14 | Hap 9 |  |
| Co41 | Cortina | 1 | 25 | 11 | 13 | 13 | 14 | Hap 9 |  |
| Co39 | Cortina | 1 | 25 | 12 | 13 | 13 | 14 | Hap 10 | 1 |
| Vg6 | Val Gardena | 1 | 22 | 11 | 13 | 13 | 15 | Hap 11 | 1 |
| Vb3 | Val Badia | 1 | 24 | 10 | 13 | 13 | 14 | Hap 12 | 1 |
| Co34 | Cortina | 1 | 23 | 12 | 13 | 13 | 14 | Hap 13 | 2 |
| Co45 | Cortina | 1 | 23 | 12 | 13 | 13 | 14 | Hap 13 |  |
| Vf9 | Val di Fassa | 1 | 24 | 10 | 11 | 12 | 16 | Hap 14 | 1 |
| Vf7 | Val di Fassa | 1 | 24 | 10 | 11 | 13 | 15 | Hap 15 | 1 |
| $\mathrm{Co5}$ | Cortina | 2 | 23 | 10 | 11 | 12 | 13 | Hap 1 | 1 |
| Co44 | Cortina | 2 | 23 | 10 | 11 | 12 | 15 | Hap 2 | 1 |
| Co54 | Cortina | 2 | 23 | 10 | 11 | 13 | 14 | Hap 3 | 1 |
| Co27 | Cortina | 2 | 23 | 10 | 11 | 14 | 14 | Hap 4 | 1 |
| Co50 | Cortina | 2 | 22 | 10 | 11 | 14 | 15 | Hap 5 | 1 |
| Vf3 | Val di Fassa | 2 | 22 | 10 | 14 | 13 | 14 | Hap 6 | 1 |
| Co26 | Cortina | 2 | 22 | 11 | 11 | 13 | 14 | Hap 7 | 1 |

Table 4.5. Complete microsatellite data for the Ladin-speaking populations

| Sample | Valley | Hg | DYS390 | DYS391 | DYS392 | DYS393 | DYS19 | Haplotype | Hap Frequency |
| :---: | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Vg7 | Val Gardena | 2 | 22 | 10 | 11 | 15 | 15 | Hap 8 | 1 |
| Vf8 | Val di Fassa | 9 | 22 | 10 | 14 | 13 | 14 | Hap 1 | 1 |
| Vf2 | Val di Fassa | 9 | 23 | 10 | 11 | 12 | 14 | Hap 2 | 1 |
| Co3 | Cortina | 9 | 24 | 10 | 11 | 12 | 16 | Hap 3 | 1 |
| Vf10 | Val di Fassa | 9 | 24 | 11 | 13 | 13 | 14 | Hap 4 | 1 |
| Co23 | Cortina | 21 | 24 | 10 | 11 | 13 | 13 | Hap 1 | 9 |
| Co28 | Cortina | 21 | 24 | 10 | 11 | 13 | 13 | Hap 1 |  |
| Co37 | Cortina | 21 | 24 | 10 | 11 | 13 | 13 | Hap 1 |  |
| Co38 | Cortina | 21 | 24 | 10 | 11 | 13 | 13 | Hap 1 |  |
| Co6 | Cortina | 21 | 24 | 10 | 11 | 13 | 13 | Hap 1 |  |
| Co33 | Cortina | 21 | 24 | 10 | 11 | 13 | 13 | Hap 1 |  |
| Co7 | Cortina | 21 | 24 | 10 | 11 | 13 | 13 | Hap 1 |  |
| Co9 | Cortina | 21 | 24 | 10 | 11 | 13 | 13 | Hap 1 |  |
| Co2 | Cortina | 21 | 24 | 10 | 11 | 13 | 13 | Hap 1 |  |
| Co4 | Cortina | 21 | 24 | 10 | 11 | 15 | 13 | Hap 2 | 1 |
| Vg3 | Val Gardena | 21 | 25 | 10 | 11 | 13 | 13 | Hap 3 | 1 |



MSY1 modular structure

| 1,3,4 |
| :---: |
| 1,3,4,3,4,3,4 |
| 3,1,3,4 |
| 1,3,4,3,4 |

## Alpine population

 Cortina忩気 Val Gardena Val Badia Vo: Val di Fassa- unobserved haplotype$n=1$

Figure 4.11. Median-joining network of hg 1 chromosomes including all four Ladin valleys
The Y-chromosomal microsatellite and minisatellite data for all four Ladin valleys, constructed into a median joining network for all haplogroup 1 chromosomes. The size of the circle is proportional to the frequency of the microsatellite haplotype detected. Each Alpine valley has been allocated a different pattern, and the different MSY1 modular structures are given different colours.

Table 4.6. Microsatellite haplotype data from 5 loci on 86 (82) chromosomes including the Ladin-speakers, other European and Asian samples

| Name |  | DYS19 | $\mathbf{3 9 0}$ | $\mathbf{3 9 1}$ | $\mathbf{3 9 2}$ | $\mathbf{3 9 3}$ | Freq. | Hap |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| Vb3 | Ladin | 14 | 24 | 10 | 13 | 13 | 6 | 1 |
| Vb8 | Ladin | 14 | 24 | 11 | 13 | 13 | 16 | 2 |
| Vf7 | Ladin | 15 | 24 | 10 | 11 | 13 | 1 | 3 |
| Vf9 | Ladin | 16 | 24 | 10 | 11 | 12 | 1 | 4 |
| Vg1 | Ladin | 14 | 24 | 11 | 13 | 12 | 2 | 5 |
| Vg2 | Ladin | 14 | 24 | 10 | 13 | 12 | 1 | 6 |
| Vg6 | Ladin | 15 | 22 | 11 | 13 | 13 | 1 | 7 |
| Co40 | Ladin | 14 | 23 | 10 | 13 | 13 | 13 | 8 |
| Co1 | Ladin | 14 | 23 | 10 | 13 | 14 | 1 | 9 |
| Co10 | Ladin | 14 | 23 | 11 | 13 | 13 | 6 | 10 |
| Co11 | Ladin | 14 | 25 | 11 | 13 | 14 | 2 | 11 |
| Co20 | Ladin | 14 | 25 | 11 | 13 | 13 | 4 | 12 |
| Co34 | Ladin | 14 | 23 | 12 | 13 | 13 | 2 | 13 |
| Co39 | Ladin | 14 | 25 | 12 | 13 | 13 | 1 | 14 |
| Co48 | Ladin | 14 | 25 | 11 | 13 | 12 | 1 | 15 |
| m419 | Indian | 14 | 23 | 10 | 10 | 14 | 4 | 16 |
| m456 | Indian | 15 | 23 | 10 | 10 | 14 | 1 | 17 |
| m464 | Indian | 13 | 23 | 10 | 10 | 14 | 2 | 18 |
| m467 | Indian | 14 | 23 | 11 | 10 | 14 | 1 | 19 |
| m437 | Indian | 15 | 23 | 10 | 13 | 13 | 1 | 20 |
| m283 | Mon. | 16 | 23 | 10 | 14 | 13 | 1 | 21 |
| m245 | Mon. | 14 | 23 | 11 | 16 | 13 | 1 | 22 |
| m507 | Indian | 13 | 22 | 10 | 16 | 13 | 1 | 23 |
| $32 c$ | Catal. | 15 | 24 | 10 | 13 | 13 | 2 | 24 |
| $43 c$ | Catal. | 14 | 23 | 11 | 14 | 14 | 1 | 25 |
| 30 v | Basq. | 14 | 25 | 10 | 13 | 12 | 1 | 26 |
| $23 v$ | Basq | 14 | 24 | 11 | 11 | 13 | 4 | 27 |
| 2001 | Basq. | 14 | 23 | 11 | 13 | 14 | 2 | 28 |
| m503 | Indian | 13 | 23 | 10 | 13 | 13 | 1 | 29 |
| $15 N W$ | Italian | 14 | 24 | 10 | 15 | 13 | 1 | 30 |

Microsatellite abbreviations include: 390 (DYS390), 391 (DYS391), 392 (DYS392), and 393 (DYS393). Hap refers to haplotypes, and Freq. is the frequency each haplotype is observed at within the data. Sample name abbreviations: Mon - Mongolian, Catal Catalan, Basq. - Basque.


Figure 4.12. Median-joining network within hg $\mathbf{1}$ including Ladins, other European and Asian chromosomes
This network includes the Ladin-speakers, plus additional European samples and a selection of Asian chromosomes. These are all haplogroup 1 chromosomes. Each circle represents a different microsatellite haplotype and the size of the circle is proportional to the observed frequency. Unobserved haplotypes are indicated as small filled black circles. The three sample types have been given different patterns, as have the different MSY1 modular structures, that have been superimposed into the relevant haplotypes within the network.
within the Ladin-speakers (Table 4.3. for a summary of the data) were analysed in the surrounding populations.

### 4.3.5. MSY1 results

A number of interesting features emerged when examining MSY1 diversity within the Ladin speakers. A high number of identical MSY1 modular structures were detected, which are characteristic of genetically isolated populations (Jobling et al., 1998a). Unusual MSY1 codes were detected in hg 1 and hg 21 chromosomes. Three haplogroups will be discussed in more detail, hg 1, hg 9 and hg 21, but for the complete list of all MSY1 codes detected in Ladin and neighbouring populations, see Table 4.8.

### 4.3.5.1. Unusual haplogroup 1 MSY1 modular structures

Unusual MSY1 codes were detected in haplogroup 1 chromosomes (Figure 4.13), and they appeared to have undergone a rearrangement at the $3^{\prime}$ end leading to an MSY1 code which is more complicated than usual. A number of hg 1 modular structures - which refers to the order of different repeat types within the minisatellite, have been previously observed, the most commonly seen ( $47 / 49$ ) European MSY1 code had a 1,3,4 modular structure, although a few (2/49) European chromosomes had a $1,3,4,3,4$ structure. Some (4/11) Asian but no ( $0 / 49$ ) European chromosomes had a 3,1,3,4 structure (Jobling et al., 1998a). The unusual Ladin codes had a $1,3,4,3,4,3,4$ modular structure, and all of these codes were only detected in a single valley (Val di Fassa). For the complete list of MSY1 codes determined from the all samples see Table 4.8.

### 4.3.5.2. Unusual haplogroup 9 MSY1 modular structures

For the range of unusual hg 9 MSY1 modular structures found, see Figure 4.14. Very few hg 9 chromosomes were detected in comparison to hg 1 (which was the most frequently observed haplogroup within European populations) and hg 21, which again are more frequent than hg 9 chromosomes. See Table 4.8. for the complete range of modular structures determined for the Ladin and some neighbouring samples.

### 4.3.5.3. Unusual haplogroup 21 MSY1 modular structures

The usual modular structure within Europe is $0,1,0,3,4$ with the " 0 " (null) type repeat being neither a 1,3 , or 4 type of repeat (see Table 4.0. to

Table 4.8. Biallelic data and corresponding MSY1 modular structures for Ladin speakers and some neighbouring European populations.


Table 4.8. Biallelic data and corresponding MSY1 modular structures for Ladin speakers and some neighbouring European populations.

| Sample | Population | Hg | MSY1 code | rpts |
| :---: | :---: | :---: | :---: | :---: |
| NW91 | Italian |  | (1)16.(3)39.(4)19 | 74 |
| Vg4 | Val Gardena | 1 | (1)16.(3)39.(4)20 | 75 |
| Vg6 | Val Gardena | 1 | (1)16.(3)39.(4)15 | 70 |
| Co21 | Cortina |  | (1)16.(3)40.(4)16 | 72 |
| Co16 | Cortina |  | (1)16.(3)40.(4)18 | 74 |
| Slov18 | Slovenian |  | (1)16.(3)41.(4)15 | 72 |
| Bav16 | Bavarian | 1 | (1)16.(3)41.(4)16 | 73 |
| NE92 | Italian | 1 | (1)16.(3)43.(4)19 | 78 |
| Co42 | Cortina | 1 | (1)16.(3)44.(4)16 | 76 |
| Co55 | Cortina | 1 | (1)16.(3)46.(4)13 | 75 |
| Co25 | Cortina | 1 | (1)16.(3)39.(4)1.(3)3.(4)2.(3)2.(4)12 | 75 |
| Vf7 | Val di Fassa | 1 | (1)16.(3)37.(4)3.(3)2.(4)2.(3)2.(4)12 | 74 |
| Vf1 | Val di Fassa | 1 | (1)16.(3)38.(4)3.(3)2.(4)2.(3)2.(4)12 | 75 |
| Vf5 | Val di Fassa |  | (1)16.(3)38.(4)3.(3)3.(4)2.(3)2.(4)12 | 75 |
| Vf6 | Val di Fassa |  | (1)16.(3)38.(4)3.(3)2.(4)2.(3)3.(4)11 | 75 |
| Vf4 | Val di Fassa | 1 | (1)17.(3)38.(4)1.(3)3.(4)1.(3)4.(4)10 | 74 |
| NW81 | Italian | 1 | (1)17.(3)35.(4)21 | 73 |
| S149 | Italian |  | (1)17.(3)35.(4)22 | 74 |
| G1 | Greek | 1 | (1)17.(3)38.(4)19 | 74 |
| NE128 | Italian | 1 | (1)17.(3)38.(4)20 | 75 |
| Slov58 | Slovenian | 1 | (1)17.(3)41.(4)19 | 77 |
| Slov80 | Slovenian | 1 | (1)17.(3)41.(4)20 | 78 |
| Co34 | Cortina | 1 | (1)18.(3)38.(4)18 | 74 |
| Co45 | Cortina |  | (1)18.(3)38.(4)18 | 74 |
| NW99 | Italian | 1 | (1)18.(3)44.(4)18 | 80 |
| S248 | Italian | 1 | (1)22.(3)47.(4)17 | 86 |
| Bav43 | Bavarian |  | (1)22.(3)55.(4)16 | 93 |
| Bav1 | Bavarian | 1 | (1)23.(3)49.(4)18 | 90 |
| Tuk29 | Turkish | 1 | (1)18.(3)5.(1/0)1.(3)37.(4)21 | 82 |
| Tuk1 | Turkish |  | (1)18.(3)8.(1)1.(3)36.(4)20 | 83 |
| Co13 | Cortina |  | (1)19.(3)35.(4)1.(3)1.(4)16 | 72 |
| NW21 | Italian | 1 | (1)17.(3)38.(4)1.(3)3.(4)16 | 75 |
| Vf9 | Val di Fassa |  | (3)3.(1)12.(3)32.(4)19 | 66 |
| Bav59 | Bavarian |  | (3)3.(1)10.(3)39.(4)18 | 70 |
| Tuk14 | Turkish |  | (3)1.(1)13.(3)3.(1/0)4.(3)37.(4)11 | 69 |
| Tuk7 | Turkish |  | (3)1.(1)14.(3)2.(1)5.(3)38.(4)10 | 70 |
| Co54 | Cortina |  | (3)2.(1)14.(3)37.(4)23 | 76 |
| Co26 | Cortina |  | (3)2.(1)14.(3)41.(4)23 | 80 |
| Co27 | Cortina | 2 | (3)4.(1)11.(3)26.(4)26 | 67 |
| Co5 | Cortina | 2 | (3)7.(1)10.(3)31.(4)20 | 68 |
| Co44 | Cortina |  | (3)4.(1)14.(3)28.(1)17.(3)17.(4)13 | 93 |
| Co50 | Cortina |  | (3)3.(1)12.(3)34.(4)1.(3)1.(4)16 | 67 |
| Vf3 | Val di Fassa |  | (3)3.(1)12.(3)31.(4)20 | 66 |

Table 4.8. Biallelic data and corresponding MSY1 modular structures for Ladin speakers and some neighbouring European populations.

| Sample | Population | Hg | MSY1 code | rpts |
| :---: | :---: | :---: | :---: | :---: |
| Vg7 | Val Gardena |  | 2 (3)4.(1)13.(3)37.(4)20 | 74 |
| Co3 | Cortina |  | 9 (3)5.(1)7.(3)2.(1)3.(3)2.(1)1.(3)23.(4)26 | 69 |
| Vf8 | Val di Fassa |  | 9 (3)5.(1)7.(3)3.(1)2.(3)27.(4)3.(0)4.(4)19 | 70 |
| Vf10 | Val di Fassa |  | 9 (3)5.(1)7.(3)31.(4)3.(0)4.(4)19 | 69 |
| Vf2 | Val di Fassa |  | 9 (3)3.(1)11.(3)18.(4)22 | 54 |
| Vb3 | Val Badia | 21 | 1 (1)20.(3)21.(4)13 | 54 |
| Vg3 | Val Gardena | 21 | 1 (1)20.(3)37.(4)13 | 70 |
| Vb7 | Val Badia | 21 | 1 (1)20.(3)38.94)11 | 69 |
| G6 | Greek |  | 1 (0)1.(1)16.(0)26.(3)11.(4)20 | 73 |
| NW27 | Italian |  | $1(0) 1 .(1) 17 .(0) 17 .(3) 19 .(4) 20$ | 74 |
| Bav81 | Bavarian | 2 | 1 (0)1.(1)17.(0)21.(3)22.(4)12 | 72 |
| NE74 | Italian |  | 1 (0)1.(1)18.(0)25.(3)18.(4)11 | 73 |
| NE19 | Italian | 21 | 1 (0)1.(1)19.(0)18.(3)23.(4)12 | 73 |
| NE22 | Italian | 21 | 1 (0)1.(1)19.(0)20.(3)21.(4)13 | 74 |
| NW89 | Italian | 21 | 1 (0)1.(1)19.(0)20.(3)23.(4)12 | 75 |
| NW45 | Italian |  | $1 \mid(0) 1 .(1) 19 .(0) 24 .(3) 17 .(4) 12$ | 73 |
| S150 | Italian | 21 | 1 (0)1.(1)20.(0)15.(3)25.(4)13 | 74 |
| S168 | Italian |  | 1 (0)1.(1)20.(0)16.(3)26.(4)11 | 74 |
| Bav88 | Bavarian | 21 | 1 (0)1.(1)20.(0)19.(3)21.(4)11 | 71 |
| NW65 | Italian | 21 | 1 (0)1.(1)20.(0)19.(3)21.(4)14 | 75 |
| NW75 | Italian | 21 | 1 (0)1.(1)20.(0)21.(3)24.(4)10 | 76 |
| Tuk5 | Turkish | 21 | 1 (0)1.(1)21.(0)20(3)22.(4)13 | 76 |
| Slov28 | Slovenian | 21 | 1 (0)1.(1)21.(0)20.(3)21.(4)12 | 74 |
| NW14 | Italian | 21 | 1 (0)1.(1)18.(0)23.(3)20.(4)1.(3)1.(4)9 | 73 |
| NW41 | Italian | 21 | 1 (0)1.(1)19.(0)20.(3)24.(4)1.(3)1.(4)9 | 75 |
| Co6 | Cortina |  | 1 (0)1.(1)12.(3)1.(1)4.(3)36.(4)14 | 65 |
| Co23 | Cortina |  | 1 (0)1.(1)12.(3)1.(1)4.(3)34.(4)2.(0)10.(4)2 | 65 |
| Co28 | Cortina |  | 1 (0)1.(1)12.(3)1.(1)4.(3)35.(4)2.(0)10.(4)2 | 66 |
| Co37 | Cortina |  | 1 (0)1.(1)12.(3)1.(1)4.(3)35.(4)2.(0)10.(4)2 | 66 |
| Co38 | Cortina |  | 1 (0)1.(1)12.(3)1.(1)4.(3)35.(4)2.(0)10.(4)2 | 66 |
| Co33 | Cortina | 2 | 1 (1)11.(3)1.(1)4.(3)35.(4)2.(0)9.(4)2 | 64 |
| Co7 | Cortina | 2 | 1 (1)12.(3)1.(1)4.(3)35.(4)2.(0)9.(4)1 | 64 |
| Co9 | Cortina | 2 | 1 (1)12.(3)1.(1)4.(3)35.(4)2.(0)9.(4)1 | 64 |
| $\mathrm{Co4}$ | Cortina |  | 1 (1)12.(3)1.(1)4.(3)35.(4)2.(0)9.(4)2 | 65 |
| Co2 | Cortina | 2 | 1 (1)13.(3)1.(1)4.(3)35.(4)2.(0)11.(4)1 | 66 |
| G36 | Greek | 2 | 1 (1)19.(0)19.(3)22.(4)13 | 73 |
| G21 | Greek |  | 1 (1)20.(0)26.(3)15.(4)12 | 73 |
| G25 | Greek |  | 1 (1)21.(0)21.(3)18.(4)15 | 75 |
| NW102 | Italian |  | 1 (0)1.(1)17.(3)33.(4)17 | 68 |
| Bav84 | Bavarian |  | 1 (0)1.(1)17.(3)38.(4)16 | 71 |
| Bav13 | Bavarian |  | 1 (0)1.(1)13.(3)1.(1)4.(3)33.(0)14.(4)1 | 66 |
| NW39 | Italian |  | 1 (0)1.(3)1.(1)14.(3)35.(4)18 | 69 |
| Slov82 | Slovenian |  | 1 (1)21.(3)49.(4)17 | 87 |

Table 4.8. Biallelic data and corresponding MSY1 modular structures for Ladin speakers and some neighbouring European populations.

| Sample | Population | Hg | MSY1 code | rpts |
| :--- | :--- | :--- | :--- | :--- |
| S194 | Italian | $21(3) 1 .(1) 14 .(0) 17 .(3) 17 .(4) 22$ | 71 |  |

## MSY1 modular structure

## 134



1000000000000000000000000000000000000000000000000000000000 evepenevepevee

## 13434



## 1343434



## 3134



## 313134

$00000000000000000000000000000000000000000000000000000000000000 勹 \omega 00$
10000000000000000000000000000000000000000000000000600 0

## Sample and modular structure

Co10 (1)15.(3)41.(4) 18
Co21 (1)16.(3)40.(4) 16
Tuk28 (1)15.(3)34.(4)22
NW99 (1)18.(3)44.(4) 18
S248 (1)22.(3)47.(4)22

Col3 (1)19.(3)35.(4)1.(3)1.(4)16

Co25 (1)16.(3)39.(4)1.(3)3.(4)2.(3)2.(4)12
Vf7 (1)16.(3)37.(4)3.(3)2.(4)2.(3)2.(4)12
Vf1 + Vf5 (1)16.(3)38.(4)3.(3)2.(4)2.(3)2.(4) 12 *
Vf6 (1)16.(3)38.(4)3.(3)2.(4)2.(3)3.(4)11
Vf4 (1)17.(3)38.(4)1.(3)3.(4)1.(3)4.(4)10

Vf9 (3)3.(1)12.(3)32.(4)19

Tuk7 (3)1.(1)14.(3)2.(1)5.(3)38.(4)10
Tuk 14 (3)1.(1)13.(3)3.(1)4.(3)37.(4)11

Figure 4.13. MSY1 modular structures within haplogroup 1 belonging to Ladin-speakers and some surrounding populations The types of repeats have a different colour - type 1 (red), type 3 (yellow), and type 4 (blue). The modular structure is written above each set of relevant codes, and examples of chromosomes with that code are indicated on the right. Identical codes are represented with an asterisk.

## MSY1 modular structure

3134


31313134
000000000000000000000000000000000000000

## 313404

## 

## 31313404

## Sample and modular structure

Co3 (3)5.(1)7.(3)2.(1)3.(3)2.(1)1.(3)23.(4)26

Vf10 (3)5.(1)7.(3)31.(4)3.(0)4.(4)19

Vf8 (3)5.(1)7.(3)3.(1)2.(3)27.(4)3.(0)4.(4)19

Figure 4.14. MSY1 modular structures within haplogroup 9 belonging to Ladin-speaking populations The types of repeats have a different colour - type 1 (red), type 3 (yellow), and type 4 (blue). The modular structure is written above each set of relevant codes, and examples of chromosomes with that code is indicated on the right.

```
134
*000<0,0,0,0,0,0,0000000000000000000000000000000000000000eveveeveve
\begin{tabular}{ll}
\(\mathrm{Vb3}\) & (1)20.(3)21.(4)13 \\
\(\mathrm{Vb7}\) & \((1) 20 .(3) 38 .(4) 11\) \\
Vg 3 & (1)20.(3)37.(4) 13
\end{tabular}
```

1034
с,

## 01313404

©0000000000000000000000000000000000000000000000000000000000000000
00000000000000000000000000000000000000000000000000000000000

## 1313404



cococcovoce0000000000000000000000000000000000000cev0000evevee
00000000000000000000000000000000000000000000000000000000

## 0103434



| G21 | (1)20.(0)26.(3)15.(4)12 |
| :--- | :--- |
| G25 | (1)21.(0)21.(3)18.(4)15 |
| G36 | (1)19.(0)19.(3)22.(4)13 |

$\mathrm{Co} 28+\mathrm{Co} 37+\mathrm{Co} 38$
(0)1.(1)12.(3)1.(1)4.(3)35.(4)2.(0)10.(4)2 2
Co 23 (0)1.(1)12.(3)1.(1)4.(3)34.(4)2.(0)10.(4)2

| Co2 | (1)13.(3)1.(1)4.(3)35.(4)2.(0)11.(4)2 |
| :---: | :---: |
| Co4 | (1)12.(3)1.(1)4.(3)35.(4)2.(0)9.(4)2 |
| Co33 | (1)11.(3)1.(1)4.(3)35.(4)2.(0)9.(4)2 |
| Co7+ | $\begin{aligned} & \mathrm{Co} 9 \\ & \text { (1)12.(3)1.(1)4.(3)35.(4)2.(0)9.(4) } 1 * \end{aligned}$ |

NW14 (0)1.(1)18.(0)23.(3)20.(4)1.(3)1.(4)9 NW41 (0)1.(1)19.(0)20.(3)24.(4)1.(3)1.(4)9

NE19 (0)1.(1)19.(0)18.(3)23.(4)12
S150 (0)1.(1)20.(0)15.(3)25.(4)13
Tuk5 (0)1.(1)21.(0)20.(3)22.(4)13
Slov28(0)1.(1)21.(0)20.(3)21.(4)12
Bav88 (0)1.(1)20.(0)19.(3)21.(4)12

Figure 4.15. Some MSY1 modular structures within haplogroup 21 belonging to the Ladin -speakers and some surrounding populations The types of repeats have a different colour - type 1 (red), type 3 (yellow), and type 4 (blue). The modular structure is written above each set of relevant codes, and examples of chromosomes with that code is indicated on the right. Identical codes are represented with an asterisk.
explain the difference between the types of repeats). The sequence fails to amplify with the repeat-specific primers, and a gap or fainter product is generated. A range of unusual modular structures were detected (Figure 4.15) within the haplogroup 21 individuals examined. For example, $0,1,3,1,3,4,0,4$ and $1,3,1,3,4,0,4$, which appear to only differ from each other by the presence or absence of the null type repeat at the 3 ' end of the array. These modular structures were only detected within the Cortina sample set. For the complete list of MSY1 codes determined from the Ladinspeaking samples see Table 4.8.

### 4.3.6. Examination of different MSY1 lineages within the Ladin speakers

We can assume a number of different sublineages are present because some of the MSY1 modular structures are so different, see Table 4.8. These modular structures can be sub-divided and considered as separate lineages within individual haplogroups. This practice is justified, since some MSY1 sublineages are found subsequently to correspond to SNP-defined lineages (Hurles, 1999). The unusual codes can be considered separate lineages to allow further dissection within certain interesting haplogroups, for example, hgs 1 and 21 (Figure 4.16). Six different lineages (1.0 to 1.5 ) were observed within the hg 1 data, and some of these lineages are shared for example, 1.0 is shared between all four populations. However, certain lineages are population-specific for example, 1.1 and 1.5 are Turkish-specific, and 1.3 is Ladin-specific. Within hg 21 a total of nine different lineages ( 21.0 to 21.8 ) were observed, and again some sharing of lineages is observed, for example 21.0 is detected within the Bavarian, Greek and Italian samples but not the Ladins. The Ladin samples show two specific lineages (21.1 and 21.6), but they also share a lineage (21.4) with only the Bavarian samples, and this may reflect some limited gene flow.


Figure 4.16. MSY1 lineage diversity within two biallelic haplogroups
The different MSY1 modular structures have been given a different colour (see key). Two main haplogroups have been examined in more detail (hg 1 and hg 21), and their locations within the maximum parsimony tree are indicated in the above figure. A number of populations within both haplogroups have been examined in more detail, see keys.

### 4.4. Discussion

### 4.4.1. General observations

The mtDNA analysis showed Ladins to be as different from other European populations as are Africans (Stenico et al., 1996). This seems hard to reconcile with the $Y$ data, which shows that, although they are significantly different from their neighbours, they nonetheless contain typical European haplogroups. It has been suggested that the mtDNA data may be due to sequencing errors (Bandelt, 1999).

### 4.4.2. Haplogroup examination

Haplogroup analysis shows apparent low diversity within the Ladins, because only four haplogroups (hgs 1, 29 and 21) are detected. It is also clear that some other lineages (hg3, hg26) are present at reasonable frequencies in neighbouring populations but are absent from Ladins. This may be due to either drift through isolation, or a founding population which lacked these haplogroups, or a combination of these two. These findings certainly suggest that gene flow from the rest of Italy in the south, and from the north (e.g. Bavaria) has not been free, and confirm the effect of isolation, perhaps both geographic and linguistic.

### 4.4.3. Multiallelic analysis

The two different multiallelic systems microsatellites and MSY1, agree with each other - the Ladin-speakers seem diverged from neighbouring Europeans. This indicates the shortcomings of biallelic markers - in that they give a low resolution of subdivision, which can miss a lot of the true diversity. Microsatellite network analysis (Bandelt et al., 1999) indicates they cluster more closely with the European than Asian samples, which may indicate a European origin, although the Asian samples are limited. The 'unusual' MSY1 structures found within the two major Ladin hgs, 1 and 21, are not found at appreciable frequency in neighbours, and this again emphasises the lack of male-mediated gene flow between these populations. These 'unusual' lineages will probably correspond to specific biallelic-defined lineages when new biallelic markers found (see Chapter 5). The major question is to try and decide whether these lineages reflect founders from non-neighbouring populations (model C), or in situ differentiation (model A). Unfortunately, the incompleteness of European and

Middle Eastern data on MSY1, microsatellites and haplogroups does not allow us to test the first idea (model C) properly, and this emphasises the need for good sampling with all potentially useful markers.

The second model (A - in situ differentiation) can be considered: would 12000 years ( $400-600$ generations) be enough time to accumulate the MSY1 modular structures and the level of microsatellite diversity observed? It may possibly. It would be extremely helpful to have more knowledge about MSY1 mutation processes to model this; and more biallelic subdivision to see if the Ladin sub-lineages correspond to specific sub-haplogroups, and if so, to ask where these are in the rest of the world. If they do correspond to biallelic-defined sublineages, which are defined by low mutation rate markers, this would suggest, as was suggested by Stenico (1996) for the mtDNA data, that we are not seeing in situ differentiation.

### 4.4.4. Conclusions

Unfortunately, at the moment it is not possible to distinguish between in situ differentiation (model A) and differentiated founder population (model C) with current data, although these may have worked in concert together rather than two separate historical events. However, it is possible to conclude that the Ladin-speakers show a strong signal of genetic isolation. The addition of the atypical Ladins to the European dataset demonstrates that, with the exception of barrier analysis, the methods of Chapter 3 are relatively robust to the introduction of new populations.

# Chapter 5: A high resolution, global study of a single lineage, YAP (Y Alu polymorphism) 

### 5.1. Introduction

This chapter is concerned with a detailed analysis of a subset of Y chromosomal lineages defined by the polymorphic insertion of an Alu element. The initial part of this introduction describes the repetitive element types within the human genome, the Alu family in particular, followed by Alu polymorphisms and in particular a $Y$ chromosome-specific example (YAP) examined in more detail within this work. The YAP lineage will be dissected into sub-lineages incorporating new biallelic markers and their distribution within three continental samples determined. After the definition of biallelic haplogroups the diversity within each will be further examined using microsatellites.

### 5.1.1. Repetitive elements

Contained within all vertebrate genomes are regions of interspersed repetitive DNA; these may have been, or may still be, mobile (Britten, 1996; Smit, 1996), but there is little evidence of active transposable elements within animal genomes. Repeated sequence elements are dispersed to thousands of locations, and include LINEs (Long Interspersed Nuclear Elements) and SINEs (Short Interspersed Nuclear Elements), each of which comprise approximately $15 \%$ of our genome (Okada, 1991; Smit, 1996). SINEs are not generally repeated in tandem arrays, but tend to appear as isolated elements, but still may be present at thousands of copies within the genome (Watson et al., 1992). They are also characterised by an internal and mobile polymerase III promoter that ensures the transcriptional activity for new copies (Okada, 1991).

The most abundant SINE in the human genome, is the Alu family (Watson et al., 1992), and it is estimated that the number of these elements range from 0.5 million to 1 million copies per haploid genome (Britten, 1996; Smit, 1996). Alu family members are found in all primate species. The 'Alu family' is so named because the (AGCT) recognition sequence for the restriction enzyme AluI (Watson et al., 1992) is contained within each element. A complete Alu sequence is approximately 281 bp in length, with a tail of between 20 to 30 adenosine (A) bases (the poly A tail) (Britten, 1996). Individual members of the Alu family are related but not identical, and a number of related sequences have also been
identified in other animals, for example, in the mouse (B1 family) and in the Chinese hamster (Alu-equivalent family) (Lewin, 1990).

Once an Alu element has inserted into a chromosomal location, it remains stable (Batzer et al., 1994), and a typical Alu insert may have undergone subsequent base substitutions particularly involving GC rich ( CpG ) regions, or an insertion/deletion event (Britten, 1996). A certain degree of clustering of Alu elements is detected within the genome, particularly around certain genes (Riccio and Rossolini, 1993). A number of Alu sequences have been identified as relatively recent insertions, because they are polymorphic in human populations, or may exhibit inter-specific differences between apes (Britten, 1996). Primate Alu sequences can be further subdivided into related sub-families that share diagnostic nucleotide substitutions (Slagel et al., 1987).

SINEs and LINEs share a variable length insertion duplication site and a 3' poly A tail, suggesting they have integrated into the human genome by a process known as reverse transcription (Smit, 1996) via a polyadenylated RNA intermediate (Watson et al., 1992). LINEs encode a reverse transcriptase and other proteins required for retrotransposition (Smit, 1996). Most Alu repeats are transcribed into RNA by RNA polymerase III, which also transcribes 5 S rRNA, all tRNAs and a number of other RNA genes, that encode RNA molecules that perform structural roles (Watson et al., 1992). The 300bp Alu repeat shows sequence similarity to the 7SL RNA, which is an abundant RNA molecule, and part of the signal-recognition particle (SRP), that aids transportation of proteins to the endoplasmic reticulum for secretion (Watson et al., 1992). The 7SL RNA corresponds to left end of Alu, and $403^{\prime}$ terminal bases of 7SL correspond to the right end, but the central 160 bases of 7SL has no homology to Alu (Lewin, 1990).

Alu family members resemble transposons, because they are flanked by short direct repeats, and the length of these repeats is different for individual family members (Lewin, 1990). Transposition can occur via reverse transcription of an RNA intermediate (retrotransposition), or via excision and reintegration of the DNA itself (DNA transposition). Most retrotransposed repeats are, SINEs, LINEs, or retrovirus-like elements (Smit, 1996). A 14bp region of the Alu sequence is nearly identical to sequence present at the origin of replication in Papova viruses (e.g. SV40) and Hepatitis B virus. A possible function of the Alu family could be some connection with the origin of replication for eukaryotic
genomes (Lewin, 1990); however this is unlikely due to their high number within the genome.

### 5.1.2. Alu elements as polymorphic markers

Each insertion event into a specific region of the human genome was unique and occurred after the divergence of humans and great apes (Batzer et al., 1994; Hammer, 1994). Since non-human primates lack Alu insertions at the homologous position, absence is the ancestral state (Batzer et al., 1994). There are many studies looking at polymorphic autosomal Alu elements (Stoneking, 1993; Batzer et al., 1994), including the examination of the PLAT locus (Tishkoff et al., 2000). The pattern of haplotype diversity including the presence or absence of an Alu element, suggests that the recent common ancestry of non-African populations lies within eastern Africa (Tishkoff et al., 2000).

### 5.1.3. The YAP polymorphism

The YAP (Y Alu polymorphism - DYS287) lineage is defined by the insertion of a 300 bp Alu element at a specific site on the long arm of the $Y$ chromosome (Yq) (Hammer, 1994), although it is not flanked by direct repeats. If a chromosome contains the Alu insertion it is designated as YAP+ (positive), and without as YAP- (negative). The YAP lineage is contained within the maximum parsimony tree, and for the position within this, see Figure 5.0.

### 5.1.4. The YAP lineage global distribution

The estimated age of Y-chromosomal diversity is approximately 156,000 years, and the YAP lineage is also considered to be extremely old and has been dated to approximately 141,000 years using human haplotype sequence divergence compared to human-chimp divergence (Hammer, 1995). Recent data indicated that the YAP lineage has an unusual global distribution (see Table 5.0 and Figure 5.1) (Karafet et al., 1999).


## Key

O Markers being analysed
Indicates ancestral (0) and derived (1) states Shows which allele is digested

Figure 5.0. Position of the YAP+ lineage within the Maximum parsimony tree of $Y$ haplogroups
Green circles indicate haplogroups being analysed during this work. Haplogroups in white circles are not involved in this work.


Figure 5.1. Global distribution of YAP chromosomes
The data sets include samples from Sub-Saharan Africa (348), European (175), North Asian (374), Central Asian (572), South Asian (229), East Asian (60), Australasia (307) and Native American (133). The two colours in the pie charts indicate the frequency (dark) of this polymorphism within the populations, and the size of the circle is uniform but the actual sample size of each population is indicated in brackets. These data are taken from Karafet et al., 1999.

Table 5.0. YAP frequency data determined for 8 different global samples.

| Global populations | Sample size | hg 4 | hg 8 | hg 21+ | YAP+ |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Sub-Saharan African | 348 | 0 | $48 \%$ | $16 \%$ | $64 \%$ |
| European | 175 | 0 | 0 | $13 \%$ | $13 \%$ |
| North Asian | 374 | 0 | 0 | 0 | 0 |
| Central Asian | 572 | $9 \%$ | 0 | 0 | $9 \%$ |
| South Asian | 229 | 0 | 0 | 0 | 0 |
| East Asian | 60 | $19 \%$ | 0 | 0 | $19 \%$ |
| Australasian | 307 | 0 | 0 | 0 | 0 |
| Native American | 133 | 0 | 0 | $0.7 \%$ | $0.7 \%$ |

Haplogroup 4 and 8 data were taken from Karafet et al. (1999), and have been superimposed onto a global landscape, see Figures 5.1 and 5.2 to give a visual impression of their unusual global geographical distribution. Haplogroup 21+ data includes haplogroups not yet described here, and has not been shown graphically.

This lineage is completely absent in the Americas, and Oceania, but is detected at low frequencies in Caucasian and Indian populations, and is very common in Sub-Saharan African populations. Within Asia it is completely absent from nearly all populations with the exceptions of Japanese and Tibetans.

### 5.1.4.1. Diversity within $Y A P+$ lineages

A number of polymorphisms were detected within the YAP region, two $C$ to $T$ transitions named PN1 and PN2 respectively, a $G$ to $A$ transition (PN3) and a variable length poly (A) tail associated with the YAP element (Hammer et al., 1997). PN3 was only observed within YAPchromosomes, but in contrast all the others (PN1, PN2, and the poly (A) tail) could be used to define YAP+ sub-lineages or haplogroups ( 3 to 5 ) (Hammer et al., 1997; Karafet et al., 1999; Hammer et al., 2000). Other researchers use a different nomenclature with regard to these haplogroups, as follows: hg 4: haplotype 3G; hg 8: haplotype 5; and hg 21: haplotypes $3 \mathrm{~A}+4$.

The PN1 derived (T) lineage was observed at a frequency of approximately $43 \%$ within Africans and $5 \%$ within West Asians (Hammer et al., 1997). In comparison the PN2 derived (T) lineage was completely absent within Asian populations but observed at frequencies of $56 \%$ and $14 \%$ within Africans and Europeans respectively (Hammer et al., 1997). Three poly (A) tail lengths were observed (long - 46bp, short - 26bp and very short -19 bp ) although a medium length (36bp) was also detected in some individuals. The distribution of these variant length polymorphisms
associated in a geographically specific manner also, the long tail mainly ( $88 \%$ ) associated with East Asian YAP+ chromosomes. The short tail was absent from East Asia, and frequently ( $>90 \%$ ) within European and Africans; and the very short tails were only associated at low frequency with East Asian YAP+ chromosomes (Hammer et al., 1997). The poly (A) tail polymorphism has not been analysed during the examination of YAP sub-lineages within this work.

Contained within this YAP Y-chromosomal lineage are a number of geographically interestingly distributed sublineages (see Table 5.0). Haplogroup 4 is the most ancestral (115,000 years old) (Hammer, 1994), and is only observed in central and eastern Asian populations (see Figure 5.2A) (Altheide and Hammer, 1997; Hammer et al., 1998; Karafet et al., 1999). The most important polymorphism with regards to the sub division of ancestral (hg 4) and derived lineages (hgs 21 and 8) is SRY-8299 (Whitfield et al., 1995a). Haplogroup 8 (defined by sY81 (DYS271) G allele), is mainly detected within Sub-Saharan African individuals (see Figure 5.2B) (Seielstad et al., 1994; Hammer et al., 1998; Karafet et al., 1999) and has been estimated to be approximately 20,000 years old (Hammer et al., 1998) from coalescence analysis.

It was proposed that the specific and unusual distribution of the ancestral (hg 4) and derived lineages (hgs 21 and 8) could be taken as evidence that the mutation occurred outside of Africa, due to the high frequency of ancestral chromosomes in Asia, and complete absence in Africa (Altheide and Hammer, 1997). So it was suggested the mutation probably occurred somewhere in Asia, then re-entered Africa at a later date (see Figure 5.3A); this has been called the 'Out of Africa and back again hypothesis' (Altheide and Hammer, 1997). The derived hg 8 lineage was thought to be subsequently spread by the Bantu farming expansion (Hammer et al., 1998). However, no archaeological evidence has been found to support the 'Out of Africa and back again hypothesis'. Another possible explanation to explain this geographical distribution could be that a previously unobserved or extinct haplogroup separates hg from the main YAP branch (Figure 5.3B), and with the increasing number of biallelic markers it may soon be possible to detect this.


## B



Figure 5.2. Global distributions of haplogroups 4 and 8
Frequency data of haplogroups within all Y-chromosomal lineages, from Karafet et al., 1999 (A) shows haplogroup 4 (corresponding to hap3G of Hammer et al., 1998) and (B) haplogroup 8 (corresponding to hap 5 of Hammer et al., 1998) lineages. The two colours in the pie charts indicate the frequency (dark) of these lineagess within the populations. The size of the circle is uniform but the actual sample size of each population is indicated below in brackets. These data include Sub-Saharan African (348), European (175), North Asian (374), Central Asian (572), South Asian (229), East Asian (60), Australasian (307) and Native American (133) samples.

A


B


Figure 5.3. Two possible ways to explain the unusual geographical distribution of YAP sub-lineages
(A) shows the geographical distribution of chromosomes that has been taken by Altheide and Hammer, 1997 to indicate the YAP mutation did not occur in Africa due to the detection of only Asian-specific hg 4 chromosomes. However (B) shows a potential alternative explanation to explain this same distribution involving an extinct or previously unobserved haplogroup.

The examination of MSY1 diversity within YAP+ chromosomes revealed a high level of repeats that failed to amplify, referred to as 'nulls', and it was suggested that this may mask a large degree of diversity previously unobserved within African chromosomes (Jobling et al., 1998a). Further analysis of hg 8 chromosomes, suggested this ancestral lineage shows a high degree of structural and length diversity, and indicated the lack of amplification was due to a single base change within the MSY1 array, which affects all the subsequent repeat types along the array (Bouzekri et al., 1998). Due to the population specificity of MSY1 modular structures, the relationship between populations can be examined, and it was found there was no strong evidence for an Asian input into the African Y chromosome pool (Jobling et al., 1998a). This influence may have been expected if the 'Out of Africa and back again hypothesis' (Hammer, 1994; Altheide and Hammer, 1997; Hammer et al., 1998) was correct.

### 5.1.5. The Bantu expansion

A number of different African dispersals have been recorded historically, and one such event occurred between 500 and 3,500 years ago, and is known as the Bantu expansion. Agricultural technology swept into Africa from the Middle East approximately 10,000 years ago (Figure 5.4A), this was followed by an expansion of Bantu-speaking people, who originated somewhere between Nigeria and Cameroon (Figure 5.4B). This migration progressed mainly in a southerly and easterly direction. The adoption of agriculture over the historically favoured Pygmy and Bushmen (Khoisan) hunter-gathering lifestyle would have probably been necessary to survive. It is possible that climate changes or overpopulation may lead to the adoption of agriculture in order to sustain the newly expanding population (Cavalli-Sforza et al., 1994). The existing $Y$ chromosomes would have been largely replaced by these invading Bantu farmers over the past 3,000 years and that may be the reason why young Y-chromosomal lineages have become predominant in Africa. Biological and archaeological evidence suggests major expansions into central and southern Africa before the Iron Age ( $\sim 2,800$ years ago) (Cavalli-Sforza et al., 1994).

The Khoisan family contains approximately 30 languages with 120,000 speakers, and these languages are confined mainly to western South Africa (Bushmen and Hottentots), except for two spoken by two Tanzanian tribes (the


Figure 5.4. The distribution of the Bantu languages within Africa A diagramatic representation of two historic migrations on the African continent. (A) shows where the Neolithic farmers would have entered approximately 9,000 years ago, and that agriculture spread and was common in the Sahara. (B) shows a much larger migration involving the Bantu language expansion, associated with the spread of agriculture within the Africa. (The figures were adapted from Cavalli-Sforza \& Cavalli-Sforza, 1994).

Hadza and Sandawe). Khoisan languages use "click" sounds, which have been borrowed by other populations, such as the Bantu tribes of Southern Africa (Cavalli-Sforza et al., 1994). Khoisanids were hunter-gatherers when the first farmers arrived in Southern Africa, but the majority of Khoisan speakers in South Africa such as the Khoikhoi (Hottentots) and the San are mainly agriculturalists today. The Bantu subgroup encompasses about 500 languages with approximately 100 million speakers and include more than half of all Africans (Cavalli-Sforza et al., 1994). There are two proposed locations for the Bantu homeland, and the first is based upon linguistic comparisons (Greenberg, 1963) of languages spoken today that are similar to Bantu. It is thought to be between Nigeria and Cameroon and this correlates well with iron culture archaeological evidence from around 500 B.C. at Nok (Flight, 1988). An alternative location for the proposed Bantu language centre is between southern Zaire and northern Zambia (Guthrie, 1967), and these locations correspond to different Iron Age expansions (Cavalli-Sforza et al., 1994).

### 5.1.6. Asian migrations

The high frequency of YAP+ chromosomes in the Japanese could reflect Asian ancestors migrating into Japan with this polymorphism, and the YAP+ chromosomes could have been lost by drift on the Asian mainland (Hammer and Horai, 1995). East Asia encompasses the most eastern part of China, as well as Korea and Japan. Two main migrations are known to have brought people from Asia to the Japanese islands (Figure 5.5) (Hammer and Horai, 1995), and these may have initially occurred when land bridges were present approximately 18,000 years ago. The first involved the Jomon people; who got their name from rope patterns introduced in clay pottery (Hammer and Horai, 1995) and are considered by some to be Japanese Neolithic; they probably arrived approximately 10,000 years ago, and were responsible for bringing pottery to Japan. It has been suggested that the Jomon would have had plentiful supplies of food, such as fish, and also cultivated a number of products such as yams (Hammer and Horai, 1995). The second proposed migration approximately 4,300 years ago, involved the Yayoi people from the Korean peninsula, who were responsible for introducing weaving, metalwork and rice into the Japanese culture (Cavalli-Sforza et al., 1994; Hammer and Horai, 1995).

An important evolutionary question is from which population, the Jomon or Yayoi, did the mainland Japanese derive? In previous analyses YAP+


Figure 5.5. The peopling of Japan following major migrations from Eastern Asia and the distribution of YAP+ lineages
Two major historical migrations from Eastern Asia brought people (the Jomon and Yayoi) into Japan. Y-chromosomal data from a previous study (Hammer and Horai,1995) have been superimposed onto this map, to show the range of YAP+ chromosomes within the Japanese archipelago.
chromosomes were observed within three Japanese populations ( $\sim 60 \%$ in Okinawa - which is known the Yayoi people had the least genetic influence from due its location), but at low frequencies in Koreans (Hammer and Horai, 1995). It has been suggested that YAP+ lineages were introduced by the Jomon, and YAPchromosomes by Yayoi from Korea, which may have diluted out YAP+.

### 5.1.7. Previous population studies

A recently published paper using a mixture of biallelic and microsatellite loci aimed to determine where the YAP+ lineages originated (Bravi et al., 2000). Some new data generated from global populations were pooled with previously published data by Hammer et al. (1997), and two models were proposed to explain the unusual distribution of the haplogroups. An Asian model proposing that 3G (hg 4) and 3A (hg 21) originated in Asia, 3A migrated into Africa, and 3A became extinct in Asia. Or an African model proposing that 3G emerged in Africa, and was lost after a migration to Asia (Altheide and Hammer, 1997; Hammer et al., 1998). However, no new evidence is provided to suggest which of these three models was the most likely (Bravi et al., 2000).

### 5.1.8. Questions to be addressed

The main aim of this work is to try and explain the unusual geographical distribution of YAP+ lineages, which involves sub-dividing the YAP lineage employing new biallelic markers. These new sublineages will be mapped geographically, and population differences and diversities will also be examined. The second aim involves the analysis of 17 Y -chromosomal specific microsatellites, to examine the relative diversity of the YAP+ lineages within different continental populations. Depending on how the microsatellites cluster with respect to haplogroup and geographic affiliations this may also help explain the unusual geographic distribution of this YAP lineage within global populations. The robustness of networks can also be examined by using different numbers of microsatellites to generate them, to see if the resolution improves. The third aim is to date the YAP sub-haplogroups, which may indicate if previous historic events such as migrations could be assigned to different lineages, which may help explain their geographical distribution.

### 5.2. Materials and methods

### 5.2.1. Genomic DNA samples

A total of 402 YAP+ DNA samples spanning 3 continents were successfully typed, and these are listed in alphabetical order including, from Europe: Bavarian, Belarussian, Bulgarian, Bulgarian gypsy, Chuvash, Cypriot, Dutch, East Anglian, Finland, French, Greek, Greenlandic Inuit, Irish, Italian, Ladin, Romanian, Russian, Sardinian, Serbian, Slovenian, Spanish, Swedish, Turkish. From Africa: East Bantu, West Bantu, Gambian, Kenyan, North African, Mbuti, Nigerian, Togolese and Zimbabwean. From Asia: Chinese, Indian, Japanese, Lebanese, Mongolian, Pakistani, and Tibetan. The Lebanese samples have been included with the Asian samples because of the lack of any other Middle Eastern samples.

YAP+ genomic DNA samples were obtained from a variety of sources the European chromosomes were the same samples used during Chapter 3. Samples from the YCC (Y Chromosome Consortium) DNA collection were from Mike Hammer (Arizona, USA). The Japanese samples were from Keiji Tamaki (Sapporo, Japan) and the Indian and Chinese samples from Chris Tyler-Smith (Oxford, UK).

### 5.2.2. Markers typed

A total of 7 new markers were typed including three (M33, M78 and M81) provided pre-publication by Peter Underhill (Stanford, USA), and DYS391 (Figure 5.6) provided by Peter de Knijff (Leiden, Netherlands). As well as published markers PN1 (Hammer et al., 1997) PN2 (Hammer et al., 1997), M15 (Underhill et al., 1997), and a number of biallelic markers currently typed via PCR length polymorphism (YAP), PCR-RFLP (SRY-8299) and allele-specific PCR (12f2).

### 5.2.3. Allele-specific oligonucleotide typing

New flanking primers and allele specific oligonucleotides (ASOs) were designed for all of the 7 markers described above. The marker-specific flanking primers for amplification were designed to have an estimated annealing temperature of approximately $54^{\circ} \mathrm{C}$, to allow the possibility of multiplexing markers in the future. The ASOs for detection were designed as 18 mers with the


Figure 5.6. The DYS391 PCR-RFLP biallelic assay
FLDO-79 sample is known to have this particular mutation (C/G). C allele loses a Tsp45I restriction site, and two fragments are detected: 98 bp and 81 bp . G allele (hg1 and hg2 chromosomes) gains a Tsp45I cut site, and three fragments are detected: $98 \mathrm{bp}, 61 \mathrm{bp}$ and 20bp. Partial digestion is observed in both FLDO79 samples, with an additional fragment visible above the 98 bp band.
single nucleotide polymorphism (SNP) being located 8 nucleotides away from the $5^{\prime}$ end, because this maximises discrimination between alleles. The following protocol for generating dot blots and hybridisation of ASOs was designed by Jeffreys et al. (1998).

### 5.2.3.1. Dot blotting

The following chemicals were used to prepare the denaturing mix, to be used during the generation of filters via dot blotting using a vacuum manifold. The denaturing mix is stored at room temperature.

| Denaturing mix | 0.5 M NaOH |
| :--- | :--- |
|  | 2 M NaCl |
|  | 25 mM EDTA |

### 5.2.3.2. Preparing dot blots

Two ( $12.5 \mathrm{~cm} \times 9 \mathrm{~cm}$ ) Hybond $\mathrm{N}+$ (Amersham Pharmacia Biotech Ltd.) membranes (replica blots) and one ( $13 \mathrm{~cm} \times 10 \mathrm{~cm}$ ) 3MM filter paper were soaked in water. The soaked 3MM filter paper was placed onto the dot blot manifold followed by a membrane (the additional thickness reduces well leakage). The dot blot apparatus was assembled, tightening the screws diagonally to ensure even pressure across the whole block, and attached to a water vacuum line.

### 5.2.3.3. Denaturing PCR products and generating the dot blots

A trace amount of Bromophenol blue (BPB - Fisher Scientific, UK) was added to an aliquot of denaturing mix, to allow the location of dots to be visualised, and to aid the loading of additional samples. A multichannel pipette was used to load 8 samples at a time. Then $30 \mu \mathrm{l}$ of denaturing mix/BPB were added to each well containing PCR products, and the vacuum was applied. A volume of $30 \mu \mathrm{l}$ of denaturing mix (containing no Bromophenol blue dye) was pipetted into each well of the manifold matrix (ensuring no bubbles were trapped), followed by $30 \mu \mathrm{l}$ of denatured DNA (with Bromophenol blue dye) into each well. A total volume of $80 \mu \mathrm{l}(2 \times 40 \mu \mathrm{l}) 2 \mathrm{x}$ SSC was added to each well in order to neutralise the DNA. The dot blot manifold was dismantled and the filter was marked in order that the array was orientated. The 3MM papers were removed and discarded, because a freshly soaked filter and 3MM paper
were used per dot blot. The newly generated dot blot filters were allowed to air dry on 3 MM filter paper, and transferred to an $80^{\circ} \mathrm{C}$ oven for 10 minutes and UV cross-linked to fix the DNA.

### 5.2.3.4. Oligonucleotide labelling

The following solutions were prepared using the chemicals below, and stored at $-20^{\circ} \mathrm{C}$.

10x Kinase mix

> 700 mM Tris- HCl pH 7.5 100 mM MgCl 2 50 mM spermidine trichloride 20 mM dithiothreitol

Kinase stop 25mM EDTA
solution $\quad 0.1 \%(w / v)$ SDS
$10 \mu \mathrm{M}$ ATP

### 5.2.3.5. Radioactive labelling of allele-specific oligonucleotides

The following labelling mix was prepared and used immediately.

Mix $\quad 1 \mu \mathrm{l} 10 \mathrm{x}$ Kinase mix
$1 \mu \mathrm{l} 1.4 \mu \mathrm{M}(8 \mu \mathrm{~g} / \mathrm{ml})$ ASO
$0.3 \mu \mathrm{l} 10 \mathrm{mCi} / \mathrm{ml} \gamma^{32} \mathrm{P}$-ATP
$7.7 \mu$ water
$0.4 \mu \mathrm{l} 10 \mathrm{u} / \mu \mathrm{l}$ T4 polynucleotide kinase

The mixture was incubated at $37^{\circ} \mathrm{C}$ for 1 hour; $20 \mu \mathrm{l}$ of kinase stop solution were added, and the radioactively labelled ASOs were used immediately for hybridisation.

### 5.2.3.6. Preparation of hybridisation solutions

The Denhardt's solution was initially prepared, filter-sterilised and stored at $-20^{\circ} \mathrm{C}$, and used to prepare both the hybridisation solutions (stored at $4^{\circ} \mathrm{C}$ ). TMAC (Tetramethylammonium chloride, Sigma-Aldrich Chemical company Ltd.) was handled in a fume hood.

| solution | 5 g polyvinyl pyrrolidone <br> 5 g BSA (fraction V, Sigma) <br> Water up to 500 ml |
| :--- | :--- |
|  |  |
| Hybridisation |  |
| solution | 60 ml 5 M TMAC |
|  | $6 \mathrm{ml} \mathrm{10} \mathrm{\%} \mathrm{(w/v)} \mathrm{SDS}$ |
|  | 0.2 ml 0.5 M EDTA |
|  | 10 ml 0.1 M sodium phosphate pH 6.8 |
|  | 10 ml 50 x Denhardt's solution |
|  | $40 \mu \mathrm{ml} 10 \mathrm{mg} / \mathrm{ml}$ yeast RNA |
|  | 13.8 ml water |
|  |  |
|  | 120 ml 5 M TMAC |
| TMAC wash | $12 \mathrm{ml} 10 \%$ (w/v) SDS |
| solution | 0.4 ml 0.5 M EDTA |
|  | 20 ml 0.1 M sodium phosphate pH 6.8 |
|  | 48 ml water |

### 5.2.3.7. Hybridisation protocol

Two dot blots were processed in parallel per SNP, one for each ASO. The dot blots were soaked in $2 x$ SSC, transferred to hybridisation bottles, to which 4 ml hybridisation solution were added (in order to prehybridise the filters), and were incubated at $52^{\circ} \mathrm{C}$ for approximately 10 minutes. This pre-hybridisation solution was discarded, and 3 ml of fresh hybridisation solution plus $15 \mu \mathrm{l} 3 \mathrm{mg} / \mathrm{ml}$ single-stranded (heat denatured for 5 minutes, and immediately cooled on ice) herring-sperm DNA were added, and incubated at $52^{\circ} \mathrm{C}$ for 5 minutes. The $\gamma^{32} \mathrm{P}$-labelled ASO plus $20 \mu \mathrm{l} 1.4 \mu \mathrm{M}$ of the complementary unlabelled ASO were added to each appropriate hybridisation bottle, and they were allowed to hybridise at $52^{\circ} \mathrm{C}$ for 1 hour.

Following hybridisation, the radioactively labelled probes were discarded; the filters were washed in 4 ml TMAC wash solution at $52^{\circ} \mathrm{C}$, and this procedure was repeated four times; this series of washes lasts for 30 minutes. The filters were transferred into a plastic box containing 30 ml TMAC solution heated to $56^{\circ} \mathrm{C}$. Multiple membranes hybridised with different probes can be pooled, and washed at $56^{\circ} \mathrm{C}$ for 30 minutes with
occasional agitation. The wash solution was discarded, and the filters were washed briefly with $3 x$ SSC at room temperature (in order to remove any residual TMAC). The filters were blotted on 3MM paper and folded inside Saran Wrap to prevent the filters drying out. The filters were subsequently exposed to X-ray film (Fuji or Kodak) in a light-proof cassette for 4 hours to 2 days at $-80^{\circ} \mathrm{C}$. Damp filters can be stored at $4^{\circ} \mathrm{C}$ to prevent them drying out.

### 5.2.3.8. Removing probes and re-using membranes

To remove previously hybridised radioactively labelled probes, the filters were washed in $4 \times 250 \mathrm{ml} 1 \%(w / v)$ SDS at $100^{\circ} \mathrm{C}$ for a total of 2 to 4 minutes. The stripping solution was discarded; the filters were rinsed in $4 x$ 250 ml of 2 x SSC at room temperature, and either stored damp or re-used immediately for another hybridisation.

### 5.2.4. Sexing genomic $D N A$ samples

A number of DNA samples of unknown sex were received, and these were sexed by typing with the YAP ASO assay. The YAP marker had previously been typed via a gel-based assay (see Chapter 3), but here was converted to an ASO assay (see Figure 5.7). An example of such genomic DNA samples were the Japanese, supplied by Keiji Tamaki, (Sapporo, Japan). A water negative control, plus two female samples were also included in a 96 -well Thermowell $M$ plate (Advanced Biotechnologies) used in PCR analysis on a MJR PTC-200 thermal cycler. For the results of the YAP ASO analysis, see Figure 5.8.

### 5.2.5. $Y$ chromosome specific microsatellite analysis

A total of 17 microsatellites were analysed and these included some routinely used, previously published and some newly identified microsatellites.

### 5.2.5.1. Microsatellite multiplex systems

Three independent microsatellite multiplex assays were carried out on all 402 YAP+ chromosomes. All three multiplex systems were set up and validated by Elena Bosch (University of Leicester). The first multiplex (Ayub et al., 2000) entitled 'CTS' includes the following six Y-chromosome specific microsatellites: DYS434, DYS435, DYS436, DYS437, DYS438, and DYS439 (Ayub et al., 2000). The second multiplex system entitled 'MS1' (Thomas et al., 1999) includes the following six Y-chromosome specific


## B

## 100bp (ancestral allele - no Alu insert)



## 109bp (derived allele - Alu insert)

Figure 5.7. The design of YAP primers and relative locations of ASOs (A) a schematic drawing showing the location of primers and ASOs used during the YAP assay. (B) shows the Y-chromosomal DNA sequence used to generate new YAP primers. A set of three primers were designed to amplify both alleles (YAP- does not have the Alu insertion, and YAP+ does have the Alu insertion). The YAP insert is indicated as a grey box, with some of the sequence represented. Primer sequences are shown in capital letters. Sequence between primer 1 and the other primers ( 2 and 3 ) is shown as a line. Two differently sized amplicons are generated during this PCR assay, of 100bp and 109bp depending if the Alu insert is present. There was no problem of amplification from other Alu sequences in the genome during this assay.

## A

G allele


A allele


B G allele - derived - YAP+ve chromosomes


C

| OO_Q OQ |  |
| :---: | :---: |
| B |  |
| C | O* Q OOO * Q O * |
| D |  |
| E |  |
| F |  |
| G |  |
| H |  |

G allele
A allele I Positive
hybridisations
No hybridisation
Female control
No DNA control
Empty well

A allele - derived - YAP-ve chromosomes


Figure 5.8. The YAP ASO assay
(A) shows the autoradiographs generated from this assay. (B) shows the schematic diagrams of the autoradiographs. (C) shows both allelic states, the female controls and the no DNA control. The samples that fail to hybridise (white circles) have been taken to be female samples.
microsatellites: DYS19, DYS388, DYS390 (Kayser et al., 1997), DYS391, DYS392 (Thomas et al., 1999), and DYS393 (Kayser et al., 1997). The third multiplex system was designed by Elena Bosch (University of Leicester), is entitled 'EBF' and includes the following five Y-chromosome specific microsatellites: DYS389I/II (Kayser et al., 1997), YA7.1, YA7.2 (White et al., 1999), G09411 (Peter de Knijff, unpublished data), as well as the Amelogenin gene sex test assay (Sullivan et al., 1993). See Table 5.1, for the repeat type, allelic length ranges, primer sequences, and the concentration of each fluorescently labelled primer pair used in the multiplex PCR reactions.

### 5.2.5.2. Microsatellite amplification protocol

The following PCR master mix and protocol was used to amplify all the multiplexes (Thomas et al., 1999; Ayub et al., 2000): 1x Super Taq Buffer (HT Biotechnology Ltd.) ( 10 mM Tris- $\mathrm{HCl} \mathrm{pH} 9.0,1.5 \mathrm{mM} \mathrm{MgCl} 2,50 \mathrm{mM} \mathrm{KCl}, 0.1 \%$ ( $\mathrm{v} / \mathrm{v}$ ) Triton $\mathrm{X}-100,0.01 \% ~(\mathrm{w} / \mathrm{v})$ stabiliser). An additional $0.7 \mathrm{mM} \mathrm{MgCl} \mathrm{m}_{2}$ was added to the $1 \times$ Super Taq Buffer, as well as $300 \mu \mathrm{M}$ dNTPs, 0.13 U Super Taq (HT Biotechnology Ltd.), and $0.375 \mu \mathrm{~g}$ TaqStart Antibody (Clontech). The Super TAQ was incubated with the TaqStart Antibody and TaqStart dilution buffer, for between 5 to 7 minutes at room temperature, before its inclusion in the PCR master mix. The final volume of each PCR reaction was $10 \mu \mathrm{l}$, and the PCR reactions were carried out in 96 -well Thermowell M plates (Advanced Biotechnologies) in an MJR PTC-200 thermal cycler. The amplified products were stored at $4^{\circ} \mathrm{C}$ until they were ready to be analysed. The following PCR conditions were used to amplify the 'CTS' multiplex (Ayub et al., 2000):
$1-94^{\circ} \mathrm{C}$ for 2 minutes
$2-94^{\circ} \mathrm{C}$ for 1 minute
$3-60^{\circ} \mathrm{C}$ for 1 minute $\left(-0.5^{\circ} \mathrm{C}\right.$ each cycle)
4-72 ${ }^{\circ} \mathrm{C}$ for 1 minute
5 - Go to 2 ( $x 8$ times)
$6-94^{\circ} \mathrm{C}$ for 1 minute
$7-56^{\circ} \mathrm{C}$ for 1 minute
8-72 ${ }^{\circ} \mathrm{C}$ for 1 minute
9 - Go to 6 (x30 times)
$10-72^{\circ} \mathrm{C}$ for 5 minutes

Table 5.1. Showing the information on all microsatellite loci amplified during this analysis.

| Loci | Multiplex | Repeat type | No. repeats | Size range (bp) | Final primer conc. $(\mu \mathrm{M})$ | 5' dye label | Primer sequence ( $5^{\prime}-3^{\prime}$ ) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| DYS434 | CTS | CTAT | 8-11 | 110-122 | 0.2 | TET | CACTCCCTGAGTGCTGGATT GGAGATGAATGAATGGATGGA |
|  |  |  |  |  |  |  |  |
| DYS435 | CTS | TGGA | 11-13 | 220-228 | 0.05 | TET | agCatctccacacagcacac |
|  |  |  |  |  | 0.05 |  | тTСТСТСТСССССТССТСТС |
| DYS436 | CTS | GTT | 10-15 | 128-143 | 0.1 | FAM | CCAGGAGAGCACACACAAAA |
|  |  |  |  |  | 0.1 |  | GCAATCCAACTTCAGCCAAT |
| DYS437 | CTS | TCAT | 8,2,4-12,2,4 | 186-202 | 0.1 | HEX | GACTATGGGCGTGAGTGCAT |
|  | CTS |  |  |  | 0.1 |  | AGACCCTGTCATTCACAGATGA |
| DYS438 |  | TTTTC | 6-12 | 203-233 | 0.25 | HEX | TGGGGAATAGTTGAACGGTAA |
|  |  |  |  |  | 0.25 |  | GTGGCAGACGCCTATAATCC TCCTGAATGGTACTTCCTAGGTTT GCCTGGCTTGGAATTCTTTT |
| DYS439 |  | GATA | 9-14 | 238-258 | $\begin{aligned} & 0.2 \\ & 0.2 \end{aligned}$ | TET |  |
|  |  |  |  |  |  |  |  |
| DYS19 | MS1 | CTAT/C | 10-19 | 173-209 | 0.35 | TET | ACTGAGTTTATGTTATAGTGTTTTT |
|  |  |  |  |  | 0.35 |  | ATGGCATGTAGTGAGGACA |
| DYS388 | MS1 | ATA | 10-16 | 123-141 | 0.25 | TET | GTGAGTTAGCCGTTTAGCGA |
|  |  |  |  |  | 0.25 |  | CAGATCGCAACCACTGCG |
| DYS390 | MS1 | CTG/AT | 18-27 | 192-228 | 0.2 |  | TATATTTTACACATTTTTGGGCC |
|  |  |  |  |  | 0.2 | FAM | TGACAGTAAAATGAACACATTGC |
| DYS391 | MS1 | CTAT | 8-13 | 156-176 | 0.23 | FAM | CTATTCATTCAATCATACACCCATAT |
|  |  |  |  |  | 0.23 |  | ACATAGCCAAATATCTCCTGGG |
| DYS392 | MS1 | ATT | 7-16 | 155-182 | 0.35 |  | AAAAGCCAAGAAGGAAAACAAA |
|  |  |  |  |  | 0.35 | HEX | CAGTCAAAGTGGAAAGTAGTCTGG |
| DYS393 | MS1 | GATA | 9-15 | 107-131 | 0.15 | HEX | GTGGTCTTCTACTTGTGTCAATAC |
|  |  |  |  |  | 0.15 | HEX | AACTCAAGTCCAAAAAATGAGG |


| Loci | Multiplex | Repeat type | No. repeats | Size range (bp) | Final primer conc. ( $\mu \mathrm{M}$ ) | 5' dye label | Primer sequence ( $5^{\prime}-3^{\prime}$ ) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| DYS389I <br> /II | EBF | CTG/AT | $\begin{aligned} & \mathrm{I}=10-16 \\ & \mathrm{II}=26-34 \end{aligned}$ | $\mathrm{I}=142-166$ | 0.35 | FAM | CCAACTCTCATCTGTATTATCTATGT |
|  |  |  |  | $\mathrm{II}=258-300$ | 0.35 |  | CCTGAGTAGCAGAAGAATGTCATA |
| YA7.1 | EBF | GATA | 9-13 | 170-186 | 0.35 | TET | GCCAAACTCTTTCCAAGAAG <br> TCATCTATCCTCTGCCTATCATT |
|  |  |  |  |  | 0.35 |  |  |
| YA7.2 | EBF | GATA | 8-12 | 174-190 | 0.25 | FAM | aGGCAGAGGATAGATGATATGGAT <br> tTCAGGTAAATCTGTCCAGTAGTGA |
|  |  |  |  |  | 0.25 |  |  |
| G09411 | EBF | TGTA | 10-14 | 179-199 | 0.3 | HEX | TTCAGGTAAATCTGTCCAGTAGTGA tGTGCTGTACCAGTTGCCTA |
|  |  |  |  |  | 0.3 |  | CCAGCCTGAGCAAGAGAGTA |
| AMEL | EBF | na | na | $\begin{aligned} & \mathrm{X}-106 \\ & \mathrm{Y}-112 \end{aligned}$ | 0.15 | HEX | CCCTGGGCTCTGTAAAGAATAGTG ATCAGAGCTTAAACTGGGAAGCTG |
|  |  |  |  |  | 0.15 |  |  |

Note: No. repeats - number of repeats observed within the microsatellites. Final primer conc. - is the actual final concentration $(\mu \mathrm{M})$ for each primer within the PCR assay. The three multiplex systems are indicated by initials, CTS (Ayub et al., 2000), MS1 (Thomas et al., 1999), and EBF (Bosch, unpublished) even though some of the loci being amplified have been published - DYS389I/II (Kayser et al., 1997), YA7.2 (White et al., 1999). An unpublished locus (G09411) was also included (Peter de Knijff, Leiden, The Netherlands). AMEL is not a microsatellite, but has been incorporated into the EBF multiplex to allow sexing and the detection of AMELY gene deletions.

The following PCR conditions were used to amplify the 'MS1' multiplex and have been modified from the original proposed conditions (Thomas et al., 1999):
$1-95^{\circ} \mathrm{C}$ for 2 minutes
$2-94^{\circ} \mathrm{C}$ for 1 minute
$3-57.5^{\circ} \mathrm{C}$ for 1 minute $\left(-0.5^{\circ} \mathrm{C}\right.$ each cycle $)$
$4-72^{\circ} \mathrm{C}$ for 1 minute
$5-\mathrm{Go}$ to $2(x 5$ times $)$
$6-94^{\circ} \mathrm{C}$ for 1 minute
$7-55^{\circ} \mathrm{C}$ for 1 minute
$8-60^{\circ} \mathrm{C}$ for 1 minute
$9-\mathrm{Go}$ to $6(\times 32$ times $)$
$10-72^{\circ} \mathrm{C}$ for 10 minutes

The following PCR conditions were used to amplify the 'EBF' multiplex (Bosch, unpublished data):

$$
\begin{aligned}
& 1-95^{\circ} \mathrm{C} \text { for } 5 \text { minutes } \\
& 2-94^{\circ} \mathrm{C} \text { for } 1 \text { minute } \\
& 3-60^{\circ} \mathrm{C} \text { for } 1 \text { minute }\left(-0.5^{\circ} \mathrm{C} \text { each cycle }\right) \\
& 4-72^{\circ} \mathrm{C} \text { for } 1 \text { minute } \\
& 5-\mathrm{Go} \text { to } 2(\times 10 \text { times }) \\
& 6-94^{\circ} \mathrm{C} \text { for } 1 \text { minute } \\
& 7-55^{\circ} \mathrm{C} \text { for } 1 \text { minute } \\
& 8-72^{\circ} \mathrm{C} \text { for } 1 \text { minute } \\
& 9-\mathrm{Go} \text { to } 6(\times 28 \text { times }) \\
& 10-72^{\circ} \mathrm{C} \text { for } 10 \text { minutes }
\end{aligned}
$$

### 5.2.5.3. Polyacrylamide gel preparation

The glass plates are rinsed in distilled water, unless they appear greasy when a detergent Alconox may be used. White Kimwipes Lite (Kimberly-Clark) must be used to dry the glass plates, to avoid contaminating the plates with coloured fibres. The gel spacers are also cleaned in the same manner, and a few drops of distilled water are placed on to the spacers, to help secure both plates. The following reagents were used when preparing the polyacrylamide gel (Ultrapure Sequagel sequencing system, National Diagnostics).

| Concentrate | 6 ml |
| :--- | :--- |
| Buffer | 2.5 ml |
| Diluent | 16.5 ml |
| $10 \%(\mathrm{w} / \mathrm{v})$ APS | $200 \mu \mathrm{l}^{*}$ |
| Temed | $30 \mu \mathrm{l}$ |

* The APS (Ammonium persulphate) was made up weekly, and stored at $4^{\circ} \mathrm{C}$.

The above reagents were mixed thoroughly and the acrylamide was introduced between the two glass plates along the top edges. Air bubbles introduced could be removed using an additional spacer, and bulldog clips were used to secure the plates and the comb. The gel was allowed to set at room temperature for 1 to 2 hours; the comb was removed, and the plates were re-washed in distilled water, and wiped dry. Care must be taken to ensure that are no visible marks on the plates in particular around where the laser scans across the gel.

### 5.2.5.4. Preparation of amplified DNA products

A $2 \mu \mathrm{l}$ aliquot of each amplified DNA product was diluted in $8 \mu \mathrm{l}$ of sterile distilled Elga water, and these diluted samples were pipetted into another 96-well plate (Advanced Biotechnologies). The GENESCAN-500 ${ }^{\mathrm{TM}}$ TAMRA internal lane size standard (Perkin-Elmer) was prepared ( $150 \mu \mathrm{l}$ of formamide, $40 \mu \mathrm{l}$ of special loading buffer, $30 \mu \mathrm{l}$ of the TAMRA standard) and this standard allows the size estimation of fragments ranging from 35 to 500 bp . A $2 \mu \mathrm{l}$ aliquot of the standard mixture was pipetted into a new 96-well plate (Advanced Biotechnologies) that had each well labelled ( 1 to 96 - to aid the loading process), and to this plate $1 \mu$ l of each diluted DNA sample was added. The samples were added to this plate in the order they were amplified and diluted, to avoid any problems with mixing up the sample order.

### 5.2.5.5. Electrophoresis of the amplified DNA products

The samples were electrophoresed on an ABI377 DNA sequencer for approximately 2 hours, and the microsatellite data were collected using the GENESCAN software (Perkin-Elmer). The gel was secured into the machine, and the plates were pre-checked with the laser to ensure no smears would interfere with the fluorescence readings. If strong fluorescence was detected, the plates were re-cleaned and re-checked. Freshly prepared 1x TBE buffer was used in both top and bottom reservoirs, and the gel was pre-run until a temperature of $51^{\circ} \mathrm{C}$ was reached. During the pre-warming stage the TAMRA standard and diluted amplified products mix can be heat denatured $\left(96^{\circ} \mathrm{C}\right.$ for 2 minutes and cooled on ice). Following this stage, the wells were washed out and $0.55 \mu \mathrm{l}$ of each sample was loaded. All the odd-numbered samples were loaded first, and the program was allowed to run for 2 minutes; paused and all the even-numbered samples were loaded. The program was started and left to run for approximately 2 hours. The standard bands were visible on the computer screen and this allowed a check if all loci had been scanned. The samples were sized using GENESCAN ${ }^{\text {TM }}$ software (Perkin-Elmer), and the alleles were assigned sizes using the GenoTyper ${ }^{\mathrm{TM}}$ software (PE Biosystems). For an example of a multiplex microsatellite gel, see Figure 5.9.


Figure 5.9. An example of a polyacrylamide gel run on the ABI Genescan machine analysing $Y$ chromosome specific microsatellites This multiplex was designed by Ayub et al., 2000, and includes 6 different loci: DYS434, DYS435, DYS436, DYS437, DYS438, and DYS439. The relative position and size for each locus is indicated on the above gel, as well as the size standard (Genescan-500 Tamra, Perkin Elmer). The samples appear staggered because they was a 2 minute delay when loading half of the samples.

### 5.3. Results

In order to investigate the phylogeography of YAP+ sublineages, it was first necessary to analyse some new biallelic markers (published and unpublished) and position them within the maximum parsimony tree.

### 5.3.1. Biallelic and allele-specific oligonucleotide analyses

Data determined from PCR-RFLP analysis were available for a number of samples, but a series of new ASO markers was also developed and analysed. This ASO protocol was more sensitive, required less DNA, and appeared more reliable than the existing RFLP detection method. One problem with RFLP detection was the digestion process; if the DNA samples were of poor quality or degraded they could generate partial data. For an example of a new ASO marker (M78) where successful allele specific discrimination was observed through the complementary hybridisation pattern detected, see Figure 5.10 (this shows the results in the 96 -well plate format to make the comparison between both alleles easier).

### 5.3.2. The positioning of new markers on the maximum parsimony tree

Once the allelic states were determined (Table 5.2) for all of the new ASO markers (M15, M78, M81, M33, DYS391, PN1, PN2, SRY-8299 and YAP), a number of additional haplogroups were generated which allowed the YAP lineage to be further sub-divided. Previously only three haplogroups were contained within the YAP branch of the tree, hg 4, and its derived sub-groups hgs 21 and 8. For an example of new marker (M78) data, and how this was used to position it within the existing maximum parsimony tree, see Figure 5.11. This marker is informative in sub-dividing an existing haplogroup (hg 21), and it was possible to separate European and non-European samples. All the European samples examined had the derived T allele, whereas all the non-European samples had the ancestral $C$ allele.

A
T allele


C allele


B
T allele

No hybridisation
Positive hybridisation

- Ambiguous result

Empty well

C allele


Figure 5.10. The M78 ASO assay
Autoradiograph and schematic diagrams to show ASO (allele specific oligonucleotide) results for a new unpublished marker, M78. The autoradiographs show the actual complementary pattern generated for the different alleles, and this is simplified in the two diagrams on the right, one for each allele ( T and C ). The diagrams show the format of the 96 -well plate, with the number and letter of each row indicated, but in the experiment only a portion of the plate was used.

A
B
Haplogroups of interest


Figure 5.11. The location of new marker M78 within the tree
(A) shows the maximum parsimony tree, with the analysed haplogroups being highlighted in yellow. (B) This section of the tree indicates where the marker M78 is positioned with relation the existing markers. This new markers splits the haplogroup 21 lineage, and a clear distinction between European and non-European samples tested is observed. The European samples have derived Talleles, and the nonEuropeans have the ancestral C allele.

Table 5.2. Eleven-marker haplotypes within YAP+ sublineages, including some new allele-specific markers.

|  | 12f2 | YAP | M15 | SRY-8299 | M33 | DYS391 | PN2 | M78 | M81 | PN1 | sy81 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| hg4 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| hg30 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| hg34 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| hg 21 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| hg 31 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| hg 25 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| hg32 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| hg33 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| hg 8 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

Note: 0 indicates an ancestral allele, whereas 1 indicates a derived allele state. The haplogroups are indicated down the side of the table and represented by hg.

The specific order of all the markers could not be determined absolutely due to absence of some intermediate haplogroups. For example, even though a total of 287 samples were examined using all of the markers described; the order of PN2 and DYS391 could not be determined, because none of the samples have an intermediate haplogroup - all samples were either ancestral, or derived for both of these mutations. This unobserved haplogroup may be absent due to sampling (but the sample distribution within these samples covers three continents), or possibly has gone extinct. The inclusion of sY81 was to test the proposal that PN1 was concordant with it, and these data supported that, and both markers have been included on the same branch of the tree diagram.

The 12f2 PCR marker had previously been used by Matt Hurles (unpublished observations) to examine a collection of global DNA samples known as the YCC (Y Chromosome Consortium) collection from Mike Hammer (Arizona, USA). It was discovered that this marker was deleted in a Japanese YAP+ chromosome (haplogroup 3G (Hammer et al., 1997) equivalent to hg 4), and that is why this marker was incorporated into this analysis. The detection of a recurrent mutation (either YAP or 12f2) (Table 5.2) during these experiments (Figure 5.12) added a potential reticulation in the new tree, which could be seen as a problem. There were four possible ways in which the recurrent mutation could be placed on the tree (see Figure 5.13). As discussed, it is extremely unlikely that the well characterised YAP polymorphism (Hammer, 1994; Hammer and Bonner, 1994; Hammer and Horai, 1995; Hammer and Zegura, 1996; Hammer et al., 1997; Hammer et al., 1998) is recurrent, and is more likely that 12 f 2 is recurrent. This hypothesis has been supported by a total of 11 new

## A $\mathbf{2 . 5} \%$ agarose gel showing the 12 f 2 short amplicon on $44 \mathrm{YAP}+$ chromosomes

| Top row (right - left) |  |  |
| :---: | :---: | :---: |
| 1. øxHaeIII | 14.832* |  |
| 2. m219 | 15. 849 |  |
| 3. ms 42 * | 16. 883 * |  |
| 4. Slov73 | 17.911 |  |
| 5. 412 | 18. 913 |  |
| 6. 426 | 19.931* |  |
| 7. 432 | 20. 959 |  |
| 8. 435 | 21. Tib1 | \# E |
| 9. 439 * | 22. Tib19 |  |
| 10.447* | 23. Tib23 |  |
| 11. 452 | 24. Tib24 |  |
| 12. 457 * | 25. Tib30 |  |
| 13. 458 * | 26. øxHaeIII | control |

all samples hg 4

Bottom row (right - left)

1. øxHaeIII 13. Tib6
2. Tib31 14. Tib9
3. Tib35 15. Tib15
4. HLZ14 16. Tib32
5. HLZ29 17. Tib34
6. QIA19 18. MAN27
7. QIA21 19. QIA29
8. QIA34 20. QIA32
9. Hui21 21. HUI10
10. Hui23 22. m40 (hg 9)
11. YGX16 23. water (-ve)
12. Tib3 24. øxHaeIII

lanes $2-10=\mathrm{hg} 4$
lanes 11-21 = hg 30

B
2.5\% agarose gel showing the 12 f 2 short amplicon on 4 hg 21 chromosomes


Figure 5.12. The identification of the $\mathbf{1 2 f} \mathbf{2}$ recurrent mutation
(A) shows the 12 f 2 small amplicon assay, on 44 YAP+ chromosomes, and some appear to be hg 9 (see asterisk). This work was carried out after Matt Hurles first indicated this may be a recurrent mutation. All the hg 21 chromosomes were analysed with the 12 f 2 small amplicon assay, to see if the recurrent mutation was branching off hg 21 or hg 4 (B).


If the ancestral state was not known for the YAP mutation, this may be the best way to position the recurrent mutation, with a reticulation. But YAP is well characterised at the molecular level.


This suggests that the (YAP) inserted during a migration out of Africa into Asia, followed by the 12 f 2 mutation and by the removal of the Alu element. No known mechanism allows the removal of such (Alu) elements, and this scenario is unlikely.


This suggests that the YAP mutation is recurrer and has occurred in two geographically distinc continents, e.g. Asia and Europe. Haplogroup is very common around the Mediterranean basir and it is unlikely this is where a 2nd YAP insertio occurred, since hg 34 is Asian-specific.


It is most likely that the 12 f 2 mutation i recurrent in the population. It is much les well characterised than the YAP polymorphism. Th geographical association also fits well, i.e. that deletion/insertion occurred in Japan, and another iv Europe. It does not rely on unlikely migrations

Figure 5.13 . Four possible ways the $\mathbf{1 2 f} \mathbf{2}$ polymorphism could be included within the tree Each (A-D) is discussed below the diagram and (D) is the most likely.
markers (Underhill and Hammer, unpublished data), because there are 3 markers (M89, PN14, 12f2) between hg 9 and hg 2, whereas only 12f2 is between hg 4 and hg 34, compared to 8 markers (YAP, M145, M174, M55, M57, M64, PN12, and PN42) between hg 2 and hg 4.

The allelic data generated can be incorporated into the existing maximum parsimony tree, to show where the new haplogroups are located (see Figure 5.14).

### 5.3.3. Continental specificity of the $Y$-chromosomal haplogroup data

For a complete list of the data for all the samples tested for the biallelic and ASO assays see Appendix 2, but for a continental breakdown of data see Table 5.3. These YAP+ data can be pooled for each continental sample (as done for Figure 5.1) and superimposed onto a global landscape, see Figure 5.15. The three continental samples encompass 402 chromosomes, and appear very different. The number of haplogroups observed varied between the three continental samples; within Africa, Europe and Asia, a total of six, four and all nine haplogroups were detected respectively. Data divided into individual lineages can be pooled into a number of distinct inter-continental groups (Table 5.4), and superimposed onto the map to see the continental specificity of certain haplogroups, see Figure 5.16. These data have been shown graphically (see Figure 5.17), to enable different patterns of haplogroup diversity to be examined within a number of different populations.

Relative diversity (h values - see Table 5.4) of continental samples show Asia to be more diverse than Africa, which is more diverse than Europe. Asia being the most diverse is unexpected, because conventionally Africans are considered the most diverse populations and the place anatomically modern humans originated. The reduced African diversity is probably due to the Bantu expansion, or the increased Asian diversity could reflect the 'Out of Africa and back again' hypothesis - indicating an Asian origin of the YAP lineage (Altheide and Hammer, 1997). The examination of sub-continental $h$ values shows that within Africa: western populations are the most diverse followed by North Africans, then east Africans. The increased diversity within western Africans may be due to undocumented Atlantic influences, or it may suggest the Bantu expansion had little influence on populations within this region. Little gene flow


Figure 5.14. The maximum parsimony tree including all new markers
The new maximum parsimony tree (B) as compared to the existing tree (A). The original haplogroups of interest are highlighted in yellow. Each circle represents a different haplogroup, and the direction of the arrow shows the direction of the mutation. (0) represents ancestral and (1) derived alleles. The fine lines represent some markers that are not being analysed during this work.

Table 5.3. The population breakdown of $402 \mathrm{YAP}+$ chromosomes distributed across three different continents.

| Geog. location | hg 4 | hg 30 | hg 34 | hg 21 | hg 31 | hg 25 | hg 32 | hg 33 | hg 8 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| North Africa | 0 | 0 | 0 | 0 | 3 | 7 | 11 | 113 | 9 |
| Sub-Sah. Africa | 0 | 0 | 0 | 3 | 6 | 3 | 1 | 1 | 45 |
| East Bantu | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 3 |
| West Bantu | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Bantu | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| Kenyan | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 10 |
| Zimbabwean | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 25 |
| Gambian | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 3 |
| Togolese | 0 | 0 | 0 | 0 | 5 | 1 | 0 | 0 | 0 |
| Nigerian | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Mbuti | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| European | 0 | 0 | 0 | 1 | 0 | 31 | 88 | 9 | 0 |
| Greenlandic | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Swedish | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Finnish | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Bul non-gyp | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Bul gyp | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 |
| Yugoslavian | 0 | 0 | 0 | 0 | 0 | 1 | 10 | 0 | 0 |
| Turkish | 0 | 0 | 0 | 0 | 0 | 3 | 8 | 0 | 0 |
| Belarussian | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 |
| Chuvash | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Romanian | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 |
| Greek | 0 | 0 | 0 | 0 | 0 | 0 | 8 | 2 | 0 |
| Greek Cypriot | 0 | 0 | 0 | 0 | 0 | 5 | 7 | 0 | 0 |
| Italian | 0 | 0 | 0 | 0 | 0 | 3 | 12 | 0 | 0 |
| Sardinian | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| English | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0 | 0 |
| Irish | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Dutch | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| French | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 |
| Spanish | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 5 | 0 |
| Bavarian | 0 | 0 | 0 | 0 | 0 | 2 | 4 | 0 | 0 |
| Ladin (Cortina) | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 |
| Ladin (Other) | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 |
| Slovenian | 0 | 0 | 0 | 1 | 0 | 0 | 4 | 0 | 0 |
| Euro. Jewish | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Asia | 24 | 11 | 8 | 1 | 0 | 15 | 5 | 1 | 6 |
| Mongolian | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Japan | 9 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 |
| India+ Pakistan | 0 | 0 | 0 | 1 | 0 | 13 | 2 | 0 | 6 |
| Lebanese | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| China | 14 | 11 | 0 | 0 | 0 | 1 | 3 | 0 | 0 |
| Total | 24 | 11 | 8 | 5 | 9 | 56 | 105 | 124 | 60 |

Note: The Euro. Jewish sample is a chromosome from an unspecified geographic location within Europe, but has been included here. Bul non-gyp and Bul. gyp describe Bulgarian gypsy and non-gypsy samples. Two of the continental samples (Europe and Asia) have been further sub-divided into separate populations.


Figure 5.15. The frequency distribution of the YAP lineages across three continents
(A) Graphically represents the different distributions of haplogroup frequencies across the three continents (Africa, Europe and Asia). The corresponding colours for each haplogroup are shown in the key and their location is indicated in the tree (B). The tree shows the direction of each mutation, by the use of an arrow, and each marker is named under an arrow. The pie charts are not to scale and their actual location within each continent is arbitrary.



## Sample division

Japan $=17$
China $=29$
Other Asian = 25
N. Africa $=143$

Sub-Sah Africa $=59$
$\mathrm{N}+\mathrm{W}$. Europe $=27$
C. Europe $=25$
$S+$ E. Europe $=76$

Figure 5.16. Haplogroup frequencies within the YAP + lineage
(A) Graphically represents the different distributions of haplogroup frequencies across the three continents (Africa, Europe and Asia). The corresponding colours for each haplogroup are shown in the key and their location is indicated in the tree (B). The tree shows the direction of each mutation, by the use of an arrow, and each marker is named under an arrow. The pie charts are not to scale and their actual location within each continent is arbitrary.


Figure 5.17. Population sub-division within the YAP+ lineages
(A) Individual population samples have been superimposed onto a map to show any patterns detectable within the sub-divided data. The size of the circle is proportional to the number of chromosomes, and each haplogroup has been given a specific colour (B). Some populations could not be assigned to specific locations (e.g. West Bantu) and these are shown in white boxes in the Sea next to the apprpriate continents.
is observed between North Africa and the rest of Africa; due to the difference in hg 33 chromosomes, and the Sahara desert probably acts as the barrier.

Within Asia, the Chinese appear more diverse than the Japanese, and this may reflect the isolation of the Japanese islands with respect to influences from external admixture. Within Europe, north and western ( $\mathrm{N}+\mathrm{W}$ ) populations are the most diverse followed by central (C) then southern and eastern (S+E) populations (see Table 5.4). If YAP+ chromosomes were spread during the Neolithic, you would expect $S+E$ to be more diverse than $C$, who would be more diverse than $N+W$ populations. However, if some of the YAP+ chromosomes were Palaeolithic, and others were Neolithic, then the expected diversity would be ( $\mathrm{N}+\mathrm{W}>\mathrm{C}>\mathrm{S}+\mathrm{E}$ ) as observed here. The diversity within these geographically affiliated samples can be examined further using microsatellite loci (see later sections).

The three continental samples have been further sub-divided (Figure 5.16), for example, Africa has been divided into North Africa and Sub-Saharan Africa, although the Sub-Saharan African samples have been further sub-divided into three, east, west and other samples. Europe has been split into three groups, north and western ( $\mathrm{N}+\mathrm{W}$ ) Europe, central, and southern and eastern ( $\mathrm{S}+\mathrm{E}$ ) Europe. Finally, Asia has been again sub-divided into three groups, Japan, China, and other Asian (including Mongolian, Indian and Pakistan). The North African samples were not sub-divided. The total number of chromosomes described in these data is 401, because the European Jewish sample was removed because of an unknown geographical affiliation within the European continent. General observations include continental and population specific lineages, for example, hg 4 is only detected in Asia, and hg 34 and hg 30 appear to be Japanese and Chinese specific respectively.

Nei's unbiased estimator of diversity was calculated for the sub-grouped populations using the following equation $\left[\mathrm{h}=1-\Sigma\left(\mathrm{x}^{2}\right)\right]$ (Nei, 1978). How the values for diversity were generated is given in an example below.

## European (N+W) samples, $\mathbf{n = 2 7}$

$$
\begin{gathered}
{[5 / 27]^{2}+[16 / 27]^{2}+[3 / 27]^{2}=(0.0343+0.3512+0.0123)=0.398} \\
1-0.398=0.602, \text { so } \mathbf{h}=\mathbf{0 . 6 0 2}
\end{gathered}
$$

The standard errors were also calculated because the sample sizes were different; the equation used was as follows: $1 / n(2 n-1) x\left[\sum x^{2}-\left(\sum x^{2}\right)^{2}+4(n-1)\left\{\sum x^{3}-\right.\right.$ $\left.\left.\left(\sum x^{2}\right)^{2}\right]\right]$ (Nei, 1978). Where $n=$ sample size, and $x^{2}$ was as before. An example is shown below:

$$
\begin{gathered}
\text { European (N+W) samples, } \mathbf{n = 2 7} \\
1 / 27(54-1) \times\left[0.398-(0.398)^{2}+4(27-1)\left\{0.251-(0.398)^{2}\right\}\right]=(1 / 1431) \times[0.24+104(0.093)] \\
=1 / 1431(9.912), \text { so } \mathbf{S E}= \pm \mathbf{0 . 0 0 6 9 3}
\end{gathered}
$$

The diversity within each continental sub-group was calculated using Nei's estimator of diversity equation (above), and all the values can be seen in Table 5.4. A clear difference in the distribution of the haplogroups is apparent when examining Figure 5.16. The diversity values indicate the east Sub-Saharan Africans have the lowest $(\mathrm{h}=0.104)$ and the west Sub-Saharan Africans have the highest ( $\mathrm{h}=0.673$ ).

These data (Table 5.4) comprising $401 \mathrm{YAP}+$ chromosomes can be superimposed on to a specific branch of the maximum parsimony tree, to illustrate how the lineages segregate geographically, see Figure 5.18. Each lineage will be briefly described in turn. The three Asian-specific lineages (hg's 4, 30 and 34) are at the more ancestral region of the branch. However, the population specificity of hgs 30 and 34 probably represents local Asian diversification. Haplogroup 21 chromosomes defined only by the SRY-8299 mutation (Whitfield et al., 1995a), are poorly represented within the 401 YAP+ samples, with only five examples being detected. Haplogroup 31 defined by M33 (Underhill, unpublished) is African-specific, and this specificity suggests the defining mutation probably arose within Africa. The globally distributed hg 25 defined by DYS391 (de Knijff, unpublished) and PN2 (Hammer et al., 1997) is only absent within Japan, and this haplogroup may represent an expansion from Africa into the rest of the world.

At the more derived end of the branch the geographic distribution varies, with hg 32 defined by M78 (Underhill, unpublished) being common (61 chromosomes) within Southern and Eastern European samples, but again absent from Japan. This may reflect the European Neolithic expansion of chromosomes

Table 5.4. The continental breakdown and subsequent grouping of Sub-Saharan African, Asian and European haplogroup data on the 401 YAP + chromosomes analysed.

|  | Hg 4 | Hg30 | Hg34 | Hg21 | Hg31 | Hg25 | Hg32 | Hg33 | Hg8 | Total | $\mathrm{h}=$ | SE |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| N. Africa | 0 | 0 | 0 | 0 | 3 (2) | 7 (5) | 11 (8) | 113 (79) | 9 (6) | 143 | 0.363 | $\pm 0.00349$ |
| Sub-Saharan Africa |  |  |  |  |  |  |  |  |  |  | $0.4 \pm \pm 0.00356$ |  |
| East | 0 | 0 | 0 | 0 | 0 | 1 (3) | 1 (3) | 0 | 35 (94) | 37 | 0.104 | $\pm 0.00243$ |
| West | 0 | 0 | 0 | 1 (7) | 6 (43) | 1 (7) | 0 | 1 (7) | 5 (36) | 14 | 0.673 | $\pm 0.01159$ |
| Other | 0 | 0 | 0 | 2 (25) | 0 | 1 (12) | 0 | 0 | 5 (63) | 8 | 0.531 | $\pm 0.02564$ |
| Asia |  |  |  |  |  |  |  |  |  |  | 0.792 | $\pm 0.00147$ |
| Japan | 9 (53) | 0 | 8 (47) | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 0.498 | $\pm 0.02539$ |
| China | 14 (48) | 11 (38) | 0 | 0 | 0 | 1 (3) | 3 (11) | 0 | 0 | 29 | 0.611 | $\pm 0.00638$ |
| Other | 1 (4) | 0 | 0 | 1 (4) | 0 | 14 (56) | 2 (8) | 1 (4) | 6 (24) | 25 | 0.382 | $\pm 0.00725$ |
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| Europe 0.289 |  |  |  |  |  |  |  |  |  |  |  | $\pm 0.00158$ |
| N+W | 0 | 0 | 0 | 0 | 0 | 5 (19) | 16 (59) | 6 (22) | 0 | 27 | 0.602 | $\pm 0.00693$ |
| S + E | 0 | 0 | 0 | 0 | 0 | 12 (16) | 61 (80) | 3 (4) | 0 | 76 | 0.329 | $\pm 0.00263$ |
| Central | 0 | 0 | 0 | 1 (4) | 0 | 13 (52) | 11 (44) | 0 | 0 | 25 | 0.534 | $\pm 0.00812$ |

Note: Actual numbers of chromosomes in each haplogroup followed by the calculated percentage in brackets. East Sub-Saharan African include Zimbabwean and Kenyan, West include Gambian, Togolese, and Nigerian. Others include Bantu, East Bantu, West Bantu and Mbuti. N+W. Europe - northern and western European samples, C. Europe - central, and S+E. Europe - southern and eastern European samples. N+W. European samples include Greenlandic Inuit, Swedish, Finnish, E. Anglian, Irish, Dutch, French, Spanish. S+E. European samples include Bulgarian, Serbian, Turkish, Belarussian, Chuvash, Romanian, Greek, Cypriot, Italian and Sardinian. The h values (Nei's estimator of diversity) and SE (standard error) values were calculated for each sub-grouping.


Figure 5.18. Continental representation of the YAP + chromosomal data
The area of the circle is proportional to the frequency of the haplogroup detected. Three continents are represented by different colours (see the key). The haplogroup numbers are written below each appropriate circle. The direction of the arrow shows the direction of the mutations; $(0)$ is ancestral, and (1) is derived.
from the Near East; further analysis of this lineage may support this hypothesis, if the Middle Eastern chromosomes appear more diverse than the European ones. More than $91 \%$ of North African chromosomes belong to hg 33, defined by M81 (Underhill, unpublished) and is completely absent from Japan, China, and central Europe. It appears likely, that this mutation occurred somewhere within North Africa, before $\sim 20,000$ years ago, and probably defines a major North African Palaeolithic component. It is unlikely that this mutation occurred elsewhere and has had enough time to be removed by effects such as drift.

Finally, the most derived haplogroup (hg 8 defined by sY81 and its phylogenetic equivalent PN1 [Hammer et al., 1997]) predominates within SubSaharan African samples, but is also detected within North African and some Asian populations. It has been previously proposed that hg 8 was a Sub-Saharan specific lineage, and was completely absent from Europe and Asia (Hammer et al., 1998; Karafet et al., 1999), and so in light of these data the Asian examples may reflect recent migrations from Sub-Saharan Africa. Additional mutations are now available and show this haplogroup is at the centre of a star-like structure, indicating an expansion event, possibly the Bantu expansion (Underhill et al., 2000).

### 5.3.4. Microsatellite analysis of $Y A P+$ chromosomes

The 402 YAP+ chromosomes were analysed using 17 Y-specific microsatellites (see materials and methods); some loci failed to amplify consistently well, and so complete and reliable data are only available for 106 samples (Table 5.5).

### 5.3.4.1. Additional microsatellite data

Some microsatellite data were generated previously by Elena Bosch (University of Leicester, England), and Michele Stenico (University of Ferrara, Italy). The generation of complete data for all samples analysed was always attempted; however, some samples failed to give data for certain assays, and that is why the additional data has been incorporated. In other cases a number of assays were repeated and these confirmed previously obtained results. For data from Elena Bosch, see samples highlighted in yellow colour, and data from Michele Stenico see samples highlighted in green, in Appendix 3.

| Sample | Hg | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS389I | DYS389II | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7.1 | YA7. 2 | G09411 | STR hap | Freg |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 412 | 4 | 16 | 12 | 26 | 10 | 11 | 13 | 12 | 29 | 10 | 11 | 12 | 8 | 10 | 11 | 13 | 11 | 12 | 1 | 1 |
| 435 | 4 | 15 | 12 | 25 | 10 | 12 | 13 | 13 | 30 | 10 | 11 | 12 | 8 | 10 | 12 | 13 | 11 | 12 | 2 | 1 |
| Tib1 | 4 | 15 | 12 | 25 | 10 | 7 | 13 | 13 | 29 | 8 | 11 | 12 | 8 | 11 | 12 | 11 | 10 | 12 | 3 | 1 |
| Tib35 | 4 | 15 | 12 | 25 | 10 | 7 | 13 | 13 | 29 | 8 | 11 | 12 | 8 | 11 | 12 | 10 | 10 | 12 | 4 | 1 |
| HLZ14 | 4 | 15 | 12 | 25 | 10 | 7 | 13 | 14 | 30 | 8 | 11 | 12 | 8 | 11 | 12 | 11 | 10 | 12 | 5 | 1 |
| RN62 | 8 | 16 | 12 | 22 | 10 | 11 | 14 | 12 | 29 | 9 | 11 | 12 | 8 | 11 | 14 | 11 | 10 | 12 | 1 | 1 |
| m67 | 8 | 15 | 12 | 21 | 10 | 11 | 13 | 14 | 31 | 9 | 11 | 12 | 8 | 12 | 12 | 10 | 10 | 12 | 2 | 1 |
| m475 | 8 | 15 | 12 | 21 | 10 | 11 | 13 | 13 | 31 | 9 | 11 | 12 | 8 | 11 | 11 | 13 | 11 | 12 | 3 | 1 |
| m719 | 8 | 15 | 12 | 21 | 10 | 11 | 14 | 12 | 30 | 9 | 11 | 12 | 8 | 11 | 11 | 10 | 10 | 13 | 4 | 1 |
| G40 | 8 | 16 | 12 | 21 | 10 | 11 | 14 | 12 | 29 | 9 | 11 | 12 | 8 | 11 | 12 | 10 | 10 | 12 | 5 | 1 |
| RN37 | 8 | 17 | 12 | 21 | 10 | 11 | 15 | 14 | 31 | 9 | 11 | 12 | 8 | 11 | 12 | 11 | 10 | 12 | 6 | 1 |
| G42 | 8 | 17 | 12 | 20 | 9 | 11 | 14 | 13 | 30 | 9 | 11 | 12 | 8 | 11 | 11 | 11 | 11 | 12 | 7 | 1 |
| R36 | 8 | 15 | 12 | 21 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 11 | 11 | 11 | 12 | 12 | 8 | 1 |
| R49 | 8 | 16 | 12 | 21 | 10 | 11 | 15 | 13 | 30 | 9 | 11 | 12 | 8 | 11 | 12 | 11 | 11 | 12 | 9 | 1 |
| R77 | 8 | 16 | 12 | 21 | 11 | 11 | 14 | 13 | 29 | 9 | 11 | 12 | 8 | 11 | 12 | 10 | 11 | 12 | 10 | 1 |
| R79 | 8 | 16 | 12 | 21 | 10 | 11 | 15 | 13 | 30 | 9 | 11 | 12 | 8 | 11 | 13 | 11 | 12 | 12 | 11 | 1 |
| MAK102 | 8 | 15 | 12 | 21 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 11 | 11 | 10 | 11 | 12 | 12 | 1 |
| RAB073 | 8 | 15 | 12 | 21 | 10 | 11 | 14 | 12 | 30 | 9 | 11 | 12 | 8 | 11 | 12 | 10 | 10 | 12 | 13 | 1 |
| RAB106 | 8 | 15 | 12 | 21 | 10 | 11 | 14 | 12 | 30 | 9 | 11 | 12 | 8 | 11 | 12 | 10 | 10 | 12 | 13 | 1 |
| RAB121 | 8 | 16 | 12 | 21 | 10 | 11 | 13 | 14 | 31 | 9 | 11 | 12 | 8 | 11 | 12 | 9 | 11 | 12 | 14 | 1 |
| RAB132 | 8 | 16 | 12 | 21 | 10 | 11 | 14 | 13 | 31 | 9 | 11 | 12 | 8 | 11 | 11 | 10 | 11 | 12 | 15 | 1 |
| MOU7 | 8 | 15 | 12 | 21 | 10 | 11 | 13 | 13 | 31 | 9 | 11 | 12 | 8 | 11 | 12 | 11 | 12 | 12 | 16 | 1 |
| MER11 | 8 | 16 | 12 | 21 | 11 | 11 | 13 | 14 | 33 | 9 | 11 | 12 | 8 | 11 | 12 | 10 | 10 | 12 | 17 | 1 |
| Slov73 | 21 | 17 | 13 | 23 | 10 | 11 | 13 | 13 | nr | 9 | 11 | 12 | 9 | 10 | 13 | 10 | 10 | 12 | 1 | 1 |
| RN39 | 21 | 14 | 13 | 23 | 10 | 11 | 14 | 14 | nr | 10 | 11 | 12 | 8 | 11 | 11 | 10 | 9 | 12 | 2 | 1 |
| m86 | 21 | 14 | 12 | 25 | 10 | 11 | 13 | 12 | nr | 9 | 11 | 12 | 8 | 11 | 11 | 11 | 11 | 12 | 3 | 1 |
| MAK108 | 21 | 14 | 13 | 23 | 10 | 10 | 14 | 13 | nr | 9 | 11 | 12 | 10 | 11 | 10 | 10 | 10 | 13 | 4 | 1 |
| Co7 | 25 | 13 | 12 | 24 | 10 | 11 | 13 | 12 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 11 | 11 | 11 | 1 | 4 |
| Co28 | 25 | 13 | 12 | 24 | 10 | 11 | 13 | 12 | 28 | 9 | 11 | 12 | 8 | 9 | 13 | 11 | 11 | 11 | 2 | 1 |
| m538 | 25 | 13 | 12 | 25 | 11 | 11 | 13 | 14 | 32 | 9 | 11 | 12 | 8 | 10 | 12 | 10 | 9 | 11 | 3 | 1 |
| LGL5197 (143) | 25 | 13 | 12 | 25 | 9 | 11 | 14 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 11 | 10 | 9 | 12 | 4 |  |
| Cy23 | 25 | 13 | 12 | 24 | 10 | 12 | 13. | 13 | 32 | 9 | 12 | 12 | 8 | 10 | 12 | 12 | 11 | 11 | 5 | 1 |
| Cy41 | 25 | 13 | 12 | 24 | 10 | 11 | 12 | 15 | 33 | 10 | 11 | 12 | 8 | 10 | 12 | 13 | 11 | 12 | 6 | 1 |
| Cy50 | 25 | 13 | 12 | 25 | 10 | 11 | 13 | 13 | 31 | 9 | 11 | 12 | 8 | 10 | 14 | 12 | 10 | 11 | 7 | 1 |
| Co31 | 25 | 13 | 12 | 24 | 10 | 11 | 13 | 12 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 11 | 11 | 11 | 1 | 1 |
| Co23 | 25 | 13 | 12 | 24 | 10 | 11 | 13 | 12 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 11 | 11 | 11 | 1 | 1 |
| Co38 | 25 | 13 | 12 | 24 | 10 | 11 | 13 | 12 | 31 | 9 | 11 | 12 | 8 | 9 | 13 | 11 | 11 | 11 | 8 | 1 |
| C 04 | 25 | 13 | 12 | 24 | 10 | 11 | 13 | 12 | 30 | 9 | 11 | 12 | 8 | 9 | 13 | 11 | 11 | 11 | 9 | 1 |
| Co 2 | 25 | 13 | 12 | 24 | 10 | 11 | 13 | 12 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 11 | 11 | 11 | 1 | 1 |
| NE74 | 25 | 14 | 14 | 24 | 11 | 13 | 13 | 12 | 29 | 9 | 11 | 12 | 8 | 10 | 12 | 10 | 10 | 12 | 10 | 1 |
| LIB39 | 25 | 13 | 12 | 24 | 10 | 11 | 13 | 13 | 31 | 10 | 11 | 12 | 8 | 10 | 12 | 10 | 10 | 11 | 11 | , |
| A22 | 25 | 13 | 12 | 22 | 9 | 11 | 12 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 9 | 10 | 12 | 12 | 1 |


| Sample | Hg | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS3891 | DYS389] | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7.2 | G09411 | STR hap | Freg |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| PRS68 | 25 | 13 | 12 | 24 | 10 | 11 | 13 | 13 | 32 | 9 | 11 | 12 | 8 | 10 | 12 | 10 | 9 | 10 | 13 | 1 |
| PRS76 6 | 25 | 13 | 12 | 23 | 10 | 11 | 13 | 12 | 31 | 9 | 11 | 12 | 8 | 10 | 11 | 10 | 9 | 10 | 14 | 1 |
| PRS88 | 25 | 13 | 12 | 24 | 10 | 11 | 13 | 14 | 31 | 9 | 11 | 12 | 8 | 10 | 12 | 11 | 10 | 10 | 15 | 1 |
| SDH55 | 25 | 13 | 12 | 22 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 10 | 10 | 11 | 16 | 1 |
| MNA4 | 25 | 13 | 12 | 22 | 9 | 11 | 13 | 13 | 29 | 9 | 11 | 12 | 8 | 10 | 11 | 11 | 12 | 12 | 17 | 1 |
| MNA20 | 25 | 13 | 12 | 24 | 11 | 11 | 12 | 13 | 31 | 9 | 10 | 12 | 8 | 10 | 12 | 10 | 11 | 12 | 18 | 1 |
| Tib9 | 30 | 15 | 12 | 26 | 10 | 10 | 12 | 12 | 29 | 9 | 11 | 12 | 8 | 10 | 11 | 11 | 9 | 12 | 1 | 1 |
| Tib34 | 30 | 15 | 12 | 24 | 10 | 11 | 12 | 13 | 29 | 8 | 11 | 12 | 8 | 10 | 10 | 10 | 9 | 12 | 2 | 1 |
| QIA32 | 30 | 15 | 12 | 25 | 10 | 10 | 12 | 12 | 29 | 9 | 11 | 12 | 8 | 10 | 11 | 10 | 10 | 12 | 3 | , |
| Bug42 | 32 | 13 | 12 | 24 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 9 | 10 | 12 | 1 | 1 |
| Cv66 | 32 | 13 | 12 | 24 | 11 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 9 | 10 | 12 | 2 | 1 |
| Cy13 | 32 | 13 | 12 | 24 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 13 | 13 | 10 | 12 | 3 | 2 |
| Cy15 | 32 | 12 | 10 | 24 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 13 | 10 | 12 | 4 | 1 |
| Cy22 | 32 | 13 | 12 | 24 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 9 | 10 | 12 | 13 | 10 | 12 | 5 | 1 |
| Cy 30 | 32 | 13 | 12 | 24 | 10 | 11 | 14 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 11 | 13 | 10 | 12 | 6 | 1 |
| Cy40 | 32 | 13 | 12 | 24 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 9 | 13 | 10 | 12 | 7 | 1 |
| G14 | 32 | 14 | 12 | 25 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 11 | 13 | 10 | 12 | 8 | 1 |
| G21 | 32 | 14 | 12 | 24 | 10 | 11 | 13 | 14 | 31 | 9 | 11 | 12 | 8 | 10 | 12 | 9 | 10 | 12 | 9 | 1 |
| G41 | 32 | 13 | 12 | 24 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 13 | 13 | 10 | 12 | 3 | 1 |
| 772 | 32 | 13 | 12 | 24 | 10 | 11 | 13 | 14 | 31 | 9 | 11 | 12 | 8 | 9 | 12 | 9 | 10 | 12 | 10 | 1 |
| 940 | 32 | 12 | 10 | 25 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 9 | 10 | 12 | 11 | 1 |
| 1324 | 32 | 13 | 12 | 24 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 9 | 11 | 12 | 12 | 2 |
| Vg3 | 32 | 13 | 12 | 25 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 9 | 10 | 12 | 13 | 1 |
| NW89 | 32 | 13 | 13 | 24 | 10 | 11 | 13 | 13 | 29 | 9 | 11 | 12 | 8 | 10 | 12 | 10 | 10 | 12 | 14 | 1 |
| S150 | 32 | 13 | 12 | 25 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 10 | 10 | 12 | 15 | 1 |
| 2B12 | 32 | 12 | 10 | 24 | 10 | 11 | 14 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 12 | 10 | 12 | 16 | 1 |
| Sp42 | 32 | 13 | 13 | 22 | 8 | 11 | 12 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 13 | 10 | 12 | 17 | 1 |
| T38 | 32 | 13 | 12 | 24 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 11 | 9 | 10 | 12 | 18 | 1 |
| SYD15 | 32 | 13 | 12 | 24 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 12 | 9 | 11 | 12 | 12 | 1 |
| SYD24 | 32 | 13 | 12 | 25 | 10 | 11 | 13 | 14 | 31 | 9 | 11 | 12 | 8 | 10 | 12 | 10 | 9 | 12 | 19 | 1 |
| RAB113 | 32 | 13 | 12 | 23 | 10 | 11 | 10 | 12 | 29 | 9 | 11 | 12 | 8 | 10 | 10 | 10 | 11 | 13 | 20 | 1 |
| RAB136 | 32 | 13 | 12 | 23 | 10 | 11 | 13 | 12 | 29 | 9 | 11 | 12 | 8 | 10 | 10 | 10 | 10 | 13 | 21 | 1 |
| RAB139 | 32 | 13 | 12 | 23 | 10 | 11 | 13 | 12 | 29 | 9 | 11 | 12 | 8 | 10 | 10 | 10 | 11 | 13 | 22 | 1 |
| m125 | 33 | 13 | 12 | 25 | 9 | 11 | 14 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 11 | 11 | 11 | 12 | 1 | 1 |
| G44 | 33 | 13 | 12 | 25 | 10 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 13 | 11 | 12 | 2 | 1 |
| Sp30 | 33 | 13 | 12 | 25 | 9 | 11 | 13 | 13 | 29 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 11 | 12 | 3 | 1 |
| Sp39 | 33 | 13 | 12 | 23 | 9 | 12 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 13 | 11 | 12 | 4 | 1 |
| AGA169 | 33 | 13 | 12 | 23 | 9 | 11 | 13 | 11 | 27 | 9 | 11 | 12 | 8 | 10 | 10 | 9 | 11 | 10 | 5 | 1 |
| RAB017 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 12 | 11 | 12 | 6 | 1 |
| RAB038 | 33 | 13 | 12 | 24 | 9 | 12 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 11 | 12 | 7 | 1 |
| RAB070 | 331 | 13 | 12 | 24 | 9 | 11 | 13 | 15 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 12 | 11 | 11 | 8 | 1 |
| RAB075 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 11 | 12 | 9 | 4 |


| Sample | Hg . | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS3891 | DYS389II | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7. 2 | C09411 | STR hap | Freq |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| RAB104 | 33 | 13 | 12 | 24 | 10 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 11 | 11 | 11 | 12 | 10 | 1 |
| RAB117 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 11 | 12 | 9 | 1 |
| RAB118 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 10 | 12 | 11 | 1 |
| RAB119 | 33 | 13 | 12 | 24 | 10 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 11 | 12 | 12 | 1 |
| RAB126 | 33 | 13 | 12 | 24 | 11 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 11 | 11 | 11 | 12 | 13 | 1 |
| RAB129 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 11 | 11 | 11 | 12 | 14 | 1 |
| RAB130 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 12 | 29 | 9 | 11 | 12 | 8 | 10 | 10 | 9 | 9 | 10 | 15 | 1 |
| RAB131 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 10 | 11 | 12 | 16 | 2 |
| RAB133 | 33 | 13 | 12 | 24 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 11 | 11 | 11 | 12 | 17 | 1 |
| RAB134 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 12 | 12 | 18 | 1 |
| RAB137 | 33 | 13 | 12 | 23 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 11 | 10 | 11 | 11 | 12 | 19 | 1 |
| RAB203 | 33 | 13 | 12 | 25 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 11 | 12 | 20 | 1 |
| RAB204 | 33 | 13 | 12 | 23 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 11 | 12 | 21 | 1 |
| MOJ2 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 11 | 12 | 9 | 1 |
| MOJ6 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 11 | 12 | 9 | 1 |
| MSM1 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 10 | 11 | 12 | 16 | 1 |
| MSM2 | 33 | 14 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 12 | 12 | 18 | 1 |
| MSM4 | 33 | 13 | 12 | 24 | 9 | 11 | 13 | 14 | 30 | 9 | 11 | 12 | 8 | 10 | 9 | 11 | 11 | 12 | 22 | 1 |
| MSM8 | 33 | 13 | 13 | 24 | 9 | 11 | 13 | 14 | 29 | 9 | 11 | 12 | 8 | 10 | 10 | 11 | 11 | 12 | 23 | 1 |
| 439 | 34 | 17 | 12 | 24 | 11 | 11 | 13 | 14 | 32 | 9 | 11 | 12 | 8 | 10 | 11 | 13 | 10 | 12 | 1 | 1 |
| 457 | 34 | 17 | 12 | 26 | 10 | 11 | 13 | 13 | 30 | 9 | 11 | 12 | 8 | 10 | 13 | 12 | 10 | 11 | 2 | 1 |
| 458 | 34 | 17 | 12 | 26 | 10 | 11 | 13 | 14 | 31 | 10 | 11 | 12 | 8 | 10 | 12 | 13 | 12 | 12 | 3 | 1 |

### 5.3.5. Distribution of microsatellite allele lengths

Microsatellite allele length diversity can be compared between the different regional samples. The total number of samples analysed that gave complete data were as follows, for Sub-Saharan Africa (11), North Africa (36), Asia (20) and Europe (35). Different sized sample sets may influence the detection of alleles, for example, fewer samples reduce the probability that all possible allele lengths will be contained. The data for the four regional samples complement some previously published population data, which adds support to this new work, including one study analysing two Spanish populations (Basque and Catalan) (Pérez-Lezaun et al., 1997b); a large study of North African populations (Bosch et al., 1999), and two European populations (German and Polish) (Kayser et al., 2000). Although, a complete comparison between these three studies is not always possible, due to the generation of new primers for certain loci (DYS19 and YA7.1), for the purpose of this study. Each multiplex system (CTS, MS1 and EBF) will be briefly discussed in turn.

### 5.3.5.1. The CTS multiplex

For the CTS (Ayub et al., 2000) multiplex, see Figure 5.19, in which all six loci and the allelic frequencies are detailed. In general the allele lengths and frequencies are consistent with previous published data (Ayub et al., 2000). Two loci are completely invariant (DYS436 and DYS437), and one other is near-invariant (DYS435). Two others (DYS434 and DYS438) show a unimodal distribution, and finally DYS439 shows a bimodal distribution with two common alleles predominantly found in North Africa and Europe populations.

### 5.3.5.2. The MS1 multiplex

For the MS1 (Thomas et al., 1999) multiplex, see Figure 5.20, in which all six loci and the allelic frequencies are detailed. Five of the loci (DYS19, DYS388, DYS390, DYS392 and DYS393) show a unimodal distribution with a predominant allele length being observed, although other allele lengths are detected at lower frequencies. These data correlate well with previous analyses examining North African (Bosch et al., 1999), Basque and Catalan (Pérez-Lezaun et al., 1997b), German and Polish populations (Kayser et al., 2000). However, previous analysis indicated in general when the most common alleles differ between continental groups, Asia alleles tend to be larger, and African and Native American alleles


Figure 5.19. Microsatellite allele frequencies observed within YAP+ global samples for the CTS multiplex system
The 6 microsatellites included within the CTS multiplex (Ayub et al., 2000). Each panel shows the allelic distribution for four different regions sampled, and each region has been allocated a different colour. The actual microsatellite length in base pairs (bp) is indicated followed by the corresponding repeat number in brackets.


Figure 5.20. Microsatellite allele frequencies observed within YAP+ global samples for the MS1 multiplex system
The 6 microsatellites included in the MS1 multiplex adapted from (Thomas et al., 1999). Eack panel shows the allelic distribution for four different regions sampled, and each region has been allocated a different colour. The actual microsatellite length in base pairs (bp) is indicated followed by the corresponding repeat number in brackets.
tend to be shorter (Kayser et al., 1997). DYS391 shows a bimodal distribution, again with two allele lengths being found in North African and European populations respectively.

### 5.3.5.3. The EBF multiplex

For the EBF (Bosch, unpublished) multiplex, see Figure 5.21, in which all five of the loci and the allelic frequencies are described. With the exception of YA7.1, the other four loci (DYS389I, DYS389II, G09411and YA7.2) appear to have a unimodal distribution of allele lengths among the four regional populations. These again support previously published work, examining North African (Bosch et al., 1999), German and Polish (Kayser et al., 2000) and 35 unrelated male individuals (White et al., 1999). However, YA7.1 shows an unusual distribution of allele lengths, in that it would appear unimodal except for the high frequency of 183bp alleles within Asian and European populaitons.

### 5.3.6. The examination of pairwise differences within haplogroups

The microsatellite haplotype data were used to calculate pairwise differences within each haplogroup, and this was done in the Arlequin 2.00 program (Schneider et al., 2000). These data can be plotted onto graphs, and have been combined for loci that were only represented by a few chromosomes (hg 4, 30,34 and 21) and loci represented more abundantly (hg 25, 32, 33 and 8). For these graphs, see Figure 5.22. However due to the lack of chromosomes contained within Figure 5.22A it is difficult to interpret these graphs. The expected pattern is a bell shape, and these graphs show the diversity within each haplogroup, because the more differences observed, the bell shape moves towards the right hand side, as seen in Figure 5.22B. Haplogroup 25 appears to show the most number of differences, and this again supports the proposal that this haplogroup is currently undefined and probably contains a number of undefined lineages.

### 5.3.7. Examination of microsatellite diversity at each locus

The examination of microsatellite diversity allows all loci to be compared individually. The values were calculated in the same was employing Nei's estimator of diversity (Nei, 1978). (See Table 5.6).


Figure 5.21. Microsatellite allele frequencies detected within YAP+ global samples for the EBF multiplex system
The 5 microsatellites included in the EBF multiplex (Bosch, unpublished). Each panel shows the allelic distribution for four different regions sampled, and each region has been allocated a different colour. The actual microsatellite length in base pairs (bp) is indicated followed by the corresponding repeat number in brackets.


Figure 5.22. Pairwise difference values calculated for eight different haplogroups The pairwise difference values were calculated within Arlequin (Schneider et al., 2000. (A) four haplogroups (hg 4, 30, 34 and 21) were plotted together, and (B) four haplogroups (hg 25, 32, 33 and 8) were also plotted together. Each haplogroup has been allocated a different colour, see key, and their relative positions within the tree are also indicated.

Table 5.6. Diversity values for each individual locus

| Loci | Repeat type | Repeats <br> (global <br> studies) | Repeats <br> observed <br> (this study) | Size range (bp) | This <br> study <br> h= | global <br> h= |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| DYS434 | CTAT | $8-11$ | $8-10$ | $110-122$ | 0.175 | 0.222 |
| DYS435 | TGGA | $11-13$ | $10-12$ | $220-228$ | 0.037 | 0.070 |
| DYS436 | GTT | $10-15$ | 12 | $128-143$ | 0 | 0.064 |
| DYS437 | TCAT | $8-10$ | $8-10$ | $186-202$ | 0.056 | 0.664 |
| DYS438 | TTTTC | $6-12$ | $9-12$ | $203-233$ | 0.419 | 0.684 |
| DYS439 | GATA | $9-14$ | $9-14$ | $238-258$ | 0.709 | 0.728 |
|  |  |  |  |  |  |  |
| DYS19 | CTAT/C | $10-19$ | $12-17$ | $173-209$ | 0.577 | 0.72 |
| DYS388 | ATA | $10-16$ | $10,12-14$ | $123-141$ | 0.176 | 0.26 |
| DYS390 | CTG/AT | $18-27$ | $20-26$ | $192-228$ | 0.702 | 0.73 |
| DYS391 | CTAT | $8-13$ | $8-11$ | $156-176$ | 0.493 | 0.49 |
| DYS392 | ATT | $7-16$ | $7,10-13$ | $155-182$ | 0.194 | 0.52 |
| DYS393 | GATA | $9-15$ | $10,12-15$ | $107-131$ | 0.393 | 0.34 |
|  |  |  |  |  |  |  |
| DYS389I/II | CTG/AT | $\mathrm{I}=10-16$ | $11-15$ | $\mathrm{I}=142-166$ | 0.661 | 0.61 |
| YA7.1 |  | GATA | $9-13$ | $97-33$ | $\mathrm{II}=258-300$ | 0.596 |
| YA7.2 | GATA | $8-12$ | $9-13$ | $170-186$ | 0.75 |  |
| G09411 | TGTA | $10-14$ | $10-13$ | $174-190$ | 0.626 | na |

Note: DYS389II has incomplete data for hg 21, and a total of 102 alleles were used when calculating that diversity value. All other loci used all 106 alleles. DYS437 shows no diversity if the incomplete hg 21 data is excluded, but this value was calculated using all the data. Standard errors (SE) were not calculated for these values because sample sizes in all cases are identical or very similar. Global h values previously calculated (Kayser et al., 1997; Ayub et al., 2000) were included within the table for comparison purposes, although these are not available for all loci, and na indicates such loci.

The values used for comparison purposes from Kayser et al. (1997) appear to be approximately similar for the loci concerned, although differences are observed for example, for DYS392 the value was 0.194 compared to 0.52 (Kayser et al., 1997), and DYS437 the value of 0.056 compared to 0.664 (Ayub et al., 2000). This difference may reflect the fact that only YAP+ loci are being examined here, when all haplogroups were examined in the (Kayser et al., 1997; Ayub et al., 2000) studies.

### 5.3.8. The dating of mutations defining haplogroups from microsatellite data

The haplotype data were used to generate a modal haplotype within each set of haplogroup data. This unobserved modal haplotype within each haplogroup was used for comparison purposes with all the observed haplotypes to enable an ASD (average squared distance) value to be determined to allow the dating of the mutations defining the haplogroups. However, not all haplogroups
are defined by specific mutations (see Figure 5.23), although this can be overcome by combining a number of different haplogroups (see Figure 5.23). What is being dated is the MRCA of a group of chromosomes, which in sometimes include several haplogroups. ASD values were calculated within an Excel table, and the process involves, the generation of an unobserved modal haplotype for all loci; the number of differences between the modal haplotype and the observed haplotypes are calculated and then squared. These squared values are summed, and then divided by the number of loci and chromosomes, to produce an ASD value. These numbers were also recalculated using the Microsat program from the Eric Minch web site. When the modal haplotype was unclear, all the possible modal haplotypes were used to calculate ASD values, and the lowest value was used, because it is most likely to be the correct evolutionary path. The following table (5.7) shows how the dates are calculated. See Table 5.8, for all the ASD values and dates generated.

Table 5.7. Estimation of the age of a lineage - ( hg 34 as an example).

| ASD | $\mu$ (15 loci) | (95\% upper CI) | (95\% lower CI) | generation time (yrs) |
| :--- | :--- | :--- | :--- | :--- |
| 0.354 | 0.0028 | 0.00172 | 0.00427 | 25 |
| ASD | $\mu$ (8 tetra loci) | (95\% upper CI) | (95\% lower CI) | generation time (yrs) |
| 0.354 | 0.00317 | 0.00189 | 0.00494 | 25 |

The mutation rate estimates are taken from Kayser et al. (2000).
hg 34: ASD/ $\boldsymbol{\mu}=$ time in generations since the MRCA.
$0.354 / 0.0028=126.4286$ generations ( $95 \%$ CI $205.8140-82.9040$ )
$0.354 / 0.00317=111.6719$ generations ( $95 \%$ CI $187.3016-71.6599$ )

Number of generations $x$ estimated generation time $=$ time to MRCA.
$126.4286 \times 25=3161$ years ( $95 \%$ CI $5145-2073$ )
$111.6719 \times 25=2792$ years ( $95 \%$ CI $4683-1791$ )

A number of different generation times have been used in the literature for example, 20, 27 (Cavalli-Sforza et al., 1994; Underhill et al., 1996; Weiss and von Haeseler, 1996) and 35 (Helgason et al., 2000; Tremblay and Vézina, 2000). The following calculation shows how these different values effect the estimated age of the mutation defining a particular haplogroup.

## Different generation times.

$126.4286 \times 27=3414$ years ( $95 \%$ CI $5557-2238$ )


Figure 5.23. The tree structure used for the dating of mutations along the YAP branch
The skeleton version of the maximum parsimony tree shows only the YAP sublineages. Some haplogroups are defined by known mutations (green circles) others are not defined by specific mutations (yellow circles). This lack of mutations causes problems when trying to date the age of mutations associated with haplogroups.
$111.6719 \times 27=3015$ years $(95 \%$ CI $5057-1935)$
$126.4286 \times 35=4425$ years ( $95 \%$ CI $7203-2902$ )
$111.6719 \times 35=3909$ years ( $95 \%$ CI $6556-2508$ )

All the values shown after this point have been calculated using the generation time of 25 .

Table 5.8. Dates calculated from microsatellite diversity within eight haplogroups

|  |  | $\boldsymbol{\mu}$ - 15 loci | 95\% upper C | 95\% lower CI | $\mu-8$ loci | 95\% upper | 95\% lower CI |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Hgs | ASD | $\mu=0.0028$ | $\mu=0.00172$ | $\mu=0.00427$ | $\mu=0.00317$ | $\mu=0.00189$ | $\mu=0.00494$ |
| Super- <br> hg 4+ | 0.851 | $\begin{aligned} & 303.9286 \\ & =7598 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 494.7674 \\ & =12369 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 199.2974 \\ & =4982 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 268.4543 \\ & =6711 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 450.2646 \\ & =11257 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 172.2672 \\ & =4307 \mathrm{yrs} \end{aligned}$ |
| hg 30 | 0.167 | $\begin{aligned} & 59.6429 \\ & =1491 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 97.0930 \\ & =2427 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 39.1101 \\ & =978 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 52.6814 \\ & =1317 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 88.3598 \\ & =2209 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 33.8057 \\ & =845 \mathrm{yrs} \end{aligned}$ |
| hg 34 | 0.354 | $\begin{aligned} & 126.4286 \\ & =3161 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 205.8140 \\ & =5145 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 82.9040 \\ & =2073 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 111.6719 \\ & =2792 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 187.3016 \\ & =4683 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 71.6599 \\ & =1791 \mathrm{yrs} \end{aligned}$ |
| Super- <br> hg 21+ | 0.775 | $\begin{aligned} & 276.7857 \\ & =6920 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 450.5814 \\ & =11265 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 181.4988 \\ & =4537 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 244.4795 \\ & =6112 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 410.0529 \\ & =10251 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 156.8826 \\ & =3922 \mathrm{yrs} \end{aligned}$ |
| hg 31 | 0.405 | $\begin{aligned} & 144.6429 \\ & =3616 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 235.4651 \\ & =5887 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 94.8478 \\ & =2371 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 127.7603 \\ & =3194 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 214.2857 \\ & =5357 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 81.9838 \\ & =2050 \mathrm{yrs} \end{aligned}$ |
| Super- $\text { hg } 25+$ | 0.676 | $\begin{aligned} & 241.4286 \\ & =6036 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 393.0233 \\ & =9826 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 158.3138 \\ & =3958 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 213.2492 \\ & =5331 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 357.6720 \\ & =8942 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 136.8421 \\ & =3421 \mathrm{yrs} \end{aligned}$ |
| hg 32 | 0.625 | $\begin{aligned} & 223.2143 \\ & =5580 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 363.3721 \\ & =9084 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 146.3700 \\ & =3659 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 197.1609 \\ & =4929 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 330.6878 \\ & =8267 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 126.5182 \\ & =3163 \mathrm{yrs} \end{aligned}$ |
| hg 33 | 0.203 | $\begin{aligned} & 72.5 \\ & =\mathbf{1 8 1 3} \mathbf{y r s} \end{aligned}$ | $\begin{aligned} & 118.0233 \\ & =2951 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 47.5410 \\ & =1189 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 64.0379 \\ & =1601 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 107.4074 \\ & =2685 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 41.0931 \\ & =1027 \mathrm{yrs} \end{aligned}$ |
| hg 8 | 0.351 | $\begin{aligned} & 125.3571 \\ & =3134 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 204.0698 \\ & =5102 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 82.2014 \\ & =2055 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 110.7256 \\ & =2768 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 185.7143 \\ & =4643 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 71.0526 \\ & =1776 \mathrm{yrs} \end{aligned}$ |

Incomplete data were available for three different hg 31 samples, with only 15 loci haplotypes being determined. Fourteen loci were used to calculate hg 4 and hg 21 dates.

These dates can be used to estimate the influence of prehistoric migrations. As discussed later, many of these dates are unexpectedly young; however, here they are taken at face value. Two haplogroups appear to be much younger than the others: the Chinese-specific mutation M15 which defines hg 30 has been dated to 1491 years ( $95 \%$ CI 2427 - 978) and may reflect local Asian (Chinese) differentiation, although this post-dates the second major Asian migration of the Yayoi people from Korea approximately 4,300 years ago (Cavalli-Sforza et al.,

1994; Hammer, 1995). The mutation defining hg 33, M81, is approximately 1813 years old ( $95 \%$ CI 2951 - 1189) and might reflect a recent mutational event that probably occurred within North Africa. The mutations sY81 and PN1 that define hg 8 have been dated to 3134 years ( $95 \%$ CI 5102 - 2055) and support the hypothesis that this lineage was probably spread by the Bantu expansion that occurred between 3,000 and 500 years ago. The Japanese-specific mutation (12f2) that defines hg 34 has been dated to 3161 years ( $95 \%$ CI 5145 - 2073) and is likely to reflect local Asian (Japanese) diversity, but again post-dates the second Yayoi migration $\sim 4,300$ years ago. The age of 3616 years ( $95 \%$ CI $5887-2371$ ) was calculated for the M33 mutation that defines hg 31, and these chromosomes were only observed within African populations, and probably reflect an African mutation. The M78 mutation defines hg 32 and has an estimated age of 5580 years ( $95 \%$ CI $9084-3659$ ); this date is too young to represent Neolithic movement which began $\sim 10,000$ years ago.

The next three haplogroups are not defined by specific mutations, and to estimate ages requires the combination of all relevant derived lineages. Haplogroup 25 and all other haplogroups derived for PN2/DYS391 in combination generate an age of 6036 years ( $95 \%$ CI $9826-3958$ ) and this 'superhg $25+^{\prime}$ is widely distributed throughout global samples, and may reflect an expansion from Africa into the rest of the world. This age pre-dates the Bantu expansion and is one of the oldest calculated ages for mutations within the YAP branch. 'Super-hg $21+$ ' composed of hgs 21, 8, 32, 33 and 31 has generated an age of approximately 6920 years ( $95 \%$ CI $11265-4537$ ). Finally, the 'super-hg $4+$ ' composed of all YAP+ lineages has been dated to 7598 years ( $95 \%$ CI 12369 4982).

Some of these estimates can be compared with estimates in the literature; however, published dates using coalescent based methods are very different to ones derived here. For example, the entire YAP+ lineage has been dates to 55 60,000 YBP (Hammer et al., 1998; Karafet et al., 1999) from coalescence analysis, and the PN1 mutation defining hg 8 has been dated to 20,000 years (Hammer et al., 1998). Those values compare with 7598 years ( $95 \%$ CI 12369 - 4982) for the YAP+ lineage ('super-hg4+'), and 3134 years ( $95 \%$ CI 5102 - 2055) for hg 8 respectively. These large differences are typical of comparisons between coalesence and microsatellite based methods, and raise two questions: which method is giving the most realistic estimate? And why are they so different? The
fact that YAP+ chromosomes are widely distributed throughout the Old world, and that several YAP+ sublineages also have wide distributions, suggests that the microsatellite-based dates for the YAP+ chromosomes is too young to be credible. The reason for these young dates may be that so many generations have passed since the common ancestor that, the rapidly mutating microsatellite loci, have become saturated. However, two of these dates correspond well with other previously published dates based on microsatellite evidence, for the mutation SRY-8299 which defines hg 21 and its sub-groups an age of 6,483 years ( $95 \%$ CI 30782-493) and for sY81, an age of 3017 ( $95 \%$ CI 18565 - 0) (Bosch et al., 1999). One source of uncertainty is the lack of information with regards to mutation rate for the majority of the new microsatellite loci included within this study. One possible way to exclude this uncertainty, is to remove all but the six loci used in the Kayser et al. (2000) study. For the values generated in the same way but with only six well characterised loci, see Table 5.9.

Table 5.9. Dates calculated from six microsatellite loci within eight haplogroups

|  |  | $\mu-8$ loci | 95\% lower | 95\% upper CI |
| :---: | :---: | :---: | :---: | :---: |
| Hgs | ASD | $\mu=0.00317$ | $\mu=0.00189$ | $\mu=0.00494$ |
| Super- <br> hg 4+ | 1.196 | $\begin{aligned} & 377.2871 \\ & =9432 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 632.8042 \\ & =15820 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 242.1053 \\ & =6053 \mathrm{yrs} \end{aligned}$ |
| hg 30 | 0.167 | $\begin{aligned} & 52.6814 \\ & =1317 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 88.3598 \\ & =2209 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 33.8057 \\ & =845 \mathrm{yrs} \end{aligned}$ |
| hg 34 | 0.444 | $\begin{aligned} & 140.0631 \\ & =3502 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 234.9206 \\ & =5873 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 89.8785 \\ & =2247 \mathrm{yrs} \end{aligned}$ |
| Superhg 21+ | 1.094 | $\begin{aligned} & 345.1104 \\ & =8628 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 578.8360 \\ & =14471 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 221.45775 \\ & =5536 \mathrm{yrs} \end{aligned}$ |
| hg 31 | 0.500 | $\begin{aligned} & 157.7287 \\ & =3943 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 264.5503 \\ & =6614 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 101.2146 \\ & =2530 \mathrm{yrs} \end{aligned}$ |
| Superhg 25+ | 1.007 | $\begin{aligned} & 317.6656 \\ & =7942 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 532.8042 \\ & =13320 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 203.8462 \\ & =5096 \mathrm{yrs} \end{aligned}$ |
| hg 32 | 0.326 | $\begin{aligned} & 102.8391 \\ & =2571 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 172.4868 \\ & =4312 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 65.9919 \\ & =1650 \mathrm{yrs} \end{aligned}$ |
| hg 33 | 0.274 | $\begin{aligned} & 86.4353 \\ & =2161 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 144.9735 \\ & =3624 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 55.4656 \\ & =1387 \mathrm{yrs} \end{aligned}$ |
| hg 8 | 0.481 | $\begin{aligned} & 151.7350 \\ & =3793 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 254.4974 \\ & =6362 \mathrm{yrs} \end{aligned}$ | $\begin{aligned} & 97.3684 \\ & =2434 \mathrm{yrs} \end{aligned}$ |

Six microsatellite loci (DYS19, DYS390, DYS391, DYS393, DYS389I and DYS389II) were used to calculate dates to MRCA. Two super-hgs (4+ and 21+) used only five loci due to incomplete data at DYS389II within hg 31.

The estimated ages of mutations which define a number of haplogroups have increased in all but two cases (hg 30 and hg 32 ). Four possible reasons have been put forward for the young ages of mutations calculated from microsatellite data, and include, (A) the mutation precedes TMRCA - and so the date may reflect a time more recent than the age of the mutation (Bosch et al., 1999). This may be because some of the original lineages produced by mutations may either have been lost via drift or were just unobserved within these data. However, the relevant age is the TMRCA, rather than the date a mutation occurred. When a date conflicts with independent evidence, for example the wide distribution of a lineage, this explanation does not help. (B) population biases can affect the ages, although these samples are widely distributed across three continents. Some samples used in a previous study (Bosch et al., 1999), were included as the North African populations, which would mainly influence the (M31) hg 33 date where these chromosomes predominate, although additional samples were included for hg 8 estimates. The date for hg 8 was very similar to dates calculated within a previous study (Bosch et al., 1999), and this cannot be due to the same samples being used. It seems unlikely that population bias plays a major role here, however, some of the sample sizes could usefully be increased to enable more accurate dates to be determined.
(C) microsatellite saturation, occurs when due to the high mutation rate these microsatellites would have reached equilibrium (Bosch et al., 1999). This may be a particularly severe problem with a very old lineage (e.g. YAP+), and may imply that young lineages are dated more accurately than older ones. (D) microsatellite rate overestimation (Bosch et al., 1999) is another potential problem, due to the fact the most loci have different mutation rates (Heyer et al., 1997; Kayser et al., 1997; Kayser et al., 2000), the use of a pooled rate will lead to errors. This may be a particular problem with loci such as DYS436 And DYS437, which are completely invariant within YAP+ chromosomes, and which therefore must have much lower mutation rates than the average. This was addressed by removing the loci with unknown mutation rates; however, this did not increase the ages dramatically (see Tables 5.8 and 5.9).

The two cases where the ages decreased could be due to reducing the overall level of diversity at the microsatellite loci. This statistical analysis, reveals how sensitive dating procedures are, and estimated dates should be quoted in relation to historical events with caution.

### 5.3.9. Generation of phylogenetic networks

The microsatellite data can be separated into a series of different lineages (haplotypes) depending upon the repeat lengths for the different loci being analysed. This allows the relationships between all the samples to be examined within haplogroups, as recently undertaken by Forster et al. (2000) and Helgason et al. (2000). These lineages can be compared using the computer program Network 2.0 (Bandelt et al., 1999). Median-joining networks can be drawn for each individual lineage, to see how the samples cluster with respect to geographic affiliation, and this may be informative when comparing a wide range of geographically distributed samples, such as being studied here. The networks may suggest an evolutionary relationship between the samples, that may be previously unsuspected.
5.3.10. Does increasing the number of loci increase the separation of samples within the networks?

During this section, the robustness of generating median-joining networks from microsatellite loci data will be tested. The samples had been previously assigned to biallelic haplogroups, and the diversity within these haplogroups was examined further using the faster mutating markers. A comparison using only seven loci routinely used during forensic analysis (DYS389I/II, DYS390, $D Y S 391, D Y S 392, D Y S 393$ and $D Y S 19$ ) and all the seventeen loci was undertaken, to see if the structure and informativeness of the networks altered. Networks generated from the seven loci will be discussed first, followed by the seventeen loci. In the following section, the networks are discussed qualitatively, as well as comparing ASD values for each haplogroup (see Table 5.10). Nei's estimator is inadequate, since most haplotypes are singletons, and $h$ does not include a measure of distance.

Table 5.10. ASD values for all YAP+ haplogroups

| Haplogroup | No. chromosomes | ASD value |
| :--- | :--- | :--- |
| hg 4 | 5 | 0.812 |
| hg 30 | 3 | 0.167 |
| hg 34 | 3 | 0.354 |
| hg 21 | 4 | 0.547 |
| hg 31 | 3 | 0.405 |
| hg 25 | 21 | 0.454 |
| hg 32 | 24 | 0.625 |
| hg 33 | 28 | 0.203 |
| hg 8 | 18 | 0.351 |

No standard errors are available for these ASD values. The values take into account the different sample sizes for each haplogroup. These values were calculated from seventeen loci, however hg 31 samples had only 15 loci data, and fourteen loci were used to calculate hg 4 and hg 21 values.

Haplogroup 4 shows the highest level of diversity, which may be expected due it being the most ancestral lineage. Haplogroup 32 is the next most diverse, and this is distributed widely across the globe, which explain this higher level of diversity. Haplogroup 8 shows quite a low level of diversity and this may be due to the Bantu expansion reducing the overall diversity. The youngest lineage (hg 30) at 1491 years ( $95 \%$ CI 2427-978) shows the lowest diversity of all the haplogroups, and this may be a direct consequence of the recent evolution of this Chinese-specific lineage.

### 5.3.10.1. Generation of networks from seven microsatellite loci

Two figures ( 5.24 and 5.25 ) summarise the seven-locus data for all eight haplogroups. Figure 5.24 (hgs $4,30,34$ and 21) shows the most ancestral lineages within the YAP branch, but the networks are small due to the small number of chromosomes contained, and this may cause potential problems when interpreting these data reliably. Three haplogroups show population and geographical specificity; hg 4 was Asian-specific with only Japanese and Chinese chromosomes detected; hg 30 and hg 34 were only detected in Chinese and Japanese populations respectively. Haplogroup 21 contains Sub-Saharan African, a Pakistani individual and a European chromosome. Almost all chromosomes are singletons, and all the networks are quite simple although many unobserved haplotypes separate most chromosomes.


Figure 5.24. Median-joining networks generated from 7 microsatellite loci within four defined biallelic haplogroups (A) shows the haplogroups in black being examined and their relative positions within the tree. (B) is the median-joining network generated for four haplogroups encompassing data from 7 Y -chromosomal microsatellites. The continent-specific colours are indicated in the key and unobserved haplotypes are shown as small black filled circles.

| $\square \mathrm{N}+$ W. Europe | $\square$ | Sub Saharan Africa |
| :--- | :--- | :--- |
| $\square$ S + E. Europe | $\square$ | Japan |
| $\square$ | C. Europe | $\square$ |
| China |  |  |
| $\square$ | North Africa | $\square$ Other Asian |

hg 32
igure 5.25. Median-joining networks generated from 7 microsatellite loci within four efined biallelic haplogroups

1) shows the haplogroups in black being examined and their relative positions within the tree. 3) is the median-joining network generated for four haplogroups encompassing data from 7 chromosomal microsatellites. The continent-specific colours are indicated in the key and nobserved haplotypes are indicated as small black filled circles.

Figure 5.25 (hgs 25, 32, 33 and 8) includes the more derived lineages within this YAP branch, and all of these networks are more complex, which is due to the larger number of chromosomes contained. The hg 25 network was quite complex and did not really appear to show the clear separation of continental samples, although a single branch contained only Asian and African samples. One particular haplotype appears to be Ladin-specific and was observed in five Cortina samples; a single-step and a two-step derivative are also Cortina-specific. A geographically close, north eastern Italian chromosome is connected to these Ladin samples but separated by six unobserved haplotypes. As the maximum parsimony tree indicates (Figure 5.25) hgs 32 and 33 are hg 25 -derived lineages; however, the chromosomes contained show very different geographical distributions. Haplogroup 32 was mainly observed within southern and eastern European samples, although a few examples of North African and other Asian chromosomes were also detected. A predominant haplotype was detected amongst eight different samples with only one of these being an Asian from Pakistan. The North African samples remained separated from the main network by a number of unobserved haplotypes on an individual branch.

Haplogroup 33 was predominantly observed within North African samples although three examples of other geographically associated chromosomes (two Spanish and one Greek) were also represented; these may reflect recent admixture with North African populations. A single haplotype was observed in eleven individuals and all the remaining haplotypes spread out from this point, including six one-step derivatives and this may represent a founding haplotype. Haplogroup 8 is the most derived lineage and was previously detected mainly within Sub-Saharan individuals (Seielstad et al., 1994; Hammer et al., 1998; Karafet et al., 1999), although during this analysis samples from North Africa and Asia were also observed. This network structure appeared simpler but did show a degree of population specificity within some regions for example, the SubSaharan African samples cluster to one side of the network, and only a few unobserved haplotypes were predicted to be included.

The definition of a cluster can be subjective, and clusters have previously been identified via a modal haplotype being 1 mutational step away from a predominant root haplotype (Thomas et al., 1998). This approach cannot be used when examining the YAP+ data because no root haplotypes are detectable, except perhaps in hg 32 and hg 33.

### 5.3.10.2. Generation of networks from seventeen microsatellite loci

Networks for same eight haplogroups (Figure 5.26 - hgs 4, 30, 34, 21 and Figure 5.27 - hgs 25, 32, 33 and 8 ) have also been determined using all seventeen microsatellite loci (see Table 5.5). As previously described, hg 30 and hg 34 are Chinese-and Japanese-specific lineages (Figure 5.26) although one sample (Qia32) was removed as the intermediate haplotype (Figure 5.24) and shown as a distinct haplotype (Figure 5.26). Although, the ASD values calculated for these haplogroups vary from 0.162 , and 0.354 for hg 30 and 34 respectively; which show an increasing level of diversity, and this may be due to the small number of chromosomes belonging to these particular haplotypes. Within hg 4 two Tibetan samples (Tib1 and 35) were previously shown to belong to the same haplotype have now been separated, and this haplogroups showed the highest level of calculated diversity (0.812). Haplogroup 21 appeared to show more of a population-specific distribution when compared to (Figure 5.24) because these two Sub-Saharan African chromosomes appear to be more closely linked, obviously due to the increased number of informative loci. However, it should be noted that the number of unobserved haplotypes separating the samples within all the networks is considerable, which is supported by the large ASD value of 0.547 .

Contained within Figure 5.27, the most ancestral haplogroup (hg 25) shows an increased level of sub-structuring (compared to Figure 5.25 supported by an ASD value of 0.454 ), in that the majority of other Asian, southern and eastern European, and central European (Cortina) chromosomes cluster together, with only a few exceptions. In particular the Cortina chromosomes form an individual cluster on a separated branch of the network, and the Italian sample (NE74) was positioned on the opposite side of this new network. The number of unobserved haplotypes within all the networks increased dramatically. The network complexity and the level of sub-structuring within hgs 32 and 33


Figure 5.26. Median-joining networks generated from 17 microsatellite loci for a number of different biallelic haplogroups (A) shows the haplogroups in black being examined and their relative positions within the tree. (B) is the median-joining network generated for three haplogroups encompassing data from 17 Y -chromosomal microsatellites, and hg 21 only includes 15 loci. The continent-specific colours are indicated in the key and unobserved haplotypes are shown as small black filled circles.


Figure 5.27. Median-joining networks generated from 17 microsatellite loci within four defined biallelic haplogroups
(A) shows the haplogroups in black being examined and their relative positions within the tree.
(B) is the median-joining network generated for four haplogroups encompassing data from 17 Y-chromosomal microsatellites. The continent specific colours are indicated in the key and all unobserved haplotypes are shown as small black filled circles. Some clusters are circled in hlark lines
increased. Within hg 32, the separation of the three North African individuals increased, and the predominant haplotype (Figure 5.25) is broken up. For example samples 2B12, Cy15 and 940 were all separated by single mutations (Figure 5.25) however, within Figure 5.27, Cy15 and 2B12 were separated by two mutations and 940 was positioned in distinct region of the network, and this level of structuring reflects the high ASD value calculated (0.625).

The predominantly North African hg 33 network increased in complexity and showed the three non-North African chromosomes (two Spanish and one Greek) as outliers. The Spanish individuals may be explained as the result of recent admixture with North Africans (Moorish occupation), however, their position as outliers suggests that they may reflect more ancient gene flow. A predominant haplotype is observed within four individuals and lies at the centre of this network, again with the other haplotypes connecting to this, which may explain the ASD value of 0.625. Two samples (Rab130 and Aga169) previously connected to the Spanish sample (Sp30 - Figure 5.25) were clear outliers separated by numerous unobserved haplotypes, however the Spanish sample was positioned in a completely different region of this network.

Finally hg 8, the most derived lineage within this part of the tree, shows a more structured network almost consisting of two sections; the upper part appears to be mainly Sub-Saharan African individuals with only two North Africans. The lower portion contains North African, SubSaharan African, and two other Asian (Indian and Pakistani) chromosomes. In summary, hg8 shows a reasonable level of diversity (ASD $=0.351$ ), with little overall geographical clustering of samples; with the exception of three East Bantu samples remaining as outliers at the tips of this network.

In general the increased number of microsatellite loci from seven up to seventeen was extremely informative and aided the improved level of sub-structuring within the networks. It has been proposed that putative ancestral haplotypes should occur at high frequency, lie near the centre of the network, and each be linked to multiple additional haplotypes. The lower frequency haplotypes should lie towards the outside of the network,
suggesting a relatively young haplogroup, because a significant proportion of the chromosomes have not yet had time to accumulate mutations at any of the loci examined (Zerjal et al., 1999). However, the networks detailed here do not appear to have a frequent common haplotype at the centre of the network, with the exception of hg 33. All the others are collections of less frequent haplotypes, and this may be due to the increased number of microsatellites examined. It must be considered that the ancestral or root haplotype may be unobserved within modern populations.

### 5.3.11. Assessing the power of assigning haplogroups from only microsatellite data

The haplogroup definitions were removed from the data encompassing 7 loci haplotype information, and all these 102 samples (the 4 hg 21 samples were removed because of incomplete DYS389II data) were used to generate a large network from 68 different haplotypes. A very complicated network is generated (Figure 5.28), with the majority of individuals being clustered together in the central region, although some degree of clustering is visible, for example amongst hg 8 and hg 34 individuals. The level of separation would probably increase if all 17 loci were included however, this amount of work was prohibitively large and was not undertaken.


Figure 5.28. Median-joining network for $\mathbf{6 2}$ haplotypes determined from 7 microsatellite loci
Each haplotype is indicated as a separate circle, and the size is uniform and does not reflect the frequency of each haplotype. The haplotypes have been coloured according to biallelic haplogroup, to determine if clusters are detectable within this network. Haplotypes shared between different haplogroups are shown with all the apprpriate haplogroup colours (e.g. haplotpye 49). Two clusters are circled with a black line. Unobserved haplotypes generated by the program are indicated as small black circles.

### 5.4. Discussion

### 5.4.1. The inclusion of new markers within the tree

The YAP+ lineage was chosen for analysis because of its unusual global distribution. The addition of new markers should hopefully always strengthen a discriminatory system. In this case seven new markers were incorporated into the existing maximum parsimony tree and this allowed the YAP lineages to be sub-divided and examined in more detail. The frequency distributions of these new haplogroups were also examined within 402 samples spanning three continents (Africa, Asia and Europe). Haplogroup 4 still remains absent from Africa as previously observed (Altheide and Hammer, 1997; Hammer et al., 1998; Karafet et al., 1999), although this could be due to a small sample size, however, this seems unlikely because this is the most comprehensive examination of the lineage to date. Other possibilities include sampling bias or effects such as drift, have removed these from their original homeland. No previously unobserved ancestral haplogroup was detected that could confirm or disprove the "out of Africa and back again hypothesis' (Altheide and Hammer, 1997) however, the most likely explanation is that this haplogroup is now extinct within modern populations and will remain unobserved.

Haplogroups still show geographical and in some cases population specificity for example, hgs 30 and 34 were only detected within Chinese and Japanese individuals respectively. Another interesting observation was the separation of the previous hg 21, into hg 21, 31, 25, 32 and 33 . Very few samples were assigned to the new hg 21 within all the samples analysed, which indicates how much of the global diversity was being masked due to the small number of markers used in previous studies.

### 5.4.2. Examination of diversity within biallelic haplogroups using microsatellite loci

Extensive studies on world wide populations have indicated a marked difference of microsatellite allele sizes when comparing different groups such as Native Americans, Asians, Europeans and Africans (Gomolka et al., 1994; Santos et al., 1996). A 'standard' set of seven loci (DYS19, DYS389I, DYS389II, DYS390, $D Y S 391, D Y S 392, D Y S 393$ ) was developed for forensic purposes, and it was suggested they would allow the discrimination ranging from $74 \%$ to $90 \%$ (within European samples) between most individuals (Kayser et al., 1997). However,
nearly complete discrimination could be achieved with the inclusion of three other loci (DYS385, YCAII, YCAIII) although, close male relatives would need further examination (Kayser et al., 1997).

Recently, it has been suggested that two loci (DYS389 and DYS390) should be sequenced directly, and each locus can be divided into four regions ( $m, n, p, q$ ) (Forster et al., 2000). When concerned with population origin, it has been proposed that due to its small size, the $D Y S 389 \mathrm{~m}$ region is highly informative, because it may retain older alleles (Forster et al., 2000). The DYS389n and q alleles do not appear to be continental specific, and it has been suggested they could obscure phylogenetic information when typed traditionally as the DYS389I/II loci (Kayser et al., 1997). The DYS390m and p regions allow discrimination between African and non-Africans, as well as between Asian, Papuans and Australians (Forster et al., 1998a; Forster et al., 2000), but the DYS390n segment was uninformative within Caucasian populations (Forster et al., 2000). Some haplotypes will be over represented within different populations, and this may cause problems, when trying to calculate statistics for forensic purposes from microsatellite data (Kayser et al., 1997).

The use of Y-chromosomal microsatellites can increase the success rate when trying to identify the male component of a mixed sample of body fluids, as well as allowing rapid screening and paternity testing in male offspring (Kayser et al., 1997). Within this study a comparison between the number of informative loci included was examined, the standard seven and a total of seventeen were used to construct networks within pre-defined biallelic haplogroups. The seventeen loci allowed much better discrimination within the network structures (see Figures 5.21 through to 5.24), and suggested that important information was not included when using only the seven standard forensic loci. However, it would be extremely unlikely that forensic laboratories would increase the number of loci used during casework when their aim is discrimination purposes compared to the aim of this work examining the relationship between populations.

### 5.4.3. Problems dating lineages and polymorphisms

The dating of lineages and mutations has always been controversial and a number of approaches are commonly used. The confidence limits generated for certain approaches suggest they are unsuitable for dating purposes because they
shed no light on the actual true date. Coalescence analysis relies upon the comparison of sequences between samples and even between humans and a suitable out-group, such as a chimpanzee, but another approach relies upon microsatellite data.

The age of the YAP+ polymorphism has been estimated as, 55-60,000 YBP (Hammer et al., 1998; Karafet et al., 1999) from coalescence analysis, although the estimate from microsatellite analysis here is only 7598 years (95\% CI 12369 4982). The striking observation is the large difference between the two ages. These differences can be compared with these calculated for another polymorphism, DYS199 (Underhill et al., 1996); an age of ~7,600 YBP was proposed for this mutation, from microsatellite analysis but is estimated as 20 30,000 YBP from coalescent analysis (Karafet et al., 1999). Coalescent dating requires a random mating population, a constant effective population size, and because of practices such as polygyny, and patrilocality, the Y-chromosome cannot be considered 'random mating'. Drift, population sub-division, and complete linkage disequilibrium of Y-chromosomal markers leads to the loss of paternal lineages (Pérez-Lezaun et al., 1999), and this has been taken to suggest that coalescent dates are probably an underestimation of age of YAP+ and an original date of $\sim 141,000$ YBP calculated from using human haplotype sequence divergence compared to human-chimp divergence (Hammer, 1995) was probably more accurate (Bravi et al., 2000). If so, the microsatellite dates are even more surprising.

### 5.4.4. The dating of $Y$-chromosomal mutations

The method chosen was ASD (average square distance) method uses the equation $(\mathrm{ASD}=\mu \mathrm{t})$ and calculates the squared mutational distance between a root haplotype and any other haplotypes within the lineage, averaged over all loci and all haplotypes (Goldstein et al., 1996). This value is divided by the mutation rate, leading to a date to the MRCA for that particular lineage in generations, and was used to date a Jewish priest specific lineage (Thomas et al., 1998). The maximum parsimony tree has been drawn in a different way (Figure 5.23 ) to show which haplogroups are defined by known mutations and which are not. This is important when trying to date mutations, because if a haplogroups is not defined by a known mutation all the derived lineages defined by mutations need to be combined to generate a date to the MRCA for the undefined haplogroup.

### 5.4.5. Deduction about the geographical distribution of $Y A P+$ lineages

All the biallelic haplogroup data and dating information from microsatellite loci were combined to produce potential explanations of how the haplogroups became distributed globally (Figure 5.29). The age of mutations that define haplogroups can be divided into two classes, ones that arose before 5,000 years ago (hgs 4, 21 and 25), and ones that arose after 5,000 years ago (hgs 30, 33, 8, 31, 34 and 32). Potential explanations such as migrations have been assigned to some of these dates, although this process is extremely speculative.

Recently, a paper suggested the prevalence of YAP+ in Asian populations may reflect a north western influence (possibly from central Asia) on the east Asian populations. It suggested that modern humans migrated to south eastern Asia from Africa, and then further migrated into China, Siberia, and southward too. It was also proposed two migrations occurred, a northerly one across the top of Siberia, and a southerly one through India towards East Asia. No suggestions were proposed to explain the large geographic gap in the YAP+ distribution (Jin and $\mathrm{Su}, 2000$ ).

### 5.4.6. Future work

Additional analysis employing networks could be undertaken using these data. For example, the parameters used to generate the median-joining networks could be altered (Forster et al., 2000; Helgason et al., 2000). The weighting of microsatellite loci due to their different mutation rates should allow a more reliable picture to emerge (Helgason et al., 2000). The networks shown within this chapter could be re-drawn after a weighting method, to see how this would alter them. MSY1 diversity within all of these samples could also be examined, to increase the amount of data available for these samples. Dating methods could also be re-done using the MSY1 data.


Figure 5.29. Continental representation of the YAP+ chromosomal data
(A) The area of the circle is proportional to the frequency of the haplogroup detected. Three continents are represented by different colours (see the key).

## Chapter 6: General Discussion

### 6.1. The Y chromosome

The Y chromosome is a useful tool for studying the human past, due to its haploid nature, paternal inheritance and lack of recombination (Ellis, 1991). Different polymorphic systems characterised by different mutation rates are contained on the non-recombining region of the $Y$ chromosome, and these include biallelic markers that can be considered 'unique', due to their extremely low mutation rate (approximately $10^{-8}$ per base per generation (Shen et al., 2000; Thomson et al., 2000). These polymorphic markers can be combined into haplogroups (Mathias et al., 1994) which show geographical and population specificity, as a result of drift and mating practices such as patrilocality. These binary markers can be combined with multiallelic markers of higher mutation rate, to examine questions on different timescales.

In this thesis, a continental survey of Y -chromosomal haplogroups has been carried out, the relationship of a linguistic isolate to its European context investigated, and the internal diversity of a major Y lineage, defined by the YAP insertion, characterised.

A number of different factors, affecting human diversity studies will be briefly discussed here, and some perspective given on the future of these studies.

### 6.2. Sampling

Despite the large number of chromosomes analysed in the European study, some areas remained underrepresented. Another potential problem is the definition of a population, as previously discussed. A general lack of Asian samples make the interpretation of results with regards to past migrations difficult. Sample sizes should be large, and as similar as possible and these studies were generally acceptable, but some were very small, for example, the Sardinian and some of the Ladin valley samples. This is important, because the general picture can be observed, but rare lineages may be missed. For example, it would be interesting to examine the internal diversity of 'Neolithic' lineages across Europe, spanning from Anatolia to Ireland. Lots of these 'Neolithic' chromosomes could be obtained easily in the south east, however, the sample
sizes needed to obtain a similar number in the west would be very large due to their lower frequency within western Europe.

### 6.3. Markers

### 6.3.1. Ascertainment bias of markers

The set of samples screened to generate new markers has a profound effect on the eventual results of diversity studies. For instance, if 100 kb of Y -chromosomal DNA were sequenced in 50 Finns to detect markers, then a subsequent survey of Europe would tend to show that Finns were the most diverse population. Although we have a lot of new markers, many populations were not represented within the screening set, and so they may appear to have low diversity, which could be an artefact. To try and remove this bias, perhaps further markers need to be ascertained using very large screening sets. Using multiallelic markers also avoids this problem to some extent, because these are variable in all populations; this has been demonstrated in Chapters 4 and 5, where microsatellites and MSY1 were used to define sublineages.

### 6.3.2. A hierarchical approach to typing - some disadvantages

In Chapters 3 and 4 of this thesis, biallelic markers were typed hierarchically to reduce work. This approach has some disadvantages, however. It might exclude very rare chromosomes, which would be interesting within these diversity studies. If a branch of the tree has several markers on it, but only one is typed, then intermediate haplotypes will not be detected. This is a problem, as these rare intermediates are important in deciding the likely geographic origin of particular lineages (see Chapter 5 for a discussion of the origins of YAP). Also, the typing of all markers reveals errors, as inconsistencies in haplotypes.

### 6.3.3. A need for parallel methods of typing

With the abundance of new markers such as SNPs detected during human genome sequencing analysis, the way they are typed must also change. Technology such as DHLPC, DNA chips and mass spectrometry (Jackson et al., 2000) will allow the greater discrimination of Y chromosomal haplogroup diversity, which may eventually lead to more
accurate and quicker population specific identifications, of interest to forensic scientists as well as anthropologists.

### 6.3.4. New $Y$-chromosomal markers now available

The isolation of new Y-chromosomal biallelic markers ( $>400$ to date) (Semino et al., 2000; Shen et al., 2000; Underhill et al., 2000), is producing a number of problems, including the need for a common nomenclature and a consensus tree, which is currently being addressed by members of the $Y$ Chromosome Consortium. A recently published tree (Semino et al., 2000; Shen et al., 2000; Underhill et al., 2000), can be seen in Figure 6.0.

### 6.4. A new study of European $Y$-chromosomal diversity

Twenty-two markers were used in a similar study to that described in Chapter 3, analysing 1007 (Mediterranean biased - with no central European, British Isles, or Scandinavian) samples, which defined ten lineages within $95 \%$ of the European samples, whose distribution was highly non-random. The overall pattern of diversity between this study (Semino et al., 2000) and ours (Rosser et al., 2000) are similar. They include some identical haplogroups - hg3, 16 - as well as others whose relationships can be approximated. However, hg2 has been broken down into 5 sub-lineages, which, as predicted in Chapter 3, show geographical coherence. The lack of definition within hg 2 was one of the reasons these data was exclude from statistical analyses (Rosser et al., 2000).

It was also suggested that there were earlier migrations along Mediterranean coast than elsewhere (Semino et al., 2000), but this may just be an artefact of the sampling bias. Migrations within other regions of Europe would be completely missed because the samples were not included. The major conclusions from the paper that that the modern population of Europe is about $80 \%$ 'Palaeolithic', relies upon the fact that $22 \%$ of European Y chromosomes belong to the 'Neolithic' lineages (Semino et al., 2000). Unfortunately due to lineage sharing between Neolithic and Palaeolithic populations, it would be extremely difficult to assign individual markers to specific pre-historic population movements.

To assign lineages to particular eras in the past, dating was carried out based on unpublished microsatellite data. The dating processes used appear


Figure 6.0. A Y-chromosomal tree incorporating more than 167 markers
The different haplogroups are indicated in different colours. The different mutations are shown on the relevant branches. This is taken from Underhill et al., 2000.
very unreliable, because single estimates for the mutation rate for the CA repeat locus (YCAII) and DYS19 were used, and most importantly no confidence interval limits were given for any of the dates. A variance-based method was employed, which generates its own confidence interval limits which also were not given. The authors assign clines to specific historical events such as migrations from distinct geographical regions, or particular archaeological cultures, such as the Aurignacian and Gravettian (Semino et al., 2000). The generation of such strict dates based on weak methodology and assignment to specific events causes great concern about its accuracy.

### 6.5. Dating

Problems determining accurate ages of mutations that define Y chromosomal lineages from microsatellite data have been seen within this thesis (Chapter 5). This uncertainty emphasizes that caution should be used when trying to connect pre-historical events and the estimated ages of mutations. Currently all dating methods employed seem unreliable. A number of studies have used different approaches to date mutations, for example, microsatellite and MSY1 data were used to date the $S R Y-2627$ mutation, and different ages were generated (Hurles et al., 1999). Another example, is the dating of the YAP lineage (hg4 and all the derived mutations), previously done using microsatellite data ( 141,000 years) (Hammer, 1995) and a coalescent based method (55-60,000 YBP) (Hammer et al., 1998; Karafet et al., 1999). A more recent approach and prehaps better called BATWING (Bayesian Analysis of trees with internal node generation) (Wilson and Balding, 1998) is a Bayesian coalescent-based method which can take into account population growth and subdivision, and appears more sophisticated than older methods.

### 6.6. The Human Genome sequence - what will be its effect?

The speed at which the human genome sequence analysis is being completed will change how this type of research will be done in the future, because of the large number of single nucleotide polymorphisms (SNPs) being detected on all chromosomes. This will allow more autosomal diversity data to be determined, which may tell us more about previous historical events. The new SNPs will have a multitude of applications such as searching for disease loci, within well characterised populations such as the Icelandic, as well as implications for the examination of population histories. Currently, 841 SNPs have been detected on the Y through the Human Genome Project, but as yet are
uncharacterised which may also need including within the new nomenclature and tree.

### 6.7. Selection

Studies are now underway to examine the effects of selection on the $Y$ chromosome. Most use a sample with a deleterious phenotype, such as subfertility, and compare haplotype distributions within this sample with a control set of samples (Jobling and Tyler-Smith, 2000). One major problem is obtaining appropriate controls (Previderé et al., 1999), especially given the high geographical stratification of the Y chromosome. However, preliminary results suggest that selective effects on modern $Y$ chromosomes may not be strong as originally thought, in which case, interpretation of $Y$ diversity in terms of population history is justified.

### 6.8. Surnames and the $Y$ chromosome

Due to the patrilinearity of the Y and surnames in many societies, and the large number of biallelic markers currently available on the Y chromosome (Semino et al., 2000; Shen et al., 2000; Underhill et al., 2000), it is now possible to examine the correlation between the Y chromosome and surnames. This has recently been undertaken on Irish samples, where statistically significant differences were found between four groups of $Y$ chromosomes each assigned to one of four Irish counties on the basis of surnames established 1000 years ago (Hill et al., 2000). Another study examined the Y-chromosomal diversity and geographical distribution of a single surname. The genetic coherence of this particular surname is probably due to the rarity of its associated haplotype in the UK, rather than the discrimination power of only four microsatellite loci (Sykes and Irven, 2000). However, this is another interesting application of the $Y$ chromosome, and will become popular as an additional method of genealogical analysis.

## References

Adams J (2000) Europe during the last 150,000 years. http://www.esd.ornl.gov/projects/qen/nercEUROPE.html:
Adams J, Otte M (2000) Did Indo-European languages spread before farming? Curr. Anthropol. 40:73-77
Altheide TK, Hammer MF (1997) Evidence for a possible Asian origin of YAP ${ }^{+}$Y chromosomes. Am. J. Hum. Genet. 61:462-466
Ammerman AJ, Cavalli-Sforza LL (1984) Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton, NJ
Ankel-Simons F, Cummins JM (1996) Misconceptions about mitochondria and mammalian fertilization: implications for theories on human evolution. Proc. Natl. Acad. Sci. USA 93:13859-13863
Armour JAL, Jeffreys AJ (1992) Recent advances in minisatellite biology. FEBS Lett. 307:113-115
Armour JAL, Harris PC, Jeffreys AJ (1993) Allelic diversity at minisatellite MS205 (D16S309) - evidence for polarized variability. Hum. Mol. Genet. 2:11371145

Armour JAL, Anttinen T, May CA, Vega EE, Sajantila A, Kidd JR, Kidd KK, Bertranpetit J, Pääbo S, Jeffreys AJ (1996) Minisatellite diversity supports a recent African origin for modern humans. Nat. Genet. 13:154-160
Ayala FJ, Escalante A, O'hUigin C, Klein J (1994) Molecular genetics of speciation and human origins. Proc. Natl. Acad. Sci. USA 91:6787-6794
Ayala FJ (1995a) The myth of Eve: molecular biology and human origins. Science 270:1930-1936
Ayala FJ (1995b) Genes and origins: the story of modern humans. J. Mol. Evol. 41:683-688
Ayub Q, Mohyuddin A, Qamar R, Mazahar K, Zerjal T, Mehdi SQ, Tyler-Smith C (2000) Identification and characterisation of novel human Y-chromosomal microsatellites from sequence database information. Nucl. Acids Res. 28:e8
Bahn PG (1998) Neanderthals emancipated. Nature 394:719-721
Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BVR, Rasanayagam A, Hammer MF (1998) Female gene flow stratifies Hindu castes. Nature 395:651-652
Bandelt H-J, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. Genetics 141:743-753

Bandelt H-J (1999) Metaphors, myths and pitfalls Presentation to the Third Biennial Euroconference of the European Human Genome Diversity Project, 'Human diversity in Europe and beyond: retrospect and prospect', Cambridge, September
Bandelt H-J, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:37-48
Bao W, Zhu S, Pandya A, Zerjal T, Xu J, Shu Q, Du R, Yang H, Tyler-Smith C (2000) MSY2: a slowly evolving minisatellite on the human Y chromosome which provides a useful polymorphic marker in Chinese populations. Gene 244:29-33
Barbujani G, Oden NL, Sokal RR (1989) Detectings regions of abrupt change in maps of biological variables. Syst. Zool. 38:376-389
Barbujani G, Sokal RR (1990) Zones of sharp genetic change in Europe are also linguistic boundaries. Proc. Natl. Acad. Sci. USA 87:1816-1819
Barbujani G (1991) What do languages tell us about human microevolution? Trends Ecol. Evol. 6:151-156
Barbujani G, Pilastro A (1993) Genetic evidence on origin and dispersal of human populations speaking languages of the Nostratic macrofamily. Proc. Natl. Acad. Sci. USA 90:4670-4673

Barbujani G, Pilastro A, De Domenico S, Renfrew C (1994) Genetic variation in North Africa and Eurasia: Neolithic Demic diffusion vs. Palaeolthic colonisation. Am. J. Phys. Anthropol. 95:137-154
Barbujani G (1997) DNA variation and language affinities. Am. J. Hum. Genet. 61:1011-1014
Barbujani G, Magagni A, Minch E, Cavalli-Sforza LL (1997) An apportionment of human DNA diversity. Proc. Natl. Acad. Sci. USA 94:4516-4519
Bates G, Lehrach H (1994) Trinucleotide repeat expansions and human genetic diseases. BioEssays 16:277-284
Batzer MA, Stoneking M, Alegria-Hartman M, Bazan H, Kass DH, Shaikh TH, Novick GE, Ioannou PA, Scheer WD, Herrera RJ, Deininger PL (1994) African origin of human-specific polymorphic Alu insertions. Proc. Natl. Acad. Sci. USA 91:12288-12292
Bertorelle G, Barbujani G (1995) Analysis of DNA diversity by spatial autocorrelation. Genetics 140:811-819
Bertranpetit J, Calafell F (1996) Genetic and geographic variability in cystic fibrosis: evolutionary considerations. In: Chadwick D, Cardew G (eds)

Variation in the human genome. John Wiley and sons, Chichester, pp 97118

Bianchi NO, Bailliet G, Bravi CM, Carnese RF, Rothhammer F, MartínezMarignac VL, Pena SDJ (1997) Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers. Am. J. Phys. Anthropol. 102:79-89
Blanco P, Shlumukova M, Sargent CA, Jobling MA, Affara N, Hurles ME (2000) Divergent outcomes of intra-chromosomal recombination on the human $Y$ chromosome: male infertility and recurrent polymorphism. J. Med. Genet. 37:752-758

Boom R, Sol CJA, Salimans MMM, Jansen CL, Wertheim-van Dillen PME, van der Noordaa J (1990) Rapid and simple method for purification of nucleic acids. J. Clin. Micro. 28:495-503

Bosch E, Calafell F, Santos FR, Pérez-Lezaun A, Comas D, Benchemsi N, TylerSmith C, Bertranpetit J (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am. J. Hum. Genet. 65:1623-1638

Bosch E, Calafell F, Pérez-Lezaun A, Clarimón J, Comas D, Mateu E, MartínezArias R, Morera B, Brakez Z, Akhayat O, Sefiani A, Hariti GC-T, A. , Bertranpetit J (2000) Genetic structure of north-west Africa revealed by STR analysis. Eur. J. Hum. Genet. 8:360-366
Bouzekri N, Taylor PG, Hammer MF, Jobling MA (1998) Novel mutation processes in the evolution of a haploid minisatellite, MSY1: array homogenization without homogenization. Hum. Mol. Genet. 7:655-659
Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455-457

Boyd R, Silk JB (1997) How humans evolved. W. W. Norton \& Co. Inc., New York, pp p469
Brassel KE, Reif D (1979) A procedure to genrate Thiessen polygons. Geogr. Anal. 11:289-303

Bravi CM, Sans M, Bailliet G, Martinez-Marignac VL, Portas M, Barreto I, Bonilla C, Bianchi NO (1997) Characterization of mitochondrial DNA and Ychromosome haplotypes in a Uruguayan population of African ancestry. Hum. Biol. 69:641-652

Bravi CM, Bailliet G, Martinez-Marignac VL, Bianchi NO (2000) Origin of YAP+ lineages of the human Y-chromosome. Am. J. Phys. Anthropol. 112:149158
Britten RJ (1996) Evolution of Alu retoposons. In: Jackson M, Strachan, T., Dover, G. (ed) Human Genome Evolution. BIOS Scientific publishers, Oxford, pp 211-227
Broglio A (1993) Mountain sites in the context of the North-east Italian Upper Palaeolithic and Mesolithic. Preistoria Alpina 28:293-310
Burgoyne PS (1982) Genetic homology and crossing over in the $X$ and $Y$ chromosomes of mammals. Hum. Genet. 61:85-90
Calafell F, Bertranpetit J (1994) Principle component analysis of gene frequencies and the origin of Basques. Am. J. Phys. Anthropol. 93:201-215
Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31-36
Cann RL, Lum IK (1991) Mitochondrial-DNA and the spread of modern people. Am. J. Hum. Genet. 49:34
Casalotti R, Simoni L, Belledi M, Barbujani G (1999) Y-chromosome polymorphisms and the origins of the European gene pool. Proc. Roy. Soc. Lond. B. 266:1959-1965
Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, Fiori G, Siniscalco M (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science 230:1403-1406
Castello-Cortes I (1998) World reference atlas. Dorling Kindersley, London, pp 731
Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J (1988) Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc. Natl. Acad. Sci. USA 85:6002-6006
Cavalli-Sforza LL (1991) Genes, peoples and languages. Scientific American:72-78
Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human evolution. Science 259:639-646
Cavalli-Sforza LL, Cavalli-Sforza F (1994) The great human diasporas: the human
Cavalli-Sforza LL, Menozzi P, Piazza A (1994) The history and Geography of human genes. Princeton University press, New Jersey
Cavalli-Sforza LL, Minch E (1997) To editor: Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am. J. Hum. Genet 61:247-251

Chai N-N, Zhou H, Hernandez J, Najmabadi H, Bhasin S, Yen PH (1998) Structure and organisation of the RBMY genes on the human $Y$ chromosome: transposition and amplification of an ancestral autosomal hnRNPG gene. Genomics 49:283-289
Champion T, Gamble C, Shennan S, Whittle A (1987) Prehistoric Europe. Academic press limited, London
Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G (1998a) Clines of nuclear DNA markers suggest a largely Neolithic ancestry of the European gene pool. Proc. Natl. Acad. Sci. USA 95:9053-9058
Chikhi L, Destro-Bisol G, Pascali V, Baravelli V, Dobosz M, Barbujani G (1998b) Clinal variation in the nuclear DNA of Europeans. Hum. Biol. 70:643-657
Ciccodicola A, D'Esposito M, Esposito T, Gianfrancesco F, Migliaccio C, Miano MG, Matarazzo MR, Vacca M, Franzè A, Cuccurese M, Cocchia M, Curci A, Terracciano A, Torino A, Cocchia S, Mercadante G, Pannone E, Archidiacono N, Rocchi M, Schlessinger D, D'Urso M (2000) Differentially regulated and evolved genes in the fully sequenced $\mathrm{Xq} / \mathrm{Yq}$ pseudoautosomal region. Hum. Mol. Genet. 9:395-401
Collins R (1986) The Basques. Blackwell, Oxford, pp100
Comas D, Calafell F, Mateu E, Perez-Lezaun A, Bosch E, Bertranpetit J (1997) Mitochondrial DNA variation and the origin of the Europeans. Hum. Genet. 99:443-449
Cooke HJ (1976) Repeated sequence specific to human males. Nature 262:182-186
Cooke HJ, Brown WRA, Rappold G (1985) Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. Nature 317:687692
Cooper G, Amos W, Hoffman D, Rubinsztein DC (1996) Network analysis of human Y microsatellite haplotypes. Hum. Mol. Genet. 5:1759-1766
Crouau-Roy B, Service S, Slatkin M, Freimer N (1996) A fine-scale comparison of the human and chimpanzee genomes: Linkage. linkage disequilibrium and sequence analysis. Hum. Mol. Genet 5:1131-1137
Cunliffe B (1998) Prehistoric Europe an illustrated history. Oxford University Press, Oxford, pp 496
de Knijff P, Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, et al (1997) Chromosome $Y$ microsatellites: population genetic and evolutionary aspects. Int. J. Legal Med. 110:134-140

Deitrich W, Katz H, Lincoln SE, Shin HS, Friedman J, Dracopoli N, Lander AS (1992) A genetic map of the mouse suitable for typing intraspecific crosses. Genetics 131:423-447
Dennell R (1983) European economic prehistory: a new approach. Academic Press, London
Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc. Natl. Acad. Sci. USA 91:3166-3170
Di Rienzo A, Donnelly P, Toomajian C, Sisk B, Hill A, Petzl-Erler ML, Haines GK, Barch DH (1998) Heterogeneity of microsatellite mutations within and bewteen loci, and implications for human demographic histories. Genetics 148:1269-1284

Dolgopolvsky AB (1988). In: Borg A, Wexler P (eds) Mediterranean language review. Vol. 3. Harrassowitz, Wiesbaden, Germany, pp 7-31
Dyen I, Kruskal JB, Black P (1992) An Indoeuropean classification: a lexicostatistical experiment. Trans. Am. Phil. Soc. 82:1-132
Ellis N, Goodfellow PN (1989) The mammalian pseudoautosomal region. Trends Genet. 5:406-410

Ellis NA (1991) The human Y chromosome. Sem. Devel. Biol. 2:231-240
Europe Co (1998) Implementation of the European charter for regional or minority languages Congress of local and regional authorities of Europe and the secretariat of the European charter for regional or minority languages in collaboration with the city of Innsbruck. Council of Europe publishing, Starsbourg
Flight CR (1988) The Bantu expansion and the SOAS network. Hist. Africa 15:261301
Forster P, Harding R, Torroni A, Bandelt H-J (1996) Origin and evolution of Native American mtDNA variation: a reappraisal. Am. J. Hum. Genet. 59:935-945

Forster P, Kayser M, Meyer E, Roewer L, Pfeiffer H, Benkmann H, Brinkmann B (1998a) Phylogenetic resolution of complex mutational features at Y-STR DYS390 in aboriginal Australians and Papuans. Mol. Biol. Evol. 15:11081114
Forster P, Toth A, Bandelt H-J (1998b) Evolutionary network analysis of word lists: visualising the relationship between alpine romance languages. J. Quant. Ling. 5:174-187

Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B (2000) A short tandem repeat-based phylogeny for the human Y chromosome. Am. J. Hum. Genet. 67:182-196
Fretwell N (1996) A search for human Y-chromosome-specific minisatellites and microsatellites Ph.D thesis, University of Leicester
Fu Y-X, Li W-H (1997) Estimating the age of the common ancestor of a sample of DNA sequeces. Mol. Biol. Evol. 14:195-199
Gagneux P, Wills C, Woodruff DS (1999) Mitochondrial sequences show diverse evolutionary histories of African hominids. Proc. Natl. Acad. Sci. USA 96:5077-5082

Gimbutas M (1970) Proto-Indo-European culture: The Kurgan culture during the fifth, fourth, and third millenia B. C. In: G. Cardona, H. M. Hoenigswald, Senn A (eds) Indo-European and Indo-Europeans. University of Pennsylvania Press, Philadelphia, pp 155-195
Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463-471
Goldstein DB, Zhivotovsky LA, Nayar K, Linares AR, Cavalli-Sforza LL, Feldman MW (1996) Statistical properties of the variation at linked microsatellite loci: implications for the history of human Y chromosomes. Mol. Biol. Evol. 13:1213-1218
Gomolka M, Hundrieser J, Nurnberg P, Roewer L, Epplen JT, Epplen C (1994) Selected dinucleotide and tetranucleotide microsatellites from chromosome 7, chromosome 12, chromosome 14, and chromosome Y in various Eurasian populations. Hum. Genet. 93:592-596
Graves JAM (1995) The origin and function of the mammalian Y chromosome and Y-bourne genes - an evolving understanding. BioEssays 17:311-321
Greenberg JH (1963) The languages of Africa. Indiana University Publication, Bloomington
Guthrie M (1967) Comparative Bantu: An introduction to the comparative linguistics and prehistory of Bantu languages. Vol. 1. Gregg Press, England, Farnborough
Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhovel W, Clegg JB, Bowden DK (1999) Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Phil. Trans. Roy. Soc. Lond. B 266:485-492
Hagelberg E, Goldman N, Lio P, Whelan S, Schiefenhovel W, Clegg JB, Bowden DK (2000) Evidence for mitochondrial DNA recombination in a human
population of island Melanesia: correction. Phil. Trans. Roy. Soc. Lond. B 267:1595-1596
Hammer M, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, SantachiaraBenerecetti S, Oppenheim A, Jobling MA, Jenkins T, Ostrer H, BonnéTamir B (2000) Jewish and middle eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. Am. J. Hum. Genet. 97:6769-6774
Hammer MF (1994) A recent insertion of an Alu element on the $Y$ chromosome is a useful marker for human population studies. Mol. Biol. Evol. 11:749-761
Hammer MF, Bonner MR (1994) Evolutionary analysis of polymorphisms in the YAP region of the human Y chromosome. Cytogenet. Cell Genet. 67:395396

Hammer MF (1995) A recent common ancestry for human Y chromosomes. Nature 378:376-378

Hammer MF, Horai S (1995) Y-chromosomal DNA variation and the peopling of Japan. Am. J. Hum. Genet. 56:951-962
Hammer MF, Zegura SL (1996) The role of the $Y$ chromosome in human evolutionary studies. Evol. Anthropol. 5:116-134
Hammer MF, Spurdle AB, Karafet T, Bonner MR, Wood ET, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, Zegura SL (1997) The geographic distribution of human Y chromosome variation. Genetics 145:787-805

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL (1998) Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation. Mol. Biol. Evol 15:427-441
Hancock JM (1996) Microsatellites and other simple sequences in the evolution of the human genome. In: Jackson M, Strachan T, Dover GA (eds) Human genome evolution. BIOS Scientific publishers, Oxford
Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, Clegg JB (1997) Archaic African and Asian lineages in the genetic ancestry of modern humans. Am. J. Hum. Genet. 60:772-789
Harpending H, Jenkins T (1973) Genetic distance among Southern African populations. In: Crawford MH, Workman PL (eds) Methods and theories of anthropologcal genetics. Uniersity of new Mexico Press, Albuquerque, pp 177-199

Hassan FA (1973) On mechanisms of population growth during the Neolithic. Curr. Anthropol. 14:535-542
Hawkes K, O'Connell JF, Rogers L (1997) The behavioural ecology of modern hunter-gatherers, and human evolution. Trends Evol. Ecol. 12:29-32
Helgason A, Sigurdardóttir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, Gulcher JR, Ward R, Stefánsson K (2000) Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. Am. J. Hum. Genet. 67:697717
Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum. Mol. Genet. 6:799-803
Hill E, Bradley D, Jobling MA (2000) Y-chromosome variation and Irish origins. Nature 404:351-352
Horai S, Satta Y, Hayasakak K, Kondo R, Inoue T, Ishida T, Hayashi S, Takahata N (1992) Man's place in Hominoidea revealed by mitochondrial DNA genealogy. J. Mol. Evol. 35:32-43
Horai S, Hayasaka K, Kondo R, Tsugane K, Takahata N (1995) Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. Proc. Natl. Acad. Sci. USA 92:532-536
Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J. Educ. Pyschol. 24:417-441, 498-520
Hurles ME, Irven C, Nicholson J, Taylor PG, Santos FR, Loughlin J, Jobling MA, Sykes BC (1998) European Y-chromosomal lineages in Polynesians. A contrast to the population structure revealed by mtDNA. Am. J. Hum. Genet. 63:1793-1806
Hurles ME (1999) Mutation and variability of the human Y chromosome Department of Genetics. Ph. D thesis, University of Leicester
Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Pérez-Lezaun A, Bosch E, Shlumukova M, Cambon-Thomsen A, McElreavey K, López de Munain A, Röhl A, Wilson IJ, Singh L, Santos FR, Tyler-Smith C, Jobling MA (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. Am. J. Hum. Genet. 65:1437-1448
Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nat. Genet. 12:417-420

Jackson PE, Scholl PF, Groopman JD (2000) Mass spectrometry for genotyping: an emerging tool for molecular medicine. Mol. Med. Today 6:271-276
Jeffreys AJ, Wilson V, Thein SL (1985) Individual-specific 'fingerprints' of human DNA. Nature 316:76-79
Jeffreys AJ, Neumann R, Wilson V (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell 60:473-485
Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG (1991) Minisatellite repeat coding as a digital approach to DNA typing. Nature 354:204-209
Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JAL (1994) Complex gene conversion events in germline mutation at human minisatellites. Nat. Genet. 6:136-145
Jeffreys AJ, Murray J, Neumann R (1998) High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. Mol. Cell 2:267-273
Jin L, Zhong Y, Shriver MD, Deka R, Chakraborty R (1994) Distribution of repeat unit differences between alleles at tandem repeat microsatellite loci. Am. J. Hum. Genet. 55:S39 (Abstract)
Jin L, Su B (2000) Natives or immigrants: modern human origin in East Asia. Nature reviews Genetics 1:126-133
Jobling MA, Fretwell N, Dover GA, Jeffreys AJ (1994) Digital coding of human Y chromosomes - MVR-PCR at Y-specific minisatellites. Cytogenet. Cell Genet. 67:390
Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. Trends Genet. 11:449-456

Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int. J. Legal Med. 110:118-124
Jobling MA, Bouzekri N, Taylor PG (1998a) Hypervariable digital DNA codes for human paternal lineages: MVR-PCR at the Y-specific minisatellite, MSY1 (DYF155S1). Hum. Mol. Genet. 7:643-653

Jobling MA, Williams G, Schiebel K, Pandya A, McElreavey K, Salas L, Rappold GA, Affara NA, Tyler-Smith C (1998b) A selective difference between human Y-chromosomal DNA haplotypes. Curr. Biol. 8:1391-1394
Jobling MA, Tyler-Smith C (2000) New uses for new haplotypes: the human Y chromosome, disease, and selection. Trends. Genet. 16:356-362
Jones J, Martin R, Pilbean D (1999) The Cambridge encylopedia of human evolution. In: Bunney S (ed). Cambridge University Press, Cambridge

Jones JS (1991) Farming is in the blood. Nature 351:97-98
Jorde LB, Rogers AR, Bamshad M, Watkins WS, Krakowiak P, Sung S, Kere J, Harpending HC (1997) Microsatellite diversity and the demographic history of modern humans. Proc .Natl. Acad. Sci. USA 94:3100-3103
Jorde LB, Bamshad M, Rogers AR (1998) Using mitochondrial and nuclear DNA markers to reconstruct human evolution. BioEssays 20:126-136
Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of new world Ychromosome founder haplotypes. Am. J. Hum. Genet. 64:817-831
Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, et al (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int. J. Legal Med. 110:125-133
Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human $Y$ chromosome, as revealed by direct observation in father/son pairs. Am. J. Hum. Genet 66:1580-1588
Krings M, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Pääbo S (1997) Neandertal DNA sequences and the origins of modern humans. Cell 90:1930

Kwok C, Tyler-Smith C, Medonca BB, Hughes I, Berkovitz GD, Goodfellow PN, Hawkins JR (1996) Mutation analysis of 2kb 5' to SRY in XY females and XX intersex subjects. J. Med. Genet. 33:465-468
Lagercrantz U, Ellegren H, Anderson L (1993) The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucl. Acids Res. 21:1111-1115

Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula P, Peltonen L, Savontaus M-L (1996) The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. Am. J. Hum. Genet. 58:1309-1322
Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. Science 286:964-967

Landers J (1992) Reconstructing ancient populations. In: Jones S, Martin R, Pilbeam D (eds) The Cambridge Encylopedia of Human Evolution. Cambridge University Press, Cambridge, pp402-405

Langaney A, Roessli D, van Blyenburgh NH, Dard P (1992) Do most human populations descend from phylogenetic trees? Hum. Evol. 7:47-61
Lewin B (1990) Genes IV. Oxford University Press, Oxford, pp 1-803
Lucotte G, Loirat F (1999) Y-chromosome DNA haplotype 15 in Europe. Hum. Biol. 71:431-437
Lyon MF (1961) Gene action in the X-chromosome of the mouse (Mus musculus 1.). Nature 190:372-373

Magenis RE, Brown MG, Dolon T, Olson SB, Sheehy R, Tomar D (1985) Structural aberrations of the Y chromosome, including the nonfluorescent Y : cytologic origin and consequences. In: Sandberg AA (ed) The Y chromosome, part a: basic characteristics of the Y chromosome. A. R. Liss, New York, pp537-574
Malaspina P, Cruciani F, Ciminelli BM, Terrenato L, Santolamazza P, Alonso A, Banyko J, Brdicka R, Garcia O, Gaudiano C, Guanti G, Kidd KK, Lavinha J, Avila M, Mandich P, Moral P, Qamar R, Mehdi SQ, Ragusa A, Sefanescu G, Caraghin M, Tyler-Smith C, Scozzari R, Novelletto A (1998) Network analyses of Y-chromosomal types in Europe, Northern Africa, and Western Asia reveal specific patterns of geographic distribution. Am. J. Hum. Genet. 63:847-860

Mantel NA (1967) The detection of disease clustering and a generalised regressional approach. Cancer Res. 27:209-220
Mathias N, Bayés M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. Hum. Mol. Genet. 3:115-123
McKeigue PM (1997) Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. Am. J. Hum. Genet. 60:188-196
Mellars P (1998) The fate of the Neanderthals. Nature 404:539-540
Menozzi P, Piazza A, Cavalli-Sforza LL (1978) Synthetic maps of human gene frequencies in Europeans. Science 201:786-792
Mitchell RJ, Hammer MF (1996) Human evolution and the Y chromosome. Curr. Opin. Genet. Dev. 6:737-742
Mullis K, Faloona F (1987) Specific synthesis of DNA in vitro via polymerase catalysed chain reaction. Meth. Enzymol. 55:335-350
Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583-590

Nei M, Graur D (1984) Extent of protein polymorphism and the neutral mutation theory. Evol. Biol 17:73-118

Nei M (1987) Molecular Evolutionary Genetics. Columbia University Press, New York
Nei M, Livshits G (1989) Genetic relationship of Europeans, Asians and Africans and the origins of modern Homo-Sapiens. Hum. Hered. 39:276-281
Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J (1986) A DNA probe detecting multiple haplotypes of the human Y chromosome. Am. J. Hum. Genet. 38:407-418
Okada N (1991) SINEs. Current Opinion in Genetics and Development 1:498-504
Ovchinnikov IV, Götherström A, Romanova GP, Kharitonov VM, Liden K, Goodwin W (2000) Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 404:490-493

Page DC (1986) Sex reversal: deletion mapping the male-determining function of the human Y chromosome. Cold Spring Harb. Symp. Quant. Biol. 51:229235

Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner R, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM (1997) A high observed substitution rate in the human mitochondrial DNA control region. Nat. Genet. 15:363-367
Pearson PL, Bobrow M (1970) Definitive evidence for the short arm of the Y chromosome associating with the X during meiosis in the human male. Nature 226:959-961
Pérez-Lezaun A, Calafell F, Mateu E, D. C, Ruiz-Pacheco R, Bertranpetit J (1997a) Microsatellite variation and the difference of modern humans. Hum. Genet 99:1-7

Pérez-Lezaun A, Calafell F, Seielstad M, Mateu E, Comas D, Bosch E, Bertranpetit J (1997b) Population genetics of Y-chromosome short tandem repeats in humans. J. Mol. Evol. 45:265-270
Pérez-Lezaun A, Calafell F, Comas D, Mateu E, Bosch E, Martinez-Arias R, Clarimón J, Fiori G, Luiselli D, Facchini F, Pettener D, Bertranpetit J (1999) Sex-specific migration patterns in central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. Am. J. Hum. Genet. 65:208-219
Piazza A (1993) Who are the Europeans? Science 260:1767-1769
Piazza A, Rendine S, Minch E, Menozzi P, Mountain J, Cavalli-Sforza LL (1995) Genetics and the origin of European languages. Proc. Natl. Acad. Sci. USA 92:5836-5840

Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup I, Langaney A, Excoffier L (1997) Human genetic affinities for Ychromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. Am. J. Hum. Genet. 61:1015-1035
Previderé C, Stuppia L, Gatta V, Fattorini P, Palka G, Tyler-Smith C (1999) Ychromosomal DNA haplotype differences in control and infertile Italian subpopulations. Eur. J. Hum. Genet. 7:733-736
Quintana-Murci L, Semino O, Minch E, Passarino G, Brega A, SantachiaraBenerecetti AS (1999) Further charactersitics of proto-European Y chromosomes. Eur. J. Hum. Genet. 7:603-608
Rendine S, Piazza A, Cavalli-Sforza LL (1986) Simulation and separation by principal components of multiple demic expansion in Europe. Am. Naturalist 128:681-706
Renfrew C (1987) Archaeology and language: the puzzle of Indo-European origins. Cambridge University press, New York
Renfrew C (1989) The origins of Indo-European languages. Sci. Am. 261:106-114
Renfrew C (1994) World linguistic diversity. Sci. Am. 270:104-110
Renfrew C (2000) At the edge of knowability: towards a prehistory of languages. Camb. Archaeol. J. 10:7-34
Riccio ML, Rossolini GM (1993) Unusual clustering of Alu repeats within the 5'flanking region of the human lysozyme gene. DNA sequence 4:129-134
Richards M, Côrte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt H-J, Sykes B (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am. J. Hum. Genet. 59:185-203
Richards M, Macaulay V, Sykes B, Pettitt P, Hedges R, Forster P, Bandelt H-J (1997) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. Am. J. Hum. Genet. 61:247-251
Richards M, Sykes B (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. Am. J. Hum. Genet. 62:491-492
Roewer L, Arnemann J, Spurr NK, Grzeschik KH, Epplen JT (1992) Simple repeat sequences on the human $Y$ chromosome are equally polymorphic as their autosomal counterparts. Hum. Genet. 89:389-394
Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Knijff P (1996) Analysis of molecular variance (AMOVA) of Y-chromosomespecific microsatellites in two closely related human populations. Hum. Mol. Genet. 5:1029-1033

Rogers J, Samollow PB, Comuzzie AG (1996) Estimating the age of the common ancestor of men from the $Z F Y$ intron. Science 272:1360-1361
Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, et al (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet. 67:1526-1543
Royle NJ, Clarkson RE, Wong Z, Jeffreys AJ (1988) Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. Genomics 3:352-360
Ruhlen M (1991) A guide to the world's languages. Volume 1: Classification. Stanford University Press, Stanford, California
Ruiz Linares A, Nayar K, Goldstein DB, Hebert JM, Seielstad MT, Underhill PA, Lin AA, Feldman MW, Cavalli Sforza LL (1996) Geographic clustering of human Y-chromosome haplotypes. Ann. Hum. Genet. 60:401-408
Saiki RK, Scharf SJ, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enymatic amplification of beta-globin sequences and restriction site analysis for diagnosis of sickle cell anaemia. Science 230:1350-1354
Sajantila A, Pääbo S (1995) Language replacement in Scandinavia. Nat. Genet. 11:359-360

Santos FR, Gerelsaikhan T, Munkhtuja B, Oyunsuren T, Epplen JT, Pena SDJ (1996) Geographic differences in the allele frequencies of the human Ylinked tetranucleotide polymorphism DYS19. Hum. Genet. 97:309-313
Santos FR, Tyler-Smith C (1996) Reading the human Y chromosome: the emerging DNA markers and human genetic history. Braz. J. Genet. 19:665670
Santos FR, Carvalho-Silvo DR, Pena SDJ (1999) PCR-based DNA profiling of human Y chromosomes. In: Epplen JT, Lubjuhn T (eds) Methods and tools in biosciences and medicine. Birkhaüser Verlag, Basel, pp 133-152
Satta Y, O'hUigin C, Takahata N, Klein J (1994) Intensity of natural-selection at the major histocompatibility complex loci. Proc. Natl. Acad. Sci. USA 91:7184-7188
Schmidt K, Lazzeroni LC, Foote S, Vollrath D, Fisher EMC, Goradia TM, Lange K, Page DC, Arnheim N (1994) Multipoint linkage map of the human pseudoautosomal region, based on single-sperm typing - do double crossovers occur during male meiosis? Am. J. Hum. Genet. 55:423-430

Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva, Switzerland
Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, CavalliSforza LL (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to $G$ transition. Hum. Mol. Genet. 3:2159-2161
Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat. Genet 20:278-280
Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS (1996) A view of the Neolithic demic diffusion in Europe through two $Y$ chromosome-specific markers. Am. J. Hum. Genet. 59:964-968
Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatasi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Palaeolithic Homo sapiens sapiens in extant Europeans: A Y chromosome perspective. Science 290:1155-1159
Shen P, Wang F, Underhill PA, Franco C, Yang W-H, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (2000) Population genetic implications from sequence variation in four $Y$ chromosomes genes. Proc. Natl. Acad. Sci. USA 97:7354-7359

Shriver MD, Jin L, Ferrell RE, Deka R (1997) Microsatelite data suport an early population expansion in Africa. Genome Res. 7:586-591
Simmler M-C, Rouyer F, Vergnaud G, Nyström-Lahti M, Ngo KY, de la Chapelle A, Weissenbach J (1985) Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. Nature 317:692-697
Simoni L, Gueresi P, Pettener D, Barbujani G (1999) Patterns of gene flow inferred from genetic distances in the Mediterranean region. Hum. Biol. 71:399-415
Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G (2000a) Geographic patterns of mtDNA diversity in Europe. Am. J. Hum. Genet. 66:262-278
Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G (2000b) Reconstruction of prehistory on the basis of genetic data. Am. J. Hum. Genet. 66:1177-1179
Slagel V, Flemington E, Traina-Dorge V, Bradshaw Hj, Deininger PL (1987) Clustering and subfamily relationships of the Alu family in the human genome. Mol. Biol. Evol. 4:19-29

Smit A, F, A. (1996) The origin of interspersed repeats in the human genome. Curr. Opin. Genet. Devel. 6:743-748
Sokal RR, Oden NL (1978) Spatial autocorrelatin analysis in biology. Biol. J. Linn. Soc. 10:199-249
Sokal RR, Oden NL, Thomson BA (1988) Genetic changes across language boundaries in Europe. Am. J. Phys. Anthropol 76:337-361
Sokal RR, Harding RM, Oden NL (1989) Spatial patterns of human gene frequencies in Europe. Am. J. Phys. Anthropol. 80:267-294
Sokal RR, Oden NL, Legendre P, Fortin MJ, Kim J, Thomson BA, Vaudor A, Harding RM, Barbujani G (1990) Genetics and language in European populations. Am. Nat. 135:157-175
Sokal RR, Oden NL, Wilson C (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. Nature 352:143-145
Sokal RR, Oden NL, Thomson BA (1997) A simulation study of microevolutionary inferences by spatial autocorrelation analysis. Biol. J. Linn. Soc. 60:73-93
Solari AJ (1980) Synaptonemal complexes and associated structures in microspread human spermatocytes. Chromosoma 81:315-337
Spurdle AB, Hammer MF, Jenkins T (1994) The Y-Alu polymorphism in Southern African populations and its relationship to other $Y$-specific polymorphisms. Am. J. Hum. Genet. 54:319-330
Stenico M, Nigro L, Bertorelle G, Calafell F, Capitanio M, Corrain C, Barbujani G (1996) High mitochondrial sequence diversity in linguistic isolates of the Alps. Am. J. Hum. Genet. 59:1363-1375
Stenico M, Nigro L, Barbujani G (1998) Mitochondrial lineages in Ladin-speaking communities of the eatern Alps. Proc. R. Soc. Lond. B 265:555-561
Stoneking M, Jorde LB, Bhatia K, Wilson AC (1990) Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124:717733
Stoneking M (1993) DNA and recent human evolution. Evol. Anthropol. 2:60-73
Stoneking M, Soodyall H (1996) Human evolution and the mitochondrial genome. Curr. Op. Genet. Devel. 6:731-736
Stoneking M (1998) Women on the move. Nat. Genet 20:219-220
Stringer CB (1992). In: Bräuer GS, F. H (ed) Continuity or replacement: Controversies in Homo sapiens evolution. Balkema, Rotterdam, pp 9-23

Sullivan KM, Mannucci A, Kimpton CP, Gill P (1993) A rapid and quantitative DNA sex test - fluorescence-based PCR analysis of X-Y homologous gene amelogenin. Biotechniques 15:636-641
Sun C, Skaletsky H, Birren B, Devon K, Tang Z, Silber S, Oates R, Page DC (1999) An azoospermic man with a de novo point mutation in the Y-chromosomal gene USP9Y. Nat. Genet. 23:429-432

Swadesh M (1952) Lexico-statistics dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. Proc. Am. Phil. Soc. 96:452-463

Swadesh M (1955) Towards greater accuracy in lexicostatistic dating. J. Am. Ling. 21:121-137

Swisher III CC, Curtis GH, JAcob T, Getty AG, Suprijo A (1994) Age of the earliest known Hominids in Java, Indonesia. Science 263:1118
Sykes B, Corte-Real H, Richards M (1996) Palaeolithic and neolithic contributions to the European gene pool. In: Boyce AJ, Mascie-Taylor CGN (eds) Molecular biology and human diversity. Cambridge University Press, Cambridge, pp 130-140
Sykes B (1999) The molecular genetics of European ancestry. Phil. Trans. R. Soc. Lond. B 354:131-139

Sykes B, Irven C (2000) Surnames and the Y chromosome. Am. J. Hum. Genet. 66:1417-1419
Takahata $\mathbf{N}$ (1991) Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. Genetics 129:585-595
Takahata N (1993) Allelic genealogy and human evolution. Mol. Biol. Evol. 10:222

Templeton AR (1993) The "Eve" hypothesis: a genetic critique and reanalysis. Am. Anthropol. 95:51-72
Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, Ambystoma tigrinum. Genet. 140:767-782
Templeton AR (1997) Out of Africa? What do genes tell us? Curr. Opin. Genet. Devel. 7:841-847
Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB (1998) Origins of Old Testament priests. Nature 384:138-140

Thomas MG, Bradman N, Flinn HM (1999) High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y chromosome. Hum. Genet. 6:577-681
Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancetry of human $Y$ chromosomes: Evidence from DNA sequence data. Proc. Natl. Acad. Sci. USA 97:7360-7365
Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonné-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M, Pääbo S, Watson E, Risch N, Jenkins T, Kidd KK (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271:1380-1387

Tishkoff SA, Pakstis AJ, Stoneking M, Kidd JR, Destro-Bisol G, Sanjantila A, Lu R-B, Deinard AS, Sirugo G, Kidd KK, Clark AG (2000) Short tandemrepeat polymorphism/Alu haplotype variation at the PLAT locus: implications for modern human evolution. Am. J. Hum. Genet. 67:901-925
Torroni A, Huoponen K, Franacalacci P, Petrozzi M, Morelli L, Scozzari R, Obinu D, Savontaus M-L, Wallace DC (1996) Classification of European mtDNAs from an analysis of three European populations. Genetcis 144:1835-1850
Torroni A, Bandelt H-J, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus M-L, Bonne-Tamir B, Scozzari R (1998) mtDNA analysis reveals a major late paleolithic population expansion from Southwestern to Northeastern Europe. Am. J. Hum. Genet. 62:1137-1152
Torroni A, Richards M, Macaulay V, Forster P, Villems R, Nørby S, Savontaus ML, Huoponen K, Scozzari R, Bandelt H-J (2000) mtDNA haplogroups and frequency patterns in Europe. Am. J. Hum. Genet. 66:1173-1177
Tremblay M, Vézina H (2000) New estimates of integenerational time intervals for the calculation of age and origins of mutations. Am. J. Hum. Genet. 66:651-658

Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL (1996) A preColumbian $Y$ chromosome-specific transition and its implications for human evolutionary history. Proc. Natl. Acad. Sci. USA 93:196-200
Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, CavalliSforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res. 7:996-1005
Underhill PA, Shen P, Lin AA, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ,

Seielstad MT, Wells SR, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y chromosome sequence variation and the history of human populations. Nat. Genet. 26:358-361
van Andel TH, Runnels CN (1995) The earliest farmers in Europe. Antiquity 69:481-500
Veitia R, Ion A, Barbaux S, Jobling MA, Souleyreau N, Ennis K, Ostrer H, Tosi M, Meo T, Chibani J, Fellous M, McElreavey K (1997) Mutations and sequence variants in the testis-determining region of the Y chromosome in individuals with a 46, XY female phenotype. Hum. Genet. 99:648-652
Vergnaud G, Mariat D, Apiou F, Aurias A, Lathrop M, Lauthier V (1991) The use of synthetic tandem repeats to isolate new VNTR loci: cloning of a human hypermutable sequence. Genomics 11:135-144
Ward R, Stringer C (1997) A molecular handle on the Neanderthals. Nature 388:225-226
Wardhaugh $R$ (1987) Languages in competition: dominence, diversity and decline. Basil Blackwell Ltd, Oxford, pp 1-275
Watson JD, Gilman M, Witkowski J, Zoller M (1992) Recombinant DNA. W. H. Freeman and Company, New York, pp 1-616
Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum. Mol. Genet. 2:1123-1128

Weiss G, von Haeseler A (1996) Estimating the age of the common ancestor of men from the $Z F Y$ intron. Science 272:1359

Weissenbach J, Levilliers J, Petit C, Rouyer F, Simmler M-C (1987) Normal and abnormal interchanges between the human $X$ and $Y$ chromosomes. Development (supplement) 101:67-74
Weissenbach J, Gyapay G, Dib C, Vignal A, Morisette J, Millaseau P, Vaysseix G, Lathrop M (1992) A second generation linkage map of the human genome. Nature 359:794-801

White PS, Tatum OL, Deaven LL, Longmire JL (1999) New, male-specific microsatellite markers from the human Y chromosome. Genomics 57:433437
Whitfield LS, Hawkins TL, Goodfellow PN, Sulston J (1995a) 41 kilobases of analyzed sequence from the pseudoautosomal and sex-determining regions of the short arm of the human $Y$ chromosome. Genomics 27:306311
Whitfield LS, Sulston JE, Goodfellow PN (1995b) Sequence variation of the human $Y$ chromosome. Nature 378:379-380

Willis KJ, Bennett KD (1994) The neolithic transition - fact or fiction? Palaeoecological evidence from the Balkans. Holocene 4:326-330
Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. Genetics 150:499-510

Wolpoff MH (1989) The human revolution: Behavioural and Biological perspectives on the origins of modern humans. In: Mellers PS, C. B (ed) The human revolution: Behavioural and Biological perspectives on the origins of modern humans. Vol. 1. Edinburgh University press, Edinburgh, pp 62-108
Wolpoff MH (1992). In: Bräuber GaS, F. H (ed) Continuity or replacement: Controversies in Homo sapiens evolution. Balkema, Rotterdam, pp 25-63
Womble WH (1951) Differential systematics. Science 114:315-322
Wood B (1996) Human evolution. BioEssays 18:945-954
Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C (1997) Genetic relationships of Asians and northern Europeans, revealed by Ychromosomal DNA analysis. Am. J. Hum. Genet. 60:1174-1183

Zerjal T, Pandya A, Santos FR, Adhikari R, Tarazona E, Kayser M, Evgrafov O, Singh L, Thangaraj K, Destro-Bisol G, Thomas MG, Qamar R, Medhi QS, Rosser ZH, Hurles ME, Jobling MA, Tyler-Smith C (1999) The use of Ychromosomal DNA variation to investigate population history. In: Papiha, Deka, Chakraborty (eds) Genomic Diversity: Applications in Human population Genetics. Kluwer Academic / Plenum Publishers, New York, pp 91-101
Zischler H, Geisert H, von Haeseler A, Pääbo S (1995) A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. Nature 378:489192
Zvelebil M (1986) Mesolithic prelude and neolithic revolution. In: Zvelebil M (ed) Hunters in transition: mesolithic societies of temperate Eurasia and their transition to farming. Cambridge University Press, Cambridge, pp 5-15
Zvelebil M, Zvelebil KV (1988) Agricultural transition and Indo-European dispersal. Antiquity 62:574-583

| Greenlandic Inuit |  |  |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| 570 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 572 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 573 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 574 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 575 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 576 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 577 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 578 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 579 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 580 | DG | nr | 0 | 0 | nr | 0 | nd | nd | nd | nd | 0 | 0 | nr |
| 581 | DG | nr | 1 | 0 | nr | 0 | nd | nd | nd | nd | 1 | 0 | 1 or 3 |
| 582 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 583 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 589 | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 610 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 611 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 612 | DG | nr | 0 | nr | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 613 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 614 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 616 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 617 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 618 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 619 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 620 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 650 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 651 | DG | nr | 0 | 0 | 1 | 1 | 1 | 0 | nd | nd | 0 | 0 | 21 |
| 652 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 653 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 654 | DG | nr | 1 | nr | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 655 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 656 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 657 | DG | nr | 0 | 0 | 1 | nr | nd | nd | nd | nd | 0 | 0 | nr |
| 658 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 659 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 660 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 670 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 671 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 672 | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 673 | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 674 | DG | nr | nr | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 675 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 676 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 677 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 678 | DG | nr | 0 | nr | nr | 0 | nd | nd | nd | nd | 0 | 0 | nr |
| 679 | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 680 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 681 | DG | nr | 0 | nr | 1 | nr | nd | nd | nd | nd | 0 | nr | nr |
| 682 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 683 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 684 | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 690 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 691 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 692 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 693 | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 694 | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 695 | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 696 | DG | nr | 0 | 0 | nr | nr | nd | nd | nd | nd | 0 | nr | nr |
| 697 | DG | nr | 1 | 0 | nr | 0 | nd | nd | nd | nd | 1 | 0 | 1 or 3 |
| 698 | DG | nr | 1 | 0 | 1 | nr | nd | nd | nd | nd | 1 | 0 | 1 |
| 699 | DG | nr | 1 | 0 | 1 | nr | nd | nd | nd | nd | 1 | nr | 1 |
| 700 | DG | nr | 1 | 0 | 1 | nr | nd | nd | nd | nd | 1 | 0 | 1 |
| 701 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 702 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 703 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12 f 2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 704 | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 1 | 9 |
| 710 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 711 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 712 | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 713 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 714 | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| Danish |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | s Y 81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| 113 | LDSD | 0 | 0 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 0 | 0 | 2 |
| 104 | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 40c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 46 c | LDSD | 0 | 0 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 0 | 0 | 2 |
| 49c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 57c | LDSD | 0 | 0 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 0 | 0 | 2 |
| 58c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 73c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 82c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 84c | LDSD | 0 | 0 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 0 | 0 | 2 |
| 90 c | LDSD | nr | nr | nr | 1 | 0 | 0 | nd | 0 | nr | 1 | 0 | 1 ? |
| 94 c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 97c | LDSD | 0 | 0 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 0 | 1 | 9 |
| 104c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 106c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 113c | LDSD | 1 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 3 |
| 114c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 116c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 117c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 125c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 126c | LDSD | 1 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 3 |
| 130c | LDSD | 1 | 1 | 0 | 1 | 0 | 0 | nd | 0 | 0 | 1 | 0 | 1 |
| 139c | LDSD | nr | nr | nr | 1 | 0 | 0 | nd | 0 | nr | 0 | 0 | 2 |
| 143c | DG | 1 | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 146c | DG | 0 | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 157c | DG | 0 | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 158c | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 161c | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 167c | DG | 0 | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 171c | DG | 0 | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 176c | DG | 0 | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 1 | 9 |
| 184c | DG | 1 | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 187c | DG | 0 | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 189c | DG | 1 | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 192c | DG | 1 | 1 | 0 | nr | nr | nd | nd | nd | nd | 1 | 0 | ? |
| 193c | DG | 1 | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 195c | DG | 1 | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 198c | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 207c | DG | nr | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 210c | DG | nr | 1 | nr | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 214 c | DG | 0 | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 216 c | DG | 1 | 1 | 0 | 1 | 0 | nd | nd | nd | nd | 1 | 0 | 1 |
| 217c | DG | 0 | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 220c | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 1 | 9 |
| 227c | DG | nr | 0 | 0 | 1 | 0 | nd | nd | nd | nd | 0 | 0 | 2 |
| 234c | DG | 0 | 0 | 0 | 1 | 0 | nd | nd | 1 | 1 | 1 | 0 | 16 |
| Finnish |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | LLy22g | M9 | 12f2 | Haplo |
| LGL514 | 135 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 2 |
| LGL514 | 136 | 0 | $0 / 1 ?$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 7or3 |
| LGL514 | 137 | 0 | 0 | nr | 1 | 0 | 0 | 0 | nr | 1 | 1 | nd | 12/16? |
| LGL515 | 138 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | nd | 1 |
| LGL517 | 139 | 0 | 0/1? | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | nd | 12 |
| LGL519 | 140 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | nd | 16 |
| LGL519 | 141 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | nr | 1 | nd | 16 |
| LGL519. | 142 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nd | 3 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| LGL519 | 143 | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | nr | 0 | nd | 21 |
| LGL519 | 144 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | nr | 1 | nd | 16 |
| LGL520 | 145 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 1 | 1 | 1 | nd | 16 |
| LGL523 | 146 | nr | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | nd | 16 |
| LGL524 | 147 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | nr | 1 | nd | 16 |
| LGL524. | 148 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nd | 3 |
| LGL524. | 149 | 0 | 0 | nr | 0 | 0 | 0 | 0 | nr | ? | 0 | nd | 7or3 |
| LGL525 | 150 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 2 |
| LGL525 | 151 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | nr | 1 | nd | 16 |
| LGL529 | 152 | 0 | 1 ? | 0 | 1 | 0 | 0 | 0 | nr | nr | 1 | nd | 26?/16? |
| LGL529 | 153 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 1 | 1 | 1 | nd | 16 |
| LGL530 | 154 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nd | 3 |
| Bulgarian gypsy |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Name | Pop | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12 f 2 | Haplo |
| Bug4 | Bug gy | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | nr |  | 2 |
| Bug5 | Bug gy | 0 | 0 | nr | 1 | 0 | 0 | 0 | nr | nr | 0 |  | 2 |
| Bug6 | Bug gy |  | 0 | nr | 1 | 0 | 0 | 0 | 0 | nr | 0 |  | 2? |
| Bug7 | Bug gy | 0 | 0 | nr | 1 | 0 | 0 | 0 | nr | 0 | nr |  | nr |
| Bug8 | Bug gy | 0 | 0 | nr | 1. | 0 | 0 | 0 | nr | 0 | 0 |  | 2 |
| Bug9 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug10 | Bug gy | 0 | 0 | nr | nr | 0 | 0 | 0 | 0 | 0 | nr |  | nr |
| Bug11 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug12 | Buggy | 0 | nr | 0 | 1 | 0 | 0 | 0 | 0 |  | 0 |  | 2 |
| Bug13 | Bug gy | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug14 | Bug gy | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | nr |  | 2 |
| Bug 15 | Bug gy | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug16 | Buggy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug17 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug18 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug19 | Bug gy |  | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug20 | Bug gy | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug21 | Buggy | 0 | 0 | nr | $\mathrm{nr} / \mathrm{nr}$ | 0 | 0 | 0 | 0 | 0 | 0 |  | $2 ?$ |
| Bug22 | Buggy | 0 | 0 | nr | 1. | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug23 | Buggy | 0 | 0 | 0 | $\mathrm{nr} / \mathrm{nr}$ | 0 | 0 | 0 | 0 | 0 | 0 |  | $2 ?$ |
| Bug24 | Bug gy | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug25 | Bug gy | 0 | 0 | nr | $\mathrm{nr} / \mathrm{nm}$ | 0 | 0 | 0 | 0 | 0 | nr |  | nr |
| Bug26 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug27 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug28 | Bug gy | . | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 |  | 2 |
| Bug29 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 |  | 2 |
| Bug30 | Bug gy | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |  | 1 |
| Bug31 | Buggy | , | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug32 | Buggy | 0 | 0 | 0 | 1. | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug 33 | Bug gy |  | 0 | 0 | $\mathrm{nr} / \mathrm{nu}$ | 0 | 0 | 0 | 0 | 0 | 0 |  | $2 ?$ |
| Bug34 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | nr | 0 | 0 |  | 2 |
| Bug35 | Bug gy | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug36 | Bug gy | 0 | 0 | 0 | - 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug37 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug38 | Buggy | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 |  | 1 |
| Bug39 | Buggy | 0 | 0 | 0 | - 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug40 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug41 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug42 | Bug gy | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 |  | 21 |
| Bug43 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug44 | Bug gy | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |  | 21 |
| Bug45 | Buggy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug46 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |  | 0 |  | 2 |
| Bug47 | Bug gy |  | 0 | nr | 1 | nr | nr | nr | 0 | nr | nr |  | nr |
| Bug48 | Bug gy |  | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug49 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug50 | Bug gy | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 |  | 1 |
| Bug51 | Buggy | 0 | nr | 0 | - 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug52 | Bug gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug53 | Romg |  | 0 | nr | 1 | 0 | 0 | 0 | 0 | nr | 0 |  | 2? |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12 f 2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Bug54 | It gyp | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |  | 2 |
| Bug55 | Ger gy | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 |  | 2 |
| Bug56 | Sp gyp | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |  | 1 |
| Bug57 | Bug gy |  | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bug58 | Bug gy |  | nr | 0 | nr | 1 | nr | 0 | 0 | nd | 0 |  | nr |
| Bug59 | Bug gy |  | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bug60 | Bug gy |  | 0 | 0 | 1 | 1 | nr | 0 | 0 | nd | 0 |  | 21? |
| Bug61 | Bug gy |  | nr | 0 | 1 | 1 | nr | 0 | 0 | nd | 0 |  | 21? |
| Bug62 | Bug gy |  | 0 | nr | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 ? |
| Bug63 | Bug gy |  | 0 | 0. | 1 | 1 | nr | 0 | 0 | nd | 0 |  | 21? |
| Bug64 | Bug gy ${ }^{\text {n }}$ |  | 1 | 0 | nr | 1 | nr | 0 | 0 | nd | 0 |  | nr |
| Bug65 | Bug gy |  | 0 | 0 | 1 | 1 | nr | 0 | 0 | nd | 0 |  | 21? |
| Bug66 | Bug gy n |  | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bug67 | Bug gy |  | nr | nr | nr | 0 | 0 | 0 | 0 | nd | nr |  | nr |
| Bulgarian non-gypsy |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Bul 1 | Bug nen |  | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 2 | Bug nct |  | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nd | 0 |  | 21 |
| Bul 3 | Bug ncn |  | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 4 | Bug nci |  | nr | 0 | 1 | nr | 0 | 0 | nr | nd | nr |  | nr |
| Bul 5 | Bug ncr |  | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 |  | 1 |
| Bul 6 | Bug ncı |  | 0 | 0 | 1 | nr | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 7 | Bug nen |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 0 |  | ?? |
| Bul 8 | Bugncr |  | 0 | 0 | 0 | nr | 0 | 0 | 0 | nd | 0 |  | ?? |
| Bul 9 | Bugncn |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 |  | 3 |
| Bul 10 | Bugnen |  | nr | 0 | 1 | nr | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 11 | Bug ncn |  | 0 | 0 | 1 | nr | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 12 | Bug ncin |  | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nd | 0 |  | 21 |
| Bul 13 | Bug non |  | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 14 | Bug non |  | 0 | 0 | 1 | 1. | nr | 0 | 0 | nd | 0 |  | 21 |
| Bul 15 | Bug nen |  | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 16 | Bug nen |  | nr | nr | nr | nr | 0 | 0 | 0 | nd | 0 |  | nr |
| Bul 17 | Bug nct |  | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 18 | Bug ncin |  | 0 | 0 | nr | nr | 0 | 0 | 0 | nd | 0 |  | nr |
| Bul 19 | Bugncr |  | nr | nr | nr | nr | 0 | 0 | nr | nd | nr |  | nr |
| Bul 20 | Bugncn |  | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 21 | Bugncn |  | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 22 | Bug nct |  | nr | nr | nr | nr | 0 | 0 | , | nd | 0 |  | nr |
| Bul 23 | Bug ncr |  | 0 | nr | nr | 0 | 0 | 0 | 0 | nd | 0 |  | nr |
| Bul 24 | Bugnen |  | 1 | 0 | 1 | 0 | 0 | 0 |  | nd | 1 |  | 1 |
| Bul 25 | Bug ncr |  | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 |  | 3 |
| Bul 26 | Bugncn |  | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 |  | 1 |
| Bul 27 | Bugncn |  | nr | 0 | 1 | 1 | nr | 0 | 0 | nd | 0 |  | 21 |
| Bul 28 | Bug ncin |  | 0 | 0 | 1 | 0 | 0 | 0 | , | nd | 0 |  | 2 |
| Bul 29 | Bugncn |  | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 30 | Bug nct |  | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 |  | 1 |
| Bul 31 | Bug nen |  | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 |  | 2 |
| Bul 32 | Bug ncr |  | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 |  | 3 |
| Turkish |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Pop | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | LLy 22 g | M9 | 12f2 | Hap |
| Tuk 1 | Turkis | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 1 | 1 |
| Tuk 2 | Turkis | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| Tuk 3 | Turkis | nr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 0 | 0 | ?? |
| Tuk 4 | Turkis | 1 | nr | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 16 |
| Tuk 5 | Turkis | 1 | nr | 0 | 1 | 1 | nr | 0 | 0 | nd | 0 | 0 | $21 ?$ |
| Tuk 6 | Turkis | 0 | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| Tuk 7 | Turkis | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 1. | 1 |
| Tuk 8 | Turkis | 0 | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 1 | 9 |
| Tuk 9 | Turkis | 1 | 1 | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Tuk 10 | Turkis | 1 | 1 | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Tuk 11 | Turkis | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Tuk 12 | Turkis | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Tuk 13 | Turkis | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1. |
| Tuk 14 | Turkis | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Tuk 15 | Turkis | 0 | 1 | nr | nr | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Tuk 16 | Turkis | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| Tuk 17 | Turkis | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| Tuk 18 | Turkis | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | ? |
| Tuk 19 | Turkis | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 1 | 9 |
| Tuk 20 | Turkis | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | nr |
| Tuk 21 | Turkis | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 |
| Tuk 22 | Turkis | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | ? |
| Tuk 23 | Turkis | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | nr |
| Tuk 24 | Turkis | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 1 | 9 |
| Tuk 25 | Turkis | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | nr | nr |
| Tuk 26 | Turkis | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| Tuk 27 | Turkis | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| Tuk 28 | Turkis | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 |
| Tuk 29 | Turkis | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Tuk 30 | Turkis | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 1 | 9 |
| Cypriot |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Pop | 92R7 | DYS257 | 2627 | 1532 | YAP | 8299 | sY81 | Tat | LLy22g | M9 | 12 f 2 | Hap |
| Cy1 | Cypriot | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy2 | Cypriot | 0 | 0 | 0 | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Cy3 | Cypriot | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Cy4 | Cypriot | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Cy5 | Cypriot | nr | nr | 0 | 0 | 0 | 0 | 0 | nr | nr | 0 | nr | nr |
| Cy6 | Cypriot | 0 | 0 | 0 | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | 2? |
| Cy7 | Cypriot | 1 | 1 | 0 | 1/0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Cy8 | Cypriot | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 0 | 1 | 9 |
| Cy9 | Cypriot | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Cy10 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy11 | Cypriot | 0 | nr | 0 | 1 | 1 | 1 | 0 | 0 | nr | 0 | 0 | 21 |
| Cy12 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy13 | Cypriot | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | 0 | nr | 0 | 21 |
| Cy14 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Cy15 | Cypriot | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | nr | 0 | 0 | 21 |
| Cy16 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy17 | Cypriot | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Cy 18 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| Cy19 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 26 |
| Cy20 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy21 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy22 | Cypriot | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Cy 23 | Cypriot | 0 | 0 | nr | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Cy24 | Cypriot | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Cy 25 | Cypriot | 0 | nr | nr | 1 | 1 | 1 | 0 | 0 | nr | 0 | 0 | 21 |
| Cy26 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Cy27 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| Cy 28 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| Cy29 | Cypriot | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Cy30 | Cypriot | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Cy31 | Cypriot | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy32 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy33 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Cy34 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy35 | Cypriot | nr | nr | 1 ? | nr | 1 r | nr | nr | nr | nr | nr | nr | nr |
| Cy36 | Cypriot | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Cy37 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 26 |
| Cy38 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 12 |
| Cy 39 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy40 | Cypriot | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Cy41 | Cypriot | nr | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nr | 0 | 0 | 21 |
| Cy42 | Cypriot | 0 | 0 | 0 | 0/nr | 0 | 0 | 0 | 0 | nr | 0 | 1 | 9 |
| Cy43 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| Cy44 | Cypriot | 0 | 0 | $1 ?$ | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy45 | Cypriot | 0 | 0 | 0 | nr | 0 | 0 | 0 | 0 | nr | 0 | nr | 2?/7? |
| Cy46 | Cypriot | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy47 | Cypriot | nr | 1 | 0 | $0 / \mathrm{nr}$ | 0 | 0 | 0 | nr | nr | 0 | nr | nr |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Cy48 | Cypriot | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Cy49 | Cypriot | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Cy50 | Cypriot | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | nr | 0 | 0 | 21 |
| Mainland Greek |  |  |  |  |  |  |  |  |  |  |  |  |  |
| G1 | Greek | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| G2 | Greek | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| G3 | Greek | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| G4 | Greek | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| G5 | Greek | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| G6 | Greek | 0 | nr | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| G7 | Greek | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ? |
| G8 | Greek | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| G9 | Greek | nr | 1 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr |
| G10 | Greek | 1 | 1 | nr | 0/nr | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| G11 | Greek | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| G12 | Greek | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| G13 | Greek | 0 | nr | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| G14 | Greek | 0 | nr | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| G15 | Greek | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| G16 | Greek | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| G17 | Greek | 1 | 1 | nr | 0/nr | 0 | 0 | 0 | nr | nr | 1 | 0 | nr |
| G18 | Greek | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| G19 | Greek | 0 | 0 | 0 | 0/nr | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 ? |
| G20 | Greek | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| G21 | Greek | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| G22 | Greek | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| G23 | Greek | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | nr | 21 |
| G24 | Greek | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | nr | 1 | nr | 26 |
| G25 | Greek | 0 | nr | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | nr | 21 |
| G26 | Greek | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 2 |
| G27 | Greek | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 2? |
| G28 | Greek | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |
| G29 | Greek | 1 | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | nr | 0 | 3? |
| G30 | Greek | 0 | 0 | 0 | 0/nr | 0 | 0 | 0 | 0 | 0 | 0 | nr | 7or3 |
| G31 | Greek | 1 | nr | nr | 1 | 0 | 0 | 0 | 0 | nr | 1 | nr | 1 |
| G32 | Greek | nr | 1 | nr | 0 | 0 | 0 | 0 | nr | nr | 1 | nr | nr |
| G33 | Greek | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | nr | 0 | nr | 21 |
| G34 | Greek | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | nr | 0 | 1 | 9 |
| G35 | Greek | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| G36 | Greek | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | nr | 21 |
| G37 | Greek | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| G38 | Greek | 0 | 0 | nr | 0 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 7? |
| G39 | Greek | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 2 |
| G40 | Greek | 1 | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | nr | nr | 1 ? |
| G41 | Greek | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| G42 | Greek | 0 | 0 | 0 | 0/nr | 0 | 0 | 0 | 0 | 0 | nr | 0 | 7 r 3 |
| G43 | Greek | 0 | 0 | nr | 0/nr | 0 | 0 | 0 | 0 | 0 | 0 | nr | nr |
| G44 | Greek | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21? |
| Slovenian |  |  |  |  |  |  |  |  |  |  |  |  |  |
| PLATE |  | 92 R 7 | DYS257 | 2627 | 1532 | YAP | 8299 | sY81 | Tat | LLy22g | M9 | 12 f 2 | Haplo |
| Slov1 |  | 1 | 1 | 0 | nr | 0 | 0 | 0 | 0 | 0 | 1 | nr | 3 |
| Slov2 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Slov 3 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Slov4 |  | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | nr | 1 | nr | 3 |
| Slov5 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Slov6 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| Slov7 |  | 1 | nr | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ? |
| Slov8 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov9 |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ? |
| Slov10 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov11 |  | 0 | 0 | 0 | 1 | 0 | 0. | 0 | 0 | 0 | 0 | nr | 2 |
| Slov12 |  | 1 | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov13 |  | nr | nr | nr | nr | 0 | 0 | 0 | nr | nr | nr | nr | nr |
| Slov14 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Slov15 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov16 |  | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Slov17 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 26 |
| Slov18 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| Slov19 |  | nr | 0 | nr | 1 | 1 | 1 | 0 | 0 | nr | 0 | 0 | 21? |
| Slov20 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Slov21 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Slov22 |  | 1 | nr | 0 | nr | nr | 0 | 0 | nr | 0 | 1 | nr | nr |
| Slov23 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 2 |
| Slov24 |  | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21? |
| Slov25 |  | nr | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |  |
| Slov26 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov27 |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |  |
| Slov28 |  | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | nr | 21? |
| Slov29 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| Slov30 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Slov31 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Slov32 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Slov33 |  | nr | nr | ? 1 | nr | 0 | 0 | 0 | nr | nr | nr | nr | nr |
| Slov34 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | nr | 0 | 0 | 0 | 2 |
| Slov35 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 2 |
| Slov36 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov37 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Slov38 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Slov39 |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nr | nr | 0 | ? |
| Slov40 |  | nr | nr | 0 | 1 | 0 | 0 | 0 | nr | nr | 0 | 0 | 2 |
| Slov41 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov42 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |  |
| Slov43 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov44 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Slov45 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 3 |
| Slov46 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov47 |  | 1 | nr | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov48 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov49 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov50 |  | nr | nr | 0 | nr | 0 | 0 | 0 | nr | nr | nr | 0 | nr |
| Slov51 |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Slov52 |  | nr | nr | 0 | nr | 0 | 0 | 0 | nr | nr | nr | 0 | nr |
| Slov53 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| Slov54 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 3 |
| Slov55 |  | 0 | 0 | 0 | 1 | 0 | 0. | 0 | 0 | nr | 0 | 1 | 9 |
| Slov56 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov57 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov58 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Slov59 |  | nr | nr | 0 | nr | 0 | 0 | 0 | nr | nr | 0 | nr | nr |
| Slov60 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |  |
| Slov61 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Slov62 |  | nr | nr | 0 | nr | 0 | 0 | 0 | nr | nr | 1 | 0 | nr |
| Slov63 |  | nr | nr | 0 | nr | 0 | 0 | 0 | nr | nr | nr | nr | nr |
| Slov64 |  | nr | nr | nr | nr | 0 | 0 | 0 | nr | nr | nr | nr | nr |
| Slov65 |  | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | nr |
| Slov66 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| Slov67 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 3 |
| Slov68 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Slov69 |  | nr | nr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | ? |
| Slov70 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 3 |
| Slov71 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |
| Slov72 |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ? |
| Slov73 |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 21? |
| Slov74 |  | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 |
| Slov75 |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |  |
| Slov76 |  | 1 | 1 | 0 | nr | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| Slov77 |  | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | nr | 2 |
| Slov78 |  | nr | 0 | 0 | nr | 0 | 0 | 0 | 0 | 0 | 0 | nr | 2? |
| Slov79 |  | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Slov80 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |
| Slov81 |  | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 3 |
| Slov82 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 3 |
| Slov83 |  | 0 | 0 | 0 | 1 | 1 ? | 1 | 0 | 0 | 0 | 0 | nr | 21 |
| Slov84 |  | nr | nr | 0 | nr | 0 | 0 | 0 | nr | nr | 1 | nr | nr |
| Slov85 |  | nr | nr | 0 | nr | 0 | 0 | 0 | nr | 0 | nr | 0 | nr |
| Slov86 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 3 |
| Slov87 |  | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ? |
| Slov88 |  | nr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 0 | 0 | ? |
| Slov89 |  | nr | nr | 0 | nr | 0 | 0 | 0 | nr | 0 | 0 | 0 | nr |
| Slov90 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Slov91 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Slov92 |  | nr | nr | ? 1 | nr | 0 | 0 | 0 | nr | nr | 1 | nr | nr |
| Slov93 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 3 |
| Slov94 |  | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 3 |
| Polish |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Pop | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | LLy22g | M9 | 12 f 2 | Hap |
| Pol 1 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 2 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 3 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 4 | Polish | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Pol 5 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 6 | Polish | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | nr | nr |
| Pol 7 | Polish. | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 8 | Polish | nr | nr | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 9 | Polish | nr | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 16 |
| Pol 10 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 11 | Polish | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | nr | nr |
| Pol 12 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | nr | 16-faint |
| Pol 13 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 14 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 15 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 16 | Polish | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 1 | 9 |
| Pol 17 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | nr | 3 |
| Pol 18 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 19 | Polish | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| Pol 20 | Polish | nr | 1 | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | nr | nr |
| Ladin |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| Vf1 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 1 |
| Vf2 | LDSD | 0 | 0 | nr | 1 | 0 | nd | nd | 0 | 0 | 0 | 1 | 9 |
| Vf3 | LDSD | 0 | 0 | 0 | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 |  |
| Vf4 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 1 |
| Vf5 | LDSD | 1 | 1 | nr | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 1 |
| Vf6 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 1 |
| Vf7 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 1 |
| Vf8 | LDSD | 0 | nr | 0 | 1 | 0 | nd | nd | 0 | 0 | 0 | 1 | 9 |
| Vf9 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | $1 ?$ |
| Vf10 | LDSD | 0 | 0 | 0 | 1 | 0 | nd | nd | 0 | 0 | 0 | 1 | 1 |
| Vb 1 | LDSD | 1 | nr | 0 | 1 | 0 | nd | nd | 0 | 0 | nr | nr | nr |
| Vb 2 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 1 |
| Vb3 | LDSD | 1 ? | $1 ?$ | 0 | 1 | 1 | nr | 0 | 0 | 0 | 0 | 0 | 26? |
| Vb4 | LDSD | 1 | nr | nr | 1 | 0 | nd | nd | 0 | nr | 1 | nr | 1 |
| Vb5 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | nr | 1 | 0 | 1 |
| Vb6 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | nr | 1 | 0 | 1 |
| Vb7 | LDSD | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | nr | 0 |  |
| $\sqrt{68}$ | LDSD | 0 | 0 | 0 | 1 | 0 | nd | nd | 0 | 0 | 0 | 0 | 2/26? |
| Vb9 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | nr | 1 | nr | 1 |
| Vg1 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | - 1 |
| Vg2 | LDSD | nr | nr | nr | nr | nr | nd | nd | nr | nr | nr | nr | female |
| Vg3 | LDSD | 0 | nr | 0 | 1 | 1 | 1 | 0 | 0 | 0 | nr | 0 | 2/26? |
| Vg4 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 1 |
| Vg5 | LDSD | 1 | 1 | 0 | 1 |  | nd | nd | 0 | 0 | 1 | 0 | 1 |
| Vg6 | LDSD | 1 | 1 |  | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 1 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Vg7 | LDSD | 0 | 0 | 0 | 1 | 0 | nd | nd | 0 | 0 | 0 | 0 | 2 |
| Sardinian |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly 22 g | M9 | 12 f 2 | Haplo |
| SAR1 | LDSD | nr | nr | nr | nr | 0 | nd | nd | 0 | nr | $0 / \mathrm{nr}$ | 0 | nr |
| SAR2 | LDSD | nr | nr | nr | nr | 0 | nd | nd | nr | nr | $0 / \mathrm{nr}$ | 0 | nr |
| SAR3 | LDSD | nr | nr | nr | nr | 0 | nd | nd | 0 | nr | $\mathrm{nr} / \mathrm{nr}$ | 0 | nr |
| SAR4 | LDSD | 1 | 1 | nr | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 1 |
| SAR5 | LDSD | nr | 0 | nr | 0 | 0 | nd | nd | 0 | nr | $0 / \mathrm{nr}$ | nr | nr |
| SAR6 | LDSD | nr | 1 | nr | 1 | 0 | nd | nd | 0 | nr | $\mathrm{nr} / \mathrm{nr}$ | nr | nr |
| SAR7 | LDSD | nr | 1 | nr | 1 | 1 | 1 | 1 | 0 | nr | 0 | 0 | 8 |
| SAR9 | LDSD | 1 | 0 | nr | 1 | 0 | nd | nd | 0 | nr | 1 | 0 | nr |
| SAR10 | LDSD | 0 | 0 | nr | 1 | 0 | nd | nd | 0 | nr | 0 | 0 | 2 |
| SAR11 | LDSD | nr | 0 | nr | 1 | 0 | nd | nd | 0 | nr | 0 | 0 | 2 |
| SAR12 | LDSD | nr | nr | nr | nr | 0 | nd | nd | nr | nr | $1 / \mathrm{nr}$ | 0 | nr |
| SAR13 | LDSD | 1 | 1 | 0 | - 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 1 |
| SAR14 | LDSD | 0 | 0 | 0 | - 1 | 1 | 1 | 0 | 0 | nr | 0 | 0 | 21 |
| SAR15 | LDSD | 1 | 1 | 0 | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 1 |
| SAR17 | LDSD | 0 | 0 | 0 | 1 | 1. | 1. | 0 | 0 | 0 | 0 | 0 | 21 |
| SAR18 | LDSD | 0 | 0 | 0 | 1 | 0 | nd | nd | 0 | nr | 0 | 0 | 2 |
| SAR19 | LDSD | 0 | 0 | 0 | 1 | 0 | nd | nd | 0 | 0 | 0 | nr | 2 |
| Bavarian |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Pop | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | LLy22g | M9 | 12 f 2 | Hap |
| Bav1 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav2 | Bavariar | 1 | nr | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |
| Bav3 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav4 | Bavariar | nr | nr | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | nr |
| Bav5 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | nr | 0 | 1 | 0 | 1 |
| Bav6 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| Bav7 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav8 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav9 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav10 | Bavariar | nr | nr | nr | 1 | nr | nr | nr | 0 | 0 | nr | nr | nr |
| Bav11 | Bavariar | 1 | nr | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |
| Bav12 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav13 | Bavariar | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nr | 0 | 0 | 21 |
| Bav14 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Bav15 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav16 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav17 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |  | 1 | nr | 1 |
| Bav18 | Bavariar | nr | nr | nr | 1 | nr | nr | nr | nr | nr | nr | nr | nr |
| Bav19 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |
| Bav20 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav21 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav22 | Bavariar | nr | nr | nr | 1 | nr | nr | nr | nr | nr | nr | nr | nr |
| Bav23 | Bavariar | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 3 |
| Bav24 | Bavariar | 1 | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav25 | Bavariar | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Bav26 | Bavariar | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Bav27 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav28 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Bav29 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav30 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav31 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav32 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |
| Bav33 | Bavariar | nr | nr | nr | 1 | 1 | 1 | 0 | 0 | nr | 0 | nr | nr |
| Bav34 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| Bav35 | Bavariar | 1 | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav36 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| Bav37 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav38 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Bav39 | Bavariar | 0 | 0 | nr | 1. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav40 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Bav41 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| Bav42 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 2 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Bav43 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav44 | Bavariar | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 3 |
| Bav45 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Bav46 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 |
| Bav47 | Bavariar | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav48 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| Bav49 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav50 | Bavariar | 1 | nr | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav51 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav52 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav53 | Bavariar | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 |
| Bav54 | Bavariar | nr | nr | nr | 0 | nr | nr | nr | nr | nr | 1 | nr | nr |
| Bav55 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav56 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |
| Bav57 | Bavariar | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 2 |
| Bav58 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav59 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav60 | Bavariar | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav61 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav62 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav63 | Bavariar | nr | nr | nr | 1 | nr | nr | nr | nr | nr | nr | nr | nr |
| Bav64 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav65 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav66 | Bavariar | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 3 |
| Bav67 | Bavariar | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav68 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav69 | Bavariar | 1 | 1 | $1 ?$ | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 22? |
| Bav70 | Bavariar | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Bav71 | Bavariar | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Bav72 | Bavariar | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Bav73 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav74 | Bavariar | 1 | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Bav75 | Bavariar | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Bav76 | Bavariar | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Bav77 | Bavariar | 1 | nr | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav78 | Bavariar | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 22? |
| Bav79 | Bavariar | nr | nr | 0 | 0 | nr | nr | nr | nr | nr | 1 | nr | nr |
| Bav80 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav81 | Bavariar | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | nr | 0 | 21 |
| Bav82 | Bavariar | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | nr | 0 | nr |
| Bav83 | Bavariar | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7? |
| Bav84 | Bavariar | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | nr | 21 |
| Bav85 | Bavariar | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Bav86 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav87 | Bavariar | 1 | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav88 | Bavariar | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Bav89 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Bav90 | Bavariar | 1 | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Bav91 | Bavariar | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Bav92 | Bavariar | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | nr | 0 | 1 ? |
| Bav93 | Bavariar | 1 | nr | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | ? |
| Bav94 | Bavariar | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 0 | 3 ? |
| Dutch |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Tube | 92R7 | DYS257 | 2627 | 1532 r | Yap | 8299 | sY81 | Tat | Lly22g | M9 | $12 \mathrm{f2}$ | Hap |
| Du2 | FLDO2 | nr | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | nr | 0 | nr |
| Du3 | FLDO3 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Du6 | FLDO6 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Du7 | FLDO7 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du9 | FLDO9 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Du11 | FLDO11 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Du13 | FLDO13 | 1 | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | 1 | rr | 3 |
| Du14 | FLDO14 | 0 | nr | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 26? |
| Du16 | FLDO16 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du17 | FLDO17 | 0 | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Du19 | FLDO19 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du20 | FLDO20 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du22 | FLDO22 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du23 | FLDO23 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du25 | FLDO25 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du26 | FLDO26 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du27 | FLDO27 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du28 | FLDO28 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du29 | FLDO29 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du30 | FLDO30 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du31 | FLDO31 | 1 | nr | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |
| Du32 | FLDO32 | 1 | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Du33 | FLDO33 | nr | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du34 | FLDO34 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du35 | FLDO35 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du36 | FLDO36 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du37 | FLDO37 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 2 |
| Du38 | FLDO38 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | nr | 0 | 1 ? |
| Du42 | FLDO42 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 2 |
| Du51 | FLDO51 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |
| Du53 | FLDO53 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du54 | FLDO54 | 1 | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du55 | FLDO55 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Du56 | FLDO56 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Du59 | FLDO59 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du60 | FLDO60 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 9 |
| Du64 | FLDO64 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Du65 | FLDO65 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du66 | FLDO66 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du68 | FLDO68 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du69 | FLDO69 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Du70 | FLDO70 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du71 | FLDO71 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 22 |
| Du72 | FLDO72 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du73 | FLD073 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du74 | FLDO74 | 0 ? | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Du75 | FLDO75 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du76 | FLDO76 | nr | 1 | nr | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | nr |
| Du77 | FLDOTT | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du78 | FLDO78 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du79 | FLDO79 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Du80 | FLDO80 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 2 |
| Du81 | FLDO81 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du82 | FLDO82 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| Du83 | FLDO83 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du84 | FLDO84 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | nr | 0 | $1 ?$ |
| Du85 | FLDO85 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | , | 9 |
| Du86 | FLD086 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du87 | FLD087 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du88 | FLDO88 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | nr | 1 | 9 |
| Du89 | FLDO89 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du90 | FLDO90 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Du91 | FLDO91 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du92 | FLDO92 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du93 | FLDO93 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du94 | FLDO94 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du95 | FLDO95 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du96 | FLDO96 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du97 | FLDO97 | 0 | 0 | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Du98 | FLDO98 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du99 | FLDO99 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du100 | FLDO10 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du101 | FLDO10 | 1 | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du102 | FLDO10 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du103 | FLDO10 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Du104 | FLDO10 | nr | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | nr | 0 | 1 ? |
| Du106 | FLDO10 | nr | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du107 | FLDO10 | 1 | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du108 | FLDO10 | 1 | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | nr | 0 | 1 ? |
| Du109 | FLDO10 | 1 | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Dul10 | FLDO11 | 0 | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | nr | 0 | 2?/26? |
| Dul11 | FLDO11 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | nr | 0 | 1 ? |
| Du112 | FLDO11] | 1 | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du113 | FLDO11 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du114 | FLDO11 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du115 | FLDO11 | 0 | 0 | 0 | nr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Du116 | FLDO11 | nr | 0 | nr | 1 | 0 | 0 | 0 | 0 | 0 | nr | 0 | nr |
| Du117 | FLDO11 | 1 | 1 | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| Du118 | FLDO11 | 0 | 0 | nr | nr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| French |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Pop | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Hap |
| 201 | French | $n \mathrm{r}$ | 1 | 1 | 0 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1? |
| 1201 | French | 0 | nr | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| 1701 | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| 2101 | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| 2301 | French | 0 | nr | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | nr | 2 |
| 2804 | French | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| 3501 | French | 1 | 1 | 1 | 1 | nr | 0 | 0 | 0 | nr | 1 | 0 | 22 |
| 3701 | French | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | nr | 2? |
| 4501 | French | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2? |
| 6601 | French | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| Jebu | French | 1 | 1 | $1 / \mathrm{nr}$ | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 22 |
| Prjb | French | 1 | nr | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| 1F | French | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 ? |
| 2 F | French | nr | nr | nr | nr | 0 | 0 | 0 | 0 | nr | 1 | 0 | nr |
| 3F | French | nr | nr | nr | 0 | nr | 0 | 0 | 0 | nr | nr | nr | nr |
| 4 F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | nr | 1 |
| 5F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| 6 F | French | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 ? |
| 7 F | French | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 3 |
| 8 F | French | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | nr | 2 |
| 9 F | French | nr | 0 | nr | 0 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 3 ? |
| 10F | French | nr | nr | 0 | nr | nr | 0 | 0 | 0 | nr | nr | nr | nr |
| 11 F | French | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | nr | 2 |
| 12F | French | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 13F | French | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| 14F | French | nr | nr | 0 | 1 | nr | 0 | 0 | 0 | 0 | nr | 1 | nr |
| 15F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 16F | French | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 21 |
| 17F | French | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1?? | 21 |
| 18 F | French | nr | nr | nr | 0 | nr | 0 | 0 | 0 | nr | 1 | nr | nr |
| 19F | French | 1 | 0 | 0 | 0 | nr | 0 | 0 | 0 | nr | nr | nr | 3 ? |
| 20 F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | , | 0 | 1 |
| 21F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |  | 0 | 1 |
| 22 F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 23F | French | 1 | 1 | 0 | 1 | nr | 0 | 0 | 0 | nr | nr | nr | 1 ? |
| 24 F | French | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| 25F | French | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 26 F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 27F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | $\underline{\mathrm{nr}}$ | 1 | 0 | 1 |
| 28F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 29F | French | 1 | 0 | 0 | 1 | nr | 0 | 0 | 0 | 0 | nr | 1?? | 1 ? |
| 30F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| 31 F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 32F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| 33F | French | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 |
| 34F | French | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 3 |
| 35F | French | nr | nr | nr | 0 | nr | 0 | 0 | 0 | nr | nr | nr | nr |
| 36F | French | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | nr | 0 | 8 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 37F | French | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | nr | 1 | 2? |
| Scottish (mainland) |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Pop | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Hap |
| 2101 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | nr | 1 |
| 2120 | Scot | 0 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 0 | nr | 2 |
| 2121 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | nr | 1 |
| 2130 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | nr | 1 |
| 2180 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | nr | 1 |
| 2446 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | nr | 1 |
| 2309 | Scot | 0 | nd | 0 | 1 | 0 | nd | nd |  | nd |  | nr | 2 |
| 2316 | Scot | 0 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 0 | 0 | 2 |
| 2318 | Scot | 1 | nd | 0 | 0 | 0 | nd | nd | 0 | nd | 1 | 0 | 3 |
| 2326 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 2339 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 2340 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | 0 | 0 | 0 | 2 |
| 2405 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 2470 | Scot | 1 | nd | 0 | 1. | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 2581 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 2741 | Scot | 1 | nd | 0 | 1. | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3094 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3529 | Scot | 0 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 0 | 0 | 2 |
| 3194 | Scot | nr | nd | nr | 1 | 0 | nd | nd | 0 | nr | 1 | 0 |  |
| 3209 | Scot | 1 | nd | 0 | 1 | 0. | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3210 | Scot | nr | nd | nr | nr | nr | nd | nd | nr | nd | nr | 0 | ? |
| 3312 | Scot | nr | nd | nr | 1 | 0 | nd | nd | 0 | nd | nr | 0 | ? |
| 3222 | Scot | 1 | nd | 0 | 1. | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3234 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3236 | Scot | 1 | nd | 0 | 0 | 0 | nd | nd | 0 | nd | 1 | 0 | 3 |
| 3244 | Scot | 0 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 0 | 0 | 2 |
| 3246 | Scot | 0 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 0 | 0 | 2 |
| 3294 | Scot | 1 | nd | 0 | 1. | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3298 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3313 | Scot | nr | nd | nr | 0 | 0 | nd | nd | 0 | nd | 1 | 0 | ? |
| 3316 | Scot | nr | nd | nr | nr | nr | nd | nd | nr | nd | nr | 0 | ? |
| 3364 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3421 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3422 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3441 | Scot | 1 | nd | 0 | 1. | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3447 | Scot | 1 | nd | 0 | 0 | 0 | nd | nd | 0 | nd | 1 | 0 | 3 |
| 3451 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3459 | Scot | 0 | nd | 0 | 1 | 0 | nd | nd | 0 | 0 | 1 | 0 | 26 |
| 3460 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3463 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | , | 0 | 1 |
| 3489 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | , | 0 | 1 |
| 3515 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3517 | Scot | 1. | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3531 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd |  | nd | 1 | 0 | 1 |
| 3540 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | - | nd | 1 | 0 | 1 |
| 3543 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd |  | nd | 1 | 0 | 1 |
| 3585 | Scot | 1 | nd | 0 | 1. | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| 3631 | Scot | nr | nd | nr | 1 | nr | nd | nd | nr | nd | nr | 0 | ? |
| 3639 | Scot | 1. | nd | 0 | 1 |  | nd | nd |  | nd | 1 | 0 | 1 |
| 3703 | Scot | 1 | nd | 0 | 1 | 0 | nd | nd | 0 | nd | 1 | 0 | 1 |
| m 2 | Eng | 1 | nd | 0 | 1 | 0 | nd | nd |  | nd | 1 | 0 | 1 |
| m10 | Scot | 1 | nd | 0 | 1 |  | nd | nd |  | nd | 1 | 0 | 1 |
| m11 | Eng | 1 | nd | 0 | 1 |  | nd | nd |  | nd | 1 | 0 | 1 |
| m13 | Eng | 1 | nd | 0 | 1 |  | nd | nd |  | nd | 1 | 0 | 1 |
| East Anglian |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Coord |  | 92R7 | DYS257 | RY 262 RY 153 |  | YAP | 8299 | sY81 | Tat | LLy22g | M9 | 12 f 2 | Hap |
| 1A1 |  | 1 | nd | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| 1A2 |  | nr | nd | nr | nr | nr | 0 | 0 | nr | nd | nr | 0 | nr |
| 1A3 |  | 1 | nd | nr | nr | nr | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| 1A4 |  | 1 | nd | nr | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1A5 |  | 1 | nd | 0 | 1 | nr | 0 | 0 | 0 | nd | 1 | 0 | 1 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1A6 |  | 0 | nd | nr | 1 | 1 | 1 | 0 | 0 | nd | 0 | 0 | 21 |
| 1A7 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1A8 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1A9 |  | 0 | nd | 0 | 0/0 | 0 | 0 | 0 | 0 | nd | 0 | 0 | $3^{*}$ |
| 1A10 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1A11 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1A12 |  | nr | nd | 0 | 1 | 1 | 1 | 0 | 0 | nd | 0 | 0 | 21 |
| 1B1 |  | 1 | nd | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| 1 B 2 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | nr | nd | 0 | 0 | 2 |
| 1B3 |  | 1 | nd | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| 1B4 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1B5 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1B6 |  | 1 | nd | 0 | 1 | nr | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1B7 |  | 1 | nd | nr | 0/nr | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| 1B8 |  | nr | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| $1 \mathrm{B9}$ |  | 1 | nd | 1 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 22 |
| 1B10 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1B11 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | nr | nd | 1 | 0 | 1 |
| 1B12 |  | 0 | nd | nr | 1 | 0 | 0 | 0 | nr | nd | 0 | 0 | 2 |
| 1C1 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1C2 |  | 1 | nd | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| 1 C 3 |  | nr | nd | nr | 1 | 0 | 0 | 0 | 0 | nd | nr | 0 | nr |
| 1C4 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 1C5 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 |
| 1C6 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 1C7 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 1C8 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1C9 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1C10 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1 C 11 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1 C 12 |  | 1 | nd | nr | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1D1 |  | 0 | nd | 0 | nr | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 1 D 2 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 1D3 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 1D4 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1D5 |  | nr | nd | nr | 1 | nr | 0 | 0 | nr | nd | nr | 0 | nr |
| 1D6 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1D7 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 1D8 |  | nr | nd | nr | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 26 or 2 |
| 1D9 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 1 D 10 |  | 0 | nd | nr | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| $1 \mathrm{D11}$ |  | 0 | nd | $\underline{\mathrm{nr}}$ | 1 | 1 |  |  | 0 | nd | 0 | 0 | 21 |
| 1D12 |  | 1 | nd | nr | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| 1 E 1 |  | nr | nd | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1/3? |
| 1 E 2 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1 E 3 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1 E4 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1 ES |  | nr | nd | nr | nr | nr | 0 | 0 | nr | nd | nr | 0 | nr |
| 1 E6 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1 E 7 |  | nr | nd | 0 | nr | nr | 0 | 0 | nr | nd | nr | 0 | nr |
| 1 E8 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 1 E 9 |  | nr | nd | nr | nr | nr | nr | nr | nr | nd | nr | 0 | ignore |
| 1 E10 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1 E11 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1 E12 |  | 1 | nd | nr | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1F1 |  | nr | nd | nr | nr | nr | 0 | 0 | nr | nd | nr | 0 | nr |
| 1 F 2 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1 F 3 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 154 |  | nr | nd | 0 | 0/0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3* |
| 1F5 |  | nr | nd | nr | nr | nr | 0 | 0 | nr | nd | nr | 0 | nr |
| 1 F 6 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1 F 7 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | ad | 1 | 0 | 1 |
| 1F8 |  | nr | nd | 0 | 0/nr | nr | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| 1F9 |  | nr | nd | nr | nr | nr | nr | nr | nr | nd | nr | 0 | ignore |
| 1F10 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12 f 2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1F11 |  | nr | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1F12 |  | 0 | nd | nr | 0 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 7 ? |
| 1G1 |  | 1 | nd | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1/3? |
| 1G2 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1G3 |  | 1 | nd | 0 | 1 | $\overline{0}$ | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1G4 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1G5 |  | nr | nd | nr | nr | nr | nr | nr | nr | nd | nr | 0 | nr |
| 1G6 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1G7 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1G8 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1G9 |  | nr | nd | nr | nr | nr | nr | nr | nr | nd | nr | 0 | ignore |
| 1G10 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 1G11 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 1G12 |  | 1 | nd | nr | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1H1 |  | nr | nd | 0 | 1 | nr | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1H2 |  | nr | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1H3 |  | nr | nd | 0 | 1 | nr | 0 | 0 | 0 | nd | nr | 0 | nr |
| 1H4 |  | nr | nd | 0 | 1 | nr | 0 | 0 | 0 | nd | 1 | 0 | ignore |
| 1H5 |  | nr | nd | nr | 1 | nr | 0 | 0 | 0 | nd | 1 | 0 | nr |
| 1H6 |  | nr | nd | nr | 1 | nr | 0 | 0 | 0 | nd | 1 | 0 | nr |
| 1H7 |  | nr | nd | nr | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1H8 |  | nr | nd | nr | 1 | nr | 0 | 0 | 0 | nd | 1 | 0 | nr |
| 1H9 |  | nr | nd | nr | nr | nr | nr | nr | nr | nd | nr | 0 | ignore |
| 1H10 |  | nr | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2? |
| 1H11 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 1H12 |  | nr | nd | nr | nr | nr | 0 | 0 | 0 | nd | 0 | 0 | nr |
| 2A1 |  | 0 | nd | 0 | 0/nr | 0 | 0 | 0 | 0 | nd | nr | 0 | 3 |
| 2A2 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2A3 |  | 1 | nd | 0 | 0 | nr | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| 2A4 |  | 1 | nd | 0 | 0 | nr | 0 | 0 | nr | nd | 1 | 0 | 3 |
| 2A5 |  | 1 | nd | 0 | 1 | nr | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2A6 |  | 0 | nd | 0 | 1 | nr | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2A7 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2A8 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2 A 9 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2A10 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2A11 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2A12 |  | 1 | nd | 0 | $\mathrm{nr} / \mathrm{nr}$ | nr | 0 | 0 | 0 | nd | 1 | nr | 1/3? |
| 2B1 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2B2 |  | $\underline{\mathrm{nr}}$ | nd | nr | nr | 0 | 0 | 0 | 0 | nd | nr | nr | nr |
| 2B3 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2B4 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2B5 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2B6 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2B7 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2B8 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2B9 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2B10 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2B11 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2B12 |  | 0 | nd | 0 | 1 | 1 | 1 | 0 | 0 | nd | 0 | 0 | 21 |
| 2 C 1 |  | 1 | nd | 0 | nr | 1 | 1 | 0 | 0 | nd | 0 | 0 | 21 |
| 2 C 2 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2 C 3 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2 C 4 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2C5 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2C6 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2 C 7 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2C8 |  | 1 | nd | 0 | 1 | nr | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2 C 9 |  | nr | nd | nr | nr | nr | 0 | 0 | nr | nd | nr | 0 | ignore |
| 2 C 10 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2 C 11 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2 C 12 |  | 1 | nd | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2D1 |  | 1 | nd | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| 2D2 |  | 1 | nd | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 |
| 2D3 |  | nr | nd | nr | nr | 0 | 0 | 0 | nr | nd | nr | 0 | nr |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12 f 2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 2D4 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2D5 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2D6 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2D7 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2D8 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2D9 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2D10 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2D11 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | nr | 2 |
| 2D12 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2 E 1 |  | 0 | nd | 0 | $\mathrm{nr} / \mathrm{nr}$ | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2 E 2 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 26 |
| 2 E 3 |  | nr | nd | nr | 1 | 0 | 0 | 0 | nr | nd | nr | nr | nr |
| 2 E 4 |  | nr | nd | 0 | 1 | 0 | 0 | 0 | nr | nd | nr | nr | nr |
| 2 E 5 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2 E 6 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2 E 7 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2 E 8 |  | nr | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 |
| $2 \mathrm{E9}$ |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2 E10 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 1 | 9 |
| 2 E11 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2 E12 |  | 0 | nd | 0 | 0/nr | 0 | 0 | 0 | 0 | nd | 0 | nr | 2 |
| 2F1 |  | 1 | nd | 0 | 0/nr | 0 | 0 | 0 | 0 | nd | 1 | nr | 3 or 1? |
| 2F2 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2F3 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2F4 |  | nr | nd | nr | nr | nr | nr | nr | nr | nd | nr | 0 | ignore |
| 2F5 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2F6 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2 F 7 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2F8 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2F9 |  | 0 | nd | 0 | 1 | nr | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2F10 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2F11 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2F12 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2G1 |  | 1 | nd | 0 | 1 | nr | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2G2 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2G3 |  | 0 | nd | 0 | $0 / \mathrm{nr}$ | nr | 0 | 0 | 0 | nd | 0 | 0 | 2? |
| 2G4 |  | 0 | nd | nr | 0 | nr | 0 | 0 | 0 | nd | 0 | 0 | 2? |
| 2 G 5 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2G6 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2G7 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | nr | 2 |
| 2 G 8 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2G9 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2G10 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| 2G11 |  | 0 | nd | 0 | 0/nr | 0 | 0 | 0 | 0 | nd | 0 | 0 | 7 |
| 2 G 12 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| 2 H 1 |  | 1 | nd | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 or 3? |
| 2 H 2 |  | 0 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | nr | 2 |
| 2H3 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 |
| 2 H 4 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 |
| 2 H 5 |  | nr | nd | 0 | 1 | nr | 0 | 0 | 0 | nd | nr | nr | nr |
| 2H6 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | nr | 1 |
| 2 H 7 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | nr | 1 |
| 2H8 |  | nr | nd | 0 | nr | nr | 0 | 0 | 0 | nd | nr | nr | nr |
| 2H9 |  | nr | nd | 0 | 1 | 0 | 0 | 0 | 0 | 0 | nr | nr | 1?/26? |
| 2H10 |  | nr | nd | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1?/26? |
| 2H11 |  | 1 | nd | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 |
| 2H12 |  | 1 | nd | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 |
| Additional East Anglian |  |  |  |  |  |  |  |  |  |  |  |  |  |
| EA1 |  | 1 | nr | nr | 1 | 0 | 0 | 0 | 0 | 0 | nr | nr | 1 ? |
| EA2 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| EA3 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| EA4 |  | nr | nr | 0 | 1 | 1 | 1 | 0 | 0 | nr | 0 | nr | 21 |
| EA5 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| EA6 |  | nr | nr | nr | 0/nr | 0 | 0 | 0 | 0 | nr | 1 | nr | nr |
| EA7 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| EA8 |  | nr | nr | $0 / \mathrm{nr}$ | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1?/26? |
| EA9 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | nr | 1 |
| EA10 |  | 1 | nr | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | nr | 1 |
| EA11 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| EA12 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| EA13 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| EA14 |  | 0 | nr | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| EA15 |  | 0 | nr | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| EA17 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| EA18 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| EA19 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | nr | 2 |
| EA20 |  | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 0 | 0 | 2 |
| EA21 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| EA22 |  | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nr | nr | 0 | 1?/2? |
| EA23 |  | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 |
| Spanish |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12 f 2 | Haplo |
| Sp1 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp2 | Arsp1 | nr | nr | nr | 1 | nr | 0 | 0 | 1 ? | nd | nr | nr | nr |
| Sp3 | Arsp1 | nr | nr | nr | 1 | nr | 0 | 0 | 0 | nd | nr | 0 | nr |
| Sp4 | ArSp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp5 | Arsp1 | nr | nr | nr | 1 | nr | 0 | 0 | 0 | nd | nr | nr | nr |
| Sp6 | Arsp1 | nr | nr | nr | 1 | nr | 0 | 0 | 0 | nd | nr | nr | nr |
| Sp7 | ArSp1 | nr | nr | nr | 1 | nr | 0 | 0 | 0 | nd | nr | nr | nr |
| Sp8 | Arspl | nr | nr | nr | 1 | nr | 0 | 0 | nr | nd | nr | nr | nr |
| Sp9 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp10 | ArSp1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| Sp11 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp12 | ArSp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp13 | ArSp1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nd | 0 | 0 | 21 |
| Sp14 | Arspl | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| SP15 | Arsp 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp16 | ArSp1 | 0 | 0 | 0 | 1 | 1 | nr | 0 | 0 | nd | 0 | 0 | $21 ?$ |
| Sp17 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp18 | Arsp 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp19 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp20 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp21 | Arspl | 1 | 1 | 1? | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 22? |
| Sp22 | Arsp1 | 1 | nr | nr | 1 | nr | 0 | 0 | 0 | nd | nr | nr | nr |
| Sp23 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | 0 | $1 ?$ |
| Sp24 | Arspl | nr | nr | nr | 1 | nr | 0 | 0 | nr | nd | 1 | nr | nr |
| Sp25 | Arspl | nr | 1 | nr | 1 | nr | 0 | 0 | nr | nd | 1 | nr | nr |
| Sp26 | ArSp1 | nr | nr | nr | 1 | nr | 0 | 0 | nr | nd | one | nr | nr |
| Sp27 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | 0 | $1 ?$ |
| Sp28 | ArSp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp29 | Arspl | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 1 | 9 |
| Sp30 | Arsp1 | nr | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nd | zero | 0 | 21 |
| Sp31 | Arspl | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nd | 0 | nr | 21 |
| Sp32 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | ? |
| Sp33 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 1? |
| Sp34 | Arspl | 0 | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | $2 ?$ |
| Sp35 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | nr | 1 ? |
| Sp36 | Arsp1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| Sp37 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp38 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp39 | Arsp1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nd | nr | 0 | 21 |
| Sp40 | Arsp 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp41 | Arsp 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp42 | Arsp 1 | nr | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nd | one | 0 | 21 |
| Sp43 | Arsp 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 |
| Sp44 | Arsp1 | nr | 1 | nr | 1 | nr | 0 | 0 | 0 | nd | 0 | nr | 2? |
| Sp45 | Arspl | nr | nr | nr | 1 | nr | 0 | 0 | 0 | nd | 0 | nr | $2 ?$ |
| Sp46 | Arsp1 | nr | nr | nr | 1 | nr | 0 | 0 | nr | nd | 1 | nr | nr |
| Sp51 | Arsp1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nd | 1 | 0 | 21 |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sp52 | ArSpl | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| Sp53 | Arspl | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2 |
| Sp54 | Arsp1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | 0 | $2 ?$ |
| Sp55 | Arspl | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 9 |
| Sp56 | Arspl | nr | 1 | nr | 1 | nr | 0 | 0 | 0 | nd | 1 | nr | 1 |
| Sp57 | Arsp1 | nr | nr | nr | 1 | nr | 0 | 0 | 0 | nd | 1 | nr | nr |
| Sp58 | Arsp1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | 0 | $2 ?$ |
| Sp59 | Arspl | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | ? |
| Sp60 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp61 | Arsp1 | nr | 1 | nr | 1 | nr | 0 | 0 | 0 | nd | 1 | nr | 1 |
| Sp62 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp63 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp64 | Arspl | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nd | 0 | 0 | 21 |
| Sp65 | Arspl | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2? |
| Sp66 | Arspl | 0 | 1 | 0 | 0 ? | 0 | 0 | 0 | 0 | nd | nr | nr | 7 ? |
| Sp67 | Arspl | nr | nr | nr | 1 | nr | 0 | 0 | 0 | nd | 1 | nr | nr |
| Sp68 | Arspl | nr | nr | nr | 1 | nr | 0 | 0 | 0 | nd | 1 | nr | nr |
| Sp69 | Arspl | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp70 | Arspl | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | 0 | 2 |
| Sp71 | Arspl | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 |
| Sp72 | Arsp1 | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 1 | 9 |
| Sp73 | ArSpl | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | ?/1 | ? | 2 |
| Sp74 | ArSpl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | nr | 1 |
| Sp75 | Arspl | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | ? | 1 |
| Sp76 | Arspl | 0 | 0 | 0 | 1 | 0 | 0 | 0 | nr | nd | 0 | ? | $2 ?$ |
| Sp77 | Arspl | nr | 1 | 1 | 1 | 0 | 0 | 0 | 0 | nd | 1/0 | 0 | 22? |
| Sp78 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1/0 | 0 | 1+26 |
| Sp79 | Arsp1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | nd | $1+1$ | 0 | 22? |
| Sp80 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp81 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp82 | Arsp1 | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | ?/nr | 0 | 1+26 |
| Sp83 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp84 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp85 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp86 | ArSp1 | 1 | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 |
| Sp87 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 26 |
| Sp88 | Arspl | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1+1 | 0 | ? |
| Sp89 | Arsp 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0/1 | 1 | 9 |
| Sp90 | Arsp1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | nr | nd | $\mathrm{nr} / 1$ | nr | ? |
| Sp91 | Arspl | nr | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 |
| Sp92 | Arspl | nr | 0 | 0 | 1 | 1 | nr | 0 | 0 | nd | 1 | 0 | 21? |
| Sp93 | Arsp1 | nr | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | ? |
| Sp94 | Arsp1 | nr | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1/? | ? | ? |
| Sp95 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1/0 | nr | 1+26 |
| Sp96 | Arspl | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 26 |
| Sp97 | Arsp1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 26 |
| Sp98 | Arspl | nr | nr | nr | 1 | 1 | ? | ? | nr | nd | 1 | nr | $\underline{\mathrm{nr}}$ |
| Sp47 | Arsp2 | nd | 1 | nr | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp48 | Arsp2 | nd | 1 | nr | 1 | 0 | 0 | 0 | nr | nd | 1 | 0 | $1 ?$ |
| Sp49 | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp50 | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp99 | AASP2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp100 | Arsp2 | nd | 1 | 0 | nr | 0 | 0 | 0 | 0 | nd | nr | 0 | ? |
| Sp101 | Arsp 2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp102 | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | $1 ?$ |
| Sp103 | Arsp2 | nd | 1 | nr | nr | nr | 0 | 0 | nr | nd | $\underline{\mathrm{nr}}$ | nr | nr |
| Sp104 | Arsp2 | nd | 1 | nr | nr | nr | 0 | 0 | nr | nd | 0 | nr | nr |
| Sp105 | Arsp2 | nd | nr | nr | nr | nr | 0 | 0 | nr | nd | 0 | 0 | nr |
| Sp106 | Arsp2 | nd | nr | nr | nr | nr | 0 | 0 | 0 | nd | 1 | 0 | nr |
| Sp107 | Arsp 2 | nd | nr | nr | nr | nr | 0 | 0 | 0 | nd | 0 | nr | nr |
| Sp108 | ArSp2 | nd | nr | nr | nr | nr | 0 | 0 | 0 | nd | nr | nr | nr |
| Sp109 | ArSp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | nr | nd | 1 | 0 | $1 ?$ |
| Sp110 | ArSp2 | nd | nr | 0 | 1 | 1 | nr | 0 | 0 | nd | 0 | nr | 21? |
| Sp111 | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1? |
| Sp112 | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |


|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sp113 | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1? |
| Sp114 | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp115 | ArSp2 | nd | 1 | 0 | 1 | 1 | 1 | 0 | 0 | nd | 0 | 0 | 21 |
| Sp116 | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1? |
| Sp117 | Arsp2 | nd | nr | 0 | nr | 0 | 0 | 0 | 0 | nd | 0 | nr | ? |
| Sp118 | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp119 | Arsp2 | nd | nr | nr | nr | nr | 0 | 0 | 0 | nd | 0 | 0 | nr |
| Sp120 | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | nr | nd | 1 | 0 | $1 ?$ |
| Sp121 | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp122 | ArSp2 | nd | nr | nr | nr | nr | 0 | 0 | 0 | nd | nr | nr | nr |
| Sp123 | Arsp2 | nd | 1 | 1 | 1 | 0 | 0 | 0 | nr | nd | 1 | 0 | 22? |
| Sp124 | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp125 | ArSp2 | nd | nr | nr | nr | nr | 0 | 0 | nr | nd | nr | 0 | nr |
| Sp126 | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp1d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | nr | nd | nr | 0 | ? |
| Sp2d | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | nr | nd | 1 | nr | 1? |
| Sp3d | Arsp 2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp4d | Arsp2 | nd | 0 | 0 | nr | 0 | 0 | 0 | 0 | nd | 0 | 0 | ? |
| Sp5d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1? |
| Sp6d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp7d | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | nr | nd | 0 | 0 | 2? |
| Sp8d | Arsp2 | nd | nr | 0 | 1 | 1 | nr | nr | 0 | nd | 0 | 0 | nr ? |
| Sp9d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1? |
| Sp10d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 ? |
| Sp11d | Arsp2 | nd | 0 | 0 | 1 | 0 | 0 | 0 | nr | nd | nr | 0 | ? |
| Sp12d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp13d | Arsp2 | nd | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2? |
| Sp14d | Arsp2 | nd | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | nr | $2 ?$ |
| Sp15d | ArSp2 | nd | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2? |
| Sp16d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1? |
| Sp17d | ArSp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | $1 ?$ |
| Sp18d | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | nr | nr |
| Sp19d | Arsp2 | nd | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2? |
| Sp20d | Arsp2 | nd | nr | 0 | 1 | 1 | nr | 0 | 0 | nd | 0 | nr | ? |
| Sp21d | Arsp2 | nd | 0 | 0 | 1 | 1 | nr | 0 | 0 | nd | 0 | nr | ? |
| Sp22d | ArSp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp23d | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp24d | Arsp2 | nd | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | $n \mathrm{r}$ | 2 ? |
| Sp25d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2? |
| Sp26d | ArSp2 | nd | nr | 0 | nr | 0 | 0 | 0 | 0 | nd | 0 | 0 | $2 ?$ |
| Sp27d | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp 28 d | ArSp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp29d | Arsp2 | nd | $\overline{\mathrm{nr}}$ | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 ? |
| Sp30d | ArSp2 | nd | nr | nr | nr | nr | 0 | 0 | nr | nd | 0 | nr | nr |
| Sp31d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp32d | ArSp2 | nd | 0 | 0 | 1 | 1 | 1 | 0 | 0 | nd | 0 | 0 | 21 |
| Sp33d | ArSp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | $1 ?$ |
| Sp34d | Arsp 2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp35d | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | nr | nr | nr |
| Sp36d | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp37d | Arsp2 | nd | 1 | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3? |
| Sp38d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp39d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1? |
| Sp40d | Arsp 2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp41d | Arsp2 | nd | nr | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | nr | nr |
| Sp42d | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | $1 ?$ |
| Sp43d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp44d | Arsp2 | nd | nr | 0 | 1 | 1 | nr | nr | 0 | nd | 0 | nr | nr |
| Sp45d | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1? |
| Sp46d | Arsp2 | nd | nr | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | nr | 1 ? |
| Sp47d | ArSp2 | nd | nr | 0 | nr | nr | 0 | 0 | 0 | nd | 1 | nr | ? |
| Sp48d | Arsp ${ }^{2}$ | nd | nr | 0 | nr | nr | 0 | 0 | 0 | nd | 1 | nr | ? |
| Sp49d | Arsp2 | nd | 1 | 0 | nr | 0 | 0 | 0 | 0 | nd | 1 | nr | ? |
| Sp50d | Arsp2 | nd | nr | nr | 1 | 0 | 0 | 0 | 0 | nd | nr | nr | nr |
| Sp51d | Arsp2 | nd | nr | 0 | nr | 0 | 0 | 0 | nr | nd | 0 | nr | ? |

## Appendix 1. Biallelic data

|  | Plate | 92R7 | DYS257 | 2627 | 1532 | Yap | 8299 | sY81 | Tat | Lly22g | M9 | 12f2 | Haplo |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sp52d | Arsp2 | nd | nr | 0 | 1 | nr | 0 | 0 | 0 | nd | 0 | nr | nr |
| Sp53d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp54d | ArSp2 | nd | 0 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | $2 ?$ |
| Sp55d | ArSp2 | nd | 0 | 0 | 1 | 1 | 1 | ? | 0 | nd | 0 | 0 | ? |
| Sp56d | Arsp2 | nd | nr | 0 | nr | 0 | 0 | 0 | nr | nd | nr | nr | nr |
| Sp57d | Arsp 2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1? |
| Sp58d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 2? |
| Sp59d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp60d | Arsp2 | nd | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 3 ? |
| Sp61d | Arsp2 | nd | 1 | 0 | 1 | 0 | 0 | 0 | 0 | nd | 1 | 0 | 1 ? |
| Sp62d | Arsp2 | nd | nr | 0 | nr | 1 | ? | ? | 0 | nd | nr | nr | nr |

Appendix 2. Complete biallelic and ASO data for all YAP + samples analysed

|  |  | Biallelic markers |  |  |  |  |  |  |  |  |  |  | ASO markers |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sample |  | 92R7 | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | LLy22g | 12f2 | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap |
| Slov82 | Slovene | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 0 | 3 |
| Co7 | Cortina | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| Co33 | Cortina | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |  | 1 | nd | 0 | 0 | 0 | 25 |
| Co28 | Cortina | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| Co37 | Cortina | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| Bav84 | Bavarian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | ? | 1 | 0 | 1 | 1 | nd | $n \mathrm{r}$ | 0 | nr | 25 |
| Du79 | Dutch | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 1 | 0 | 0 | 33 |
| RN39 | E.Bantu | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nd | 0 | 0 | 0 | 21 |
| ALB74 | E.Bantu | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| ALB55 | E.Bantu | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |  | 1 | 1 | nd | nr | nr | nr | nr |
| LD156 | W.Bantu | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | nr | 8 |
| m67 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m77 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | nd | 0 | 0 | 0 | ? |
| m125 | N.Africa | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 1 | 0 | 0 | 33 |
| m219 | Mongolian | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nd |  | 0 | 1 | 4 |
| NW102 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| RN38 | E.Bantu | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | nd | nr | 0 | 0 | ? |
| JW85 | Tibet | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | nd | nr | 0 | 0 | ? |
| G36 | Greek | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| S194 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| NW27 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Bav81 | Bavarian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | nd | 0 | 0 | 0 | ? |
| NE22 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| NW14 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |  | 1 | nd | 0 | 0 | 0 | 32 |
| NW41 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| m44 | Austarlian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Cy3 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Cy17 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| G6 | Greek | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 |  | 0 | 32 |
| G25 | Greek | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Bav25 | Bavarian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | ? | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Bul2 | Bul non-gyp | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | ? | nr | 0 | 1 | 0 ? | nd | nr | nr | 0 | nr |
| Du97 | Dutch | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | nr | 0 | nr | 32 |
| m211 | Bulgarian | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Bug42 | Bul gyp | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| 16F | French | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| G19 | Gambia | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | nd | 0 | 0 | 0 | 31 |
| G29 | Gambia | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | nd | 0 | 0 | nr | 31 |

Appendix 2. Complete biallelic and ASO data for all YAP+ samples analysed

| Sample |  | 92R7 | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | LLy22g | 12f2 | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| G49 | Gambian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 1 | 0 | 0 | 33 |
| m716 | Bantu | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | ? | 1 | 1 | nd | 0 | 1 | 0 | eight |
| m86 | Mbuti | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nd | 0 | 0 | 0 | 21 |
| Tuk5 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | ? | 1 | 0 | 1 | 0 | nd | nr | 0 | nr | 32 |
| m715 | Bantu | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| G38 | Gambia | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nd | nr | 0 | 0 | 21 |
| RN62 | E.Bantu | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| G39 | Gambia | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m73 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | nr | nr | 0 | 8 |
| m75 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m714 | Bantu | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | nd | nr | nr | 0 | ? |
| MS42 | Japan | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 0 | 4 |
| OK77 | Japan | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | ? | ? | 1 | ? | 1 | 1 | nd | nr | 0 | nr | nr |
| SELK129 | Russia | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | ? | nr | 0 | 1 | 1 | nd | nr | nr | 0 | nr |
| m69 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m538 | Indian | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | ? | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| 17F | French | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | ? | 1 | ? | 1 | 1 | nd | 0 | nr | 0 | ? |
| G10 | Gambian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | ns |
| m66 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | nd | nd | nd | nd | nd | nd | nd | no | nod | nod | ns |
| n8 | Egyptian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | ns |
| BK30 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | ns |
| 651 | Greenland | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| LGL5197 | Finnish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | nr | 0 |  | 25 |
| BE2 | Belarussian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| BE16 | Belarussian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| BE30 | Belarussian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| BE34 | Belarussian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Cv66 | Chuvash | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Cy4 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| Cy11 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| Cy13 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Cy15 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | nr | 0 | nr | 32 |
| Cy22 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Cy23 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| Cy30 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Cy40 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Cy41 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| Cy50 | Cypriot | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| G8 | Greek | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 1 | 0 | 0 | 33 |

Appendix 2. Complete biallelic and ASO data for all YAP+ samples analysed

| Sample |  | 92R7 | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | LLy22g | 12 f 2 | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| G14 | Greek | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| G21 | Greek | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| G23 | Greek | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| G33 | Greek | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| G41 | Greek | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | nr | 32 |
| G44 | Greek | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 1 | 0 | 0 | 33 |
| Bul12 | Bul non-gyp | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Bul14 | Bul non-gyp | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | nr | 0 | 32 |
| Bul27 | Bul non-gyp | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| 750 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | nr | 0 | 0 | 32 |
| 772 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| 773 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 1 | nr | nr | nr | nr | nr | nr | nr | nr |
| 940 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| 948 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| 952 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| 1320 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| 1324 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| 1348 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| 1357 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr |
| 2616 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | nr | 0 | 0 | 32 |
| 2618 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | nr | 0 | 0 | 32 |
| 2635 | Serbian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| Slov19 | Slovenian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | nr | 0 | 32 |
| Slov24 | Slovenian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Slov28 | Slovenian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Slov73 | Slovenian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nd | 0 | 0 | 0 | 4 |
| Slov83 | Slovenian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | nr | 0 | 32 |
| Vb 3 | Ladin | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| Vb7 | Ladin | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| Vg3 | Ladin | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| Co31 | Ladin | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| Co6 | Ladin | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| Co23 | Ladin | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| Co38 | Ladin | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| Co9 | Ladin | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| $\mathrm{Co4}$ | Ladin | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| Co2 | Ladin | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| NE74 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| NE19 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | nr | nr | 0 | 32 |

Appendix 2. Complete biallelic and ASO data for all YAP + samples analysed

| Sample |  | 92R7 | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | LLy22g | 12 f 2 | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| NW89 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| NW45 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| S150 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| S168 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| NW65 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| NW75 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| NW39 | Italian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 |  | 0 | 0 | 25 |
| Sar14 | Sardinian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | nr | 0 | nr | 1 | nd | 1 | 0 | 0 | 33 |
| Sar17 | Sardinian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | nr | nr | 0 | nd | 0 ? | nr | 0 | nr |
| Bav13 | Bavarian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 1 | 0 | 1 | 1 | 1 | nr | 0 | 0 | 25 |
| Bav88 | Bavarian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| Bav91 | Bavarian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | nr | 0 | 0 | 32 |
| Du90 | Dutch | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| 6601 | French-CEPH | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | nr | nr | $n \mathrm{r}$ | nr | nr | 1 | nr | nr | nr | nr | nr |
| 1A6 | E. Anglian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| 1A12 | E. Anglian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| 1D11 | E. Anglian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| 2B12 | E. Anglian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | nr | 0 | 0 | 32 |
| 2 C 1 | E. Anglian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| EA4 | E. Anglian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Bug44 | Bulg gyp | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | nr | nr | nr | nr |
| Bug60 | Bulg gyp | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Bug63 | Bulg gyp | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Bug65 | Bulg gyp | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Sp13 | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| Sp16 | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Sp30 | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 1 | 0 | 0 | 33 |
| Sp31 | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 1 | 0 | 0 | 33 |
| Sp39 | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 1 | 0 | 0 | 33 |
| Sp42 | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Sp51 | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | nr | 0 | 32 |
| Sp64 | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| Sp92 | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | nr | 0 | 0 | 25 |
| Sp110 | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| Sp115 | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 1 | 0 | 0 | 33 |
| Sp32d | Spanish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | nr | 1 | nd | nr | 0 | 0 | nr |
| Mad111 | Madrid | 0 | 0 | 1 | 0 | 1 | nr | nr | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | nd | 0 | nr | 0 | nr |
| Mad189 | Madrid | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | nd | 1 | 0 | 0 | 33 |
| m486 | Indian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |

Appendix 2. Complete biallelic and ASO data for all YAP+ samples analysed

| Sample |  | 92R7. | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | LLy22g | 12 f 2 | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| m538 | Indian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | nd | nr | 0 | 0 | ? |
| m475 | Indian | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m117 | Nigerian | 0 | 0 | 1 | 0 | 1 | 1 | nr | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | nr | 1 ? | nd | nr | nr | nr | nr |
| m118 | Nigerian | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | nr | 0 | 8 |
| m719 | E. Bantu | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m58 | Euro Jewish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| ROMA61 | Romanian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| ROVR207 | Romanian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | nr | 1 | nd | nr | 0 | 0 | 32 |
| ROVR225 | Romanian | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| LIB7 | Lebanese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 1 | nr | nr | 33 |
| LIB39 | Lebanese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | nr | 0 | nr | 25 |
| A04 | Togolese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | nd | 0 | 0 | 0 | 31 |
| A14 | Togolese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | nd | 0 | nr | 0 | ? |
| A21 | Togolese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | nd | 0 | 0 | 0 | 31 |
| A22 | Togolese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| A25 | Togolese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | nd | 0 | 0 | 0 | 31 |
| A26 | Togolese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | nd | 0 | 0 | 0 | 31 |
| IRE743 | Irish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| T05 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | nr | nr | nr | 32 |
| T07 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| T09 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| T17 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| T21 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | nr | 0 | 32 |
| T22 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| T23 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| T25 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| T28 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 25 |
| T38 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| T40 | Turkish | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | nd | 0 | 0 | 0 | 32 |
| G40 | Gambian | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| RN37 | E. Bantu | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| G42 | Gambian | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R9 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R12 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R14 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R15 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R18 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R22 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R24 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |

Appendix 2. Complete biallelic and ASO data for all YAP+ samples analysed

| Sample |  | 92R7 | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | LLy22g | 12 f 2 | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| R26 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 1? | 0 | 1 | 0 | 1 | 1 | nd | nr | 1 | 0 | 32 |
| R27 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R28 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R36 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R37 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R38 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R39 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | nd | nr | nr | nr | nr |
| R44 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R47 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R49 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R53 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R55 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R59 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R60 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R65 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R68 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | $1 ?$ | 0 | 1 | 1 | 1 | 1 | nd | nr | 1 | nr | 8 |
| R72 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R77 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R79 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| R83 | Zimbabwean | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m66 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | ns |
| m68 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | ns |
| m70 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m71 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m72 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m74 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m76 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | nd | 0 | 1 | 0 | 8 |
| m78 | Kenyan | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |  | 1 | nd | 0? | 1 | 0 | 8 |
| 412 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| 426 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| 432 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| 435 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| 439 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 ? | 0 | 0 | 0 | 0 | 1 | nr | 0 | 0 | 4 |
| 447 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 1 ? | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4? |
| 452 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| 457 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | nr | 4 |
| 458 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| 832 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| 849 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |

Appendix 2. Complete biallelic and ASO data for all YAP+ samples analysed

| Sample |  | 92R7 | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | LLy22g | 12 f | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 883 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| 911 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| 913 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| 931 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| 959 | Japanese | 0 | 0 | 1 | 0 | 1 | nd | nd | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| p187/23 | Swede | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | nr | 0 | 0 | 32 |
| PRS15 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| PRS45 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| PRS68 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | nr | 0 | 25 |
| PRS76 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| PRS88 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| MAK12 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |
| MAK26 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| MAK102 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |
| MAK108 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | nr | 0 | 21 |
| MAK119 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 ? | nr | nr | 0 | 8 |
| MAK132 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |
| SYD15 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| SYD24 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| SYD32 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| SYD40 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | nr | 0 | 0 | 25 |
| SDH53 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| SDH55 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| SDH57 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| BRU369 | Pakistani | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |
| YGX16 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ? | 0 | 0 | 1 | 30 |
| Tib1 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | nr | 0 | 4 |
| Tib3 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 30 |
| Tib6 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 30 |
| Tib9 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | nr | 1 | 30 |
| Tib15 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | nr | 1 | 30 |
| Tib19 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | nr | 4 |
| Tib23 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 1 | 0 | 0 | 0 | 4 |
| Tib24 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 ? | nr | 0 | 0 | 4 |
| Tib30 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 1 | nr | 0 | 0 | 4 |
| Tib31 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | nr | 0 | 4 |
| Tib32 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 30 |
| Tib34 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 0 | 1 | 30 |
| Tib35 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |

Appendix 2. Complete biallelic and ASO data for all YAP+ samples analysed

| Sample |  | 92R7 | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | Lly22g. | 12 f 2 | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| HLZ14 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | nr | 0 | 4 |
| HLZ29 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| MAN27 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 30 |
| QIA19 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| QIA21 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| QIA27 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | nr | 0 | $n \mathrm{n}$ | 0 | 1 | 0 | nr | nr | nr |
| QIA29 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 0 | 0 | 1 | 30 |
| QIA32 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 30 |
| QIA34 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 0 | nr | 4 |
| UYG15 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 32 |
| UYG20 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| HUI10 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 30 |
| HUI15 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| HUI21 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | nr | 0 | 0 | 4 |
| HUI23 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 4 |
| HUI24 | Chinese | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 ? | 1 | 1 | 0 | 0 | 0 | 32? |
| SAH2 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH3 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH5 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH6 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH8 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH10 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH11 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH12 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH13 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH14 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH20 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH22 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH23 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH31 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH35 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH40 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH44 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH48 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH49 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH52 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH59 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| SAH60 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1. | 1 | 0 | 0 | 33 |
| ARA3 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |

Appendix 2. Complete biallelic and ASO data for all YAP+ samples analysed

| Sample |  | 92R7 | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | LLy22g | 12 f 2 | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| ARA5 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| ARA6 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| ARA7 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| ARA8 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| BER3 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| INT2 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| INT3 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| INT6 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| INT7 | BERSP | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA9 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA18 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA19 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA23 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA24 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA25 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA26 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA31 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA32 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA33 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA34 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA37 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA39 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | no | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA40 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA150 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA151 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA152 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA155 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA157 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA161 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA163 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA165 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | nr | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA166 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA167 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| AGA169 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB014 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB017 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB019 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB038 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB039 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |

Appendix 2. Complete biallelic and ASO data for all YAP+ samples analysed

| Sample |  | 92R7 | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | LLy22g | 12f2 | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| RAB040 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB050 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB059 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB067 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB070 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB075 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB78 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB104 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB117 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB118 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB119 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB126 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB129 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB130 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB131 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB133 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB134 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB137 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB203 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| RAB204 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MOJ1 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MOJ2 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | nr | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MOJ5 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MOJ4 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MOJ6 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MOJ7 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MOJ8 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MOJ11 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MOJ12 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MRA11 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MNA3 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MNA6 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MNA8 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MNA28 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | nr | 0 | 1 | 1 | 1 | 1 | . 0 | 0 | 33 |
| MRA13 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MOU4 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MOU5 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MSM1 | BERSP | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MSM2 | BERSP | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |

Appendix 2. Complete biallelic and ASO data for all YAP+ samples analysed

| Sample |  | 92R7 | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | LLy 22 g | 12f2 | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| MSM4 | BERSP | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MSM6 | BERSP | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MSM8 | BERSP | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | nr | 1 | 1 | 1 | 0 | 0 | 33 |
| MSM9 | BERSP | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MSM7 | BERSP | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | nr | nr | 1 | 1 | 1 | 0 | 0 | ? |
| MSM13 | BERSP | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MER12 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MER13 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MER17 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| MER18 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | $1 ?$ | ? |
| MNA15 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | nr | 1 | 1 | 1 | 1 | 0 | 0 | 33 |
| ARA14 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| BER2 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| AGA14 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 1 | 0 | $n \mathrm{r}$ | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32? |
| AGA29 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| AGA35 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| AGA159 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| AGA160 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| RAB113 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| RAB115 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| RAB136 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| RAB139 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 32 |
| INT4 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | nr | 0 | 0 | ? |
| AGA28 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| AGA154 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| AGA168 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| MOJ3 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nr | 0 | 1 | 0 | 1 | 1 | 1 | nr | 0 | 0 | ? |
| MNA4 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1. | 0 | 0 | 0 | 25 |
| MNA20 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 25 |
| MSM11 | BERSP | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | nr | 1 | 0 | 1 | 1 | 1 | 0? | 0 | 0 | 25? |
| MER14 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 0 | $1 ?$ | 1 | 1 | 0 ? | 0 | 0 | 25? |
| SAH7 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 31 |
| MNA9 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 31 |
| MNA23 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 31 |
| SAH41 | SAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |
| AGA41 | TAH | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | nr | 0 | 1 | 1 | 1 | 0 | 1 | 0 | ? |
| RAB073 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |
| RAB106 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |
| RAB121 | ARA | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |

Appendix 2. Complete biallelic and ASO data for all YAP+ samples analysed

| Sample |  | 92R7 | 257 | 1532 | 2627 | YAP | 8299 | sY81 | Tat | LLy22g | 12 f 2 | M9 | M78 | M33 | DYS 391 | PN1 | PN2 | 8299 | YAP | M81 | sY81 | M15 | Hap, |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| RAB132 | NBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |
| MOU3 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1? | ? |
| MOU7 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |
| MER11 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |
| MER16 | CBER | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | nd | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 8 |

Appendix 3. Complete microsatellite data for all YAP + chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS3891 | DYS3891I' | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7. 2 | G09411 | Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| m219 | 4 | 189 | 129 | 228 | 168 | 167 | 119 | nr | nr | 114 | 220 | 134 | 186 | 223 | 258 | nr | nr | nr | 106/112 |
| 412 | 4 | 197 | 129 | 224 | 164 | 167 | 123 | 150 | 270 | 118 | 220 | 134 | 186 | 223 | 246 | 183 | 187 | 187 | 106/112 |
| 426 | 4 | nr | nr | 224 | 168 | nr | 127 | 154 | 274 | 118 | 220 | 134 | 186 | 228 | 250 | nr | 183 | 187 | 106/112 |
| 432 | 4 | nr | 129 | 220 | 164 | 167 | 127 | 154 | 270 | 122 | 220 | 140 | 190 | 228 | 250 | 183 | 183 | 187 | '106/112 |
| 435 | 4 | 193 | 129 | 220 | 164 | 170 | 123 | 154 | 274 | 118 | 220 | 134 | 186 | 223 | 250 | 183 | 187 | 187 | nr |
| 452 | 4 | nr | nr | 224 | 164 | 167 | 123 | 154 | 274 | 122 | 220 | 140 | 190 | 223 | 250 | 183 | 187 | 187 | 106/112 |
| 849 | 4 | nr | nr | 216 | 160 | 164 | 119 | nr | nr | 118 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | nr |
| 911 | 4 | nr | nr | nr | nr | nr | 123 | nr | nr | 110 | 216 | 134 | 186 | 218 | 250 | nr | nr | nr | nr |
| 913 | 4 | nr | nr | 218 | 168 | nr | 123 | 158 | nr | 118 | 220 | 134 | 186 | 223 | 254 | 183 | 187 | 187 | nr |
| 959 | 4 | 205 | 129 | 220 | 164 | 167 | 123 | 150 | nr | 118 | 220 | 134 | 186 | 223 | 250 | 183 | 187 | 187 | 106/112 |
| Tib1 | 4 | 193 | 129 | 220 | 164 | 155 | 123 | 154 | 270 | 110 | 220 | 134 | 186 | 228 | 250 | 175 | 183 | 187 | 106/112 |
| Tib19 | 4 | nr | nr | nr | nr | nr | 119 | nr | nr | 110 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | 106/112 |
| Tib23 | 4 | nr | nr | nr | nr | nr | nr | nr | nr | 110 | 220 | 134 | 186 | 228 | 250 | nr | nr | nr | nr |
| Tib24 | 4 | 193 | 129 | 220 | 168 | 155 | 119 | nr | nr | 110 | 220 | 134 | 186 | 228 | 250 | nr | nr | nr | 106/112 |
| Tib30 | 4 | 193 | 129 | 220 | 164 | 155 | 123 | nr | nr | 110 | 220 | 134 | 186 | 228 | 250 | nr | nr | nr | 106/112 |
| Tib31 | 4 | 193 | 129 | 220 | 164 | 155 | 123 | nr | nr | 110 | 220 | 134 | 186 | 228 | 254 | nr | nr | nr | 106/112 |
| Tib35 | 4 | 193 | 129 | 220 | 164 | 155 | 123 | 154 | 270 | 110 | 220 | 134 | 186 | 228 | 250 | 171 | 183 | 187 | 106/112 |
| HLZ14 | 4 | 193 | 129 | 220 | 164 | 155 | 123 | 158 | 274 | 110 | 220 | 134 | 186 | 228 | 250 | 175 | 183 | 187 | 106/112 |
| HLZ29 | 4 | nr | nr | nr | nr | nr | 123 | 158 | nr | 110 | 220 | 134 | 186 | 228 | 250 | 175 | 183 | 187 | 106/112 |
| QIA19 | 4 | 193 | 129 | 220 | 164 | 155 | 123 | nr | nr | 114 | 220 | 134 | 186 | 228 | 250 | nr | nr | nr | 106/112 |
| QIA21 | 4 | 193 | 129 | 216 | 164 | 155 | 127 | $n \mathrm{n}$ | nr | 110 | 220 | 134 | 186 | 228 | 250 | nr | nr | nr | 106/112 |
| QIA34 | 4 | nr | nr | nr | nr | nr | 127 | 158 | nr | 110 | 220 | 134 | 186 | 228 | 250 | 175 | 183 | 187 | 106/112 |
| HUI21 | 4 | 193 | 129 | 220 | 164 | 155 | 127 | nr | nr | 110 | 220 | 134 | 186 | 228 | 250 | nr | nr | nr | 106/112 |
| HUI23 | 4 | 189 | 129 | 212 | 164 | 155 | 127 | nr | nr | 110 | 220 | 134 | 186 | 228 | 250 | nr | nr | nr | 106/112 |
| ALB74 | 8 | 193 | 129 | 204 | 164 | 167 | 123 | 158 | nr | 114 | 220 | 134 | 186 | 228 | 246 | nr | nr | nr | 106/112 |
| LD156 | 8 | nr | 123 | nr | nr | nr | nr | nr | nr | 110 | 220 | 128 | 182 | 228 | 242 | nr | nr | nr | 106/112 |
| m67 | 8 | 193 | 129 | 204 | 164 | 167 | 123 | 158 | 278 | 114 | 220 | 134 | 186 | 233 | 250 | 171 | 183 | 187 | 106/112 |
| RN62 | 8 | 197 | 129 | 208 | 164 | 167 | 127 | 150 | 270 | 114 | 220 | 134 | 186 | 228 | 258 | 175 | 183 | 187 | 106/112 |
| G39 | 8 | 201 | 121 | 204 | 164 | 167 | 123 | 146 | nr | 110 | 220 | 134 | 186 | 228 | 250 | nr | nr | 187 | 106/112 |
| m73 | 8 | nr | nr | nr | nr | nr | nr | nr | nr | 114 | 220 | 134 | 186 | 228 | 250 | nr | nr | nr | 106/112 |
| m75 | 8 | 197 | 129 | 204 | 164 | 164 | 131 | nr | nr | 114 | 220 | 134 | 186 | 228 | 250 | nr | nr | nr | nr |
| m69 | 8 | nr | nr | nr | nr | nr | nr | nr | $n \mathrm{n}$ | 110 | 220 | 140 | nr | nr | 250 | nr | nr | nr | 106/112 |
| m475 | 8 | 193 | 129 | 204 | 164 | 167 | 123 | 154 | 278 | 114 | 220 | 134 | 186 | 228 | 246 | 183 | 187 | 187 | 106/112 |
| m118 | 8 | nr | nr | 204 | 164 | 164 | 127 | 154 | 274 | 114 | 220 | 134 | 186 | 228 | 250 | 183 | 187 | 187 | 106/112 |

Appendix 3. Complete microsatellite data for all YAP + chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS3891 | DYS3891I | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7.1 | YA7.2 | G09411 | Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| m719 | 8 | 193 | 129 | 204 | 164 | 167 | 127 | 150 | 274 | 114 | 220 | 134 | 186 | 228 | 246 | 171 | 183 | 191 | 106/112 |
| G40 | 8 | 197 | 129 | 204 | 164 | 167 | 127 | 150 | 270 | 114 | 220 | 134 | 186 | 228 | 250 | 171 | 183 | 187 | 106/112 |
| RN37 | 8 | 201 | 129 | 204 | 164 | 167 | 131 | 158 | 278 | 114 | 220 | 134 | 186 | 228 | 250 | 175 | 183 | 187 | 106/112 |
| G42 | 8 | 201 | 129 | 200 | 160 | 167 | 127 | 154 | 274 | 114 | 220 | 134 | 186 | 228 | 246 | 175 | 187 | 187 | 106/112 |
| R9 | 8 | nr | nr | 204 | 164 | nr | nr | 154 | 274 | nr | 216 | 128 | nr | nr | 246 | 171 | 187 | 187 | 106/112 |
| R12 | 8 | 201 | 129 | 204 | 164 | nr | 123 | 154 | 282 | 114 | 220 | 134 | 186 | 228 | 250 | nr | 187 | 187 | 106/112 |
| R14 | 8 | nr | nr | 204 | 164 | nr | 135 | 154 | 270 | 114 | 220 | 134 | nr | nr | 250 | 175 | 187 | 187 | 106/112 |
| R15 | 8 | nr | nr | 204 | 164 | 167 | 127 | 154 | 274 | 114 | 220 | 134 | nr | nr | 250 | 171 | 187 | 187 | 106/112 |
| R18 | 8 | nr | nr | nr | nr | nr | 123 | 154 | 278 | 114 | 220 | 134 | 186 | 228 | 246 | 175 | 187 | 187 | 106/112 |
| R22 | 8 | nr | nr | 204 | 164 | nr | 123 | 154 | 278 | nr | 220 | 134 | nr | nr | 250 | 171 | 191 | 187 | 106/112 |
| R24 | 8 | nr | nr | 204 | 164 | 167 | 127 | 154 | 274 | 114 | 220 | 134 | nr | nr | 246 | 175 | 187 | 187 | 106/112 |
| R27 | 8 | nr | nr | 204 | 164 | nr | 127 | 154 | nr | 114 | 220 | 134 | nr | nr | 250 | nr | 183 | 187 | 106/112 |
| R28 | 8 | nr | nr | 204 | 164 | nr | 135 | 154 | 274 | 114 | 220 | 134 | 186 | 228 | 250 | 175 | 187 | 187 | 106/112 |
| R36 | 8 | 193 | 129 | 204 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 228 | 246 | 175 | 191 | 187 | 106/112 |
| R37 | 8 | nr | nr | 204 | 164 | 167 | 131 | 154 | 274 | 114 | 220 | 134 | 186 | 228 | 250 | 171 | 183 | 187 | 106/112 |
| R38 | 8 | $n \mathrm{r}$ | nr | 204 | 164 | 167 | 123 | 154 | 278 | 114 | 220 | 134 | 186 | 228 | 250 | 171 | 183 | 187 | 106/112 |
| R44 | 8 | nr | nr | 204 | 164 | 164 | 123 | 154 | 274 | nr | nr | 128 | nr | nr | nr | nr | 187 | 187 | 106/112 |
| R47 | 8 | 193 | 129 | 204 | 164 | 167 | 127 | 154 | 274 | nr | nr | 140 | nr | nr | nr | 171 | 187 | 187 | 106/112 |
| R49 | 8 | 197 | 129 | 204 | 164 | 167 | 131 | 154 | 274 | 114 | 220 | 134 | 186 | 228 | 250 | 175 | 187 | 187 | 106/112 |
| R53 | 8 | 193 | 129 | 204 | 164 | 167 | 127 | 154 | 278 | 114 | 220 | 134 | nr | nr | 254 | 171 | 187 | 187 | 106/112 |
| R55 | 8 | 193 | 129 | 204 | 164 | 167 | 123 | 154 | 278 | nr | nr | nr | nr | nr | nr | 171 | 187 | 187 | 106/112 |
| R59 | 8 | 201 | 129 | 208 | 168 | 167 | nr | 154 | 274 | nr | nr | 125 | nr | nr | nr | 175 | 187 | 187 | 106/112 |
| R60 | 8 | 197 | 129 | 204 | 164 | 167 | 127 | 157 | nr | 114 | 220 | 134 | 186 | 228 | 250 | nr | 183 | 187 | 106/112 |
| R65 | 8 | nr | nr | 204 | 164 | 167 | 127 | 154 | 278 | 118 | 220 | 134 | nr | nr | 250 | 175 | 183 | 187 | 106/112 |
| R68 | 8 | nr | nr | 204 | 168 | 167 | 131 | 154 | 274 | 114 | 216 | 128 | nr | nr | 242 | 171 | 183 | 187 | 106/112 |
| R72 | 8 | nr | nr | 204 | 164 | 167 | 123 | 154 | 278 | 114 | 220 | 134 | 186 | 228 | 246 | 171 | 187 | 187 | 106/112 |
| R77 | 8 | 197 | 129 | 204 | 168 | 167 | 127 | 154 | 270 | 114 | 220 | 134 | 186 | 228 | 250 | 171 | 187 | 187 | 106/112 |
| R79 | 8 | 197 | 129 | 204 | 164 | 167 | 131 | 154 | 274 | 114 | 220 | 134 | 186 | 228 | 254 | 175 | 191 | 187 | 106/112 |
| R83 | 8 | 193 | 129 | nr | 164 | 167 | 123 | 150 | 274 | nr | nr | 128 | nr | nr | nr | 171 | 187 | 187 | 106/112 |
| m70 | 8 | 193 | 129 | 204 | 164 | 167 | 123 | 154 | 278 | 114 | 220 | 134 | 186 | 228 | 246 | 171 | 187 | 187 | 106/112 |
| m71 | 8 | 193 | 129 | 204 | 164 | 167 | 123 | 158 | 278 | 114 | 220 | 134 | 186 | 228 | 254 | nr | 183 | 187 | 106/112 |
| m72 | 8 | 197 | 129 | 208 | 164 | 167 | nr | nr | nr | 114 | 220 | 134 | 182 | nr | 254 | nr | nr | nr | 112? |
| m74 | 8 | 193 | 129 | 204 | 164 | 170 | nr | 154 | $n \mathrm{r}$ | 114 | 220 | 134 | 186 | 228 | 246 | nr | 187 | 187 | 106/112 |
| m76 | 8 | 193 | 129 | 204 | 164 | 170 | 123 | 154 | 278 | 114 | 220 | 134 | 186 | 228 | 246 | 171 | 187 | 187 | 106/112 |

Appendix 3. Complete microsatellite data for all YAP+ chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS3891 | DYS3891I. | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7.2 | G09411 | Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| m78 | 8 | 189 | 129 | 204 | 168 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | nr | 191 | nr | nr |
| MAK12 | 8 | 193 | 129 | 204 | 168 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 228 | 250 | nr | nr | nr | ${ }^{\mathrm{nr}}$ |
| MAK102 | 8 | 193 | 129 | 204 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 228 | 246 | 171 | 187 | 187 | 106/112 |
| MAK119 | 8 | 201 | 129 | 204 | 164 | 167 | 127 | nr | nr | 118 | 220 | 134 | 186 | 228 | 246 | nr | nr | nr | 106/112 |
| MAK132 | 8 | nr | nr | 204 | 164 | 167 | 127 | 150 | 270 | 118 | 220 | 134 | 186 | 228 | 246 | 171 | 187 | 187 | 106/112 |
| BRU369 | 8 | 195 | 129 | 204 | 168 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 228 | 250 | nr | nr | nr | 106/112 |
| SAH41 | 8 | 193 | 129 | 204 | 164 | 167 | 123 | 138 | 262 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr |
| RAB073 | 8 | 193 | 129 | 204 | 164 | 167 | 127 | 150 | 274 | 114 | 220 | 134 | 186 | 228 | 250 | 171 | 183 | 187 | 106/112 |
| RAB106 | 8 | 193 | 129 | 204 | 164 | 167 | 127 | 150 | 274 | 114 | 220 | 134 | 186 | 228 | 250 | 171 | 183 | 187 | 106/112 |
| RAB121 | 8 | 197 | 129 | 204 | 164 | 167 | 123 | 158 | 278 | 114 | 220 | 134 | 186 | 228 | 250 | 167 | 187 | 187 | 106/112 |
| RAB132 | 8 | 197 | 129 | 204 | 164 | 167 | 127 | 154 | 278 | 114 | 220 | 134 | 186 | 228 | 246 | 171 | 187 | 187 | 106/112 |
| MOU7 | 8 | 193 | 129 | 204 | 164 | 167 | 123 | 154 | 278 | 114 | 220 | 134 | 186 | 228 | 250 | 175 | 191 | 187 | 106/112 |
| MER11 | 8 | 197 | 129 | 204 | 168 | 167 | 123 | 158 | 286 | 114 | 220 | 134 | 186 | 228 | 250 | 171 | 183 | 187 | 106/112 |
| MER16 | 8 | nr | nr | nr | nr | nr | nr | 158 | 278 | nr | nr | nr | nr | nr | nr | 171 | 187 | 187 | 106/112 |
| m716 | eight | 193 | 129 | 204 | 164 | 167 | 123 | 158 | nr | 114 | 220 | 134 | 186 | 228 | 246 | nr | nr | nr | nr |
| RN39 | 21 | 189 | 132 | 212 | 164 | 167 | 127 | 158 | nr | 118 | 220 | 134 | 186 | 228 | 246 | 171 | 179 | 187 | 106/112 |
| m86 | 21 | 189 | 129 | 220 | 164 | 167 | 123 | 150 | nr | 114 | 220 | 134 | 186 | 228 | 246 | 175 | 187 | 187 | 106/112 |
| G38 | 21 | 193 | 135 | 224 | 168 | 170 | 127 | 154 | nr | 123? | 220 | 134 | 186 | $n \mathrm{r}$ | nr | 175 | 183 | 184 | 106/112 |
| MAK108 | 21 | 189 | 132 | 212 | 164 | 164 | 127 | 154 | nr | 114 | 220 | 134 | 194 | 228 | 242 | 171 | 183 | 191 | 106/112 |
| Slov73 | 21 | 201 | 132 | 212 | 164 | 167 | 123 | 154 | nr | 114 | 220 | 134 | 190 | 223 | 254 | 171 | 183 | 187 | nr |
| Co7 | 25 | 185 | 129 | 216 | 164 | 167 | 123 | 150 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 175 | 187 | 183 | 106/112 |
| Co33 | 25 |  |  | nr | nr |  | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | 106/112 |
| Co28 | 25 | 185 | 129 | 216 | 164 | 167 | 123 | 150 | 266 | 114 | 220 | 134 | 186 | 218 | 254 | 175 | 187 | 183 | 106/112 |
| Co37 | 25 |  | 129 | nr | nr |  |  | nr | nr | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | nr | 106/112 |
| Bav84 | 25 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 134 | nr | nr | nr | nr | nr | nr | 106/112 |
| NW102 | 25 | 189 | 129 | 216 | nr | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 254 | nr | nr | nr | 106/112 |
| m715 | 25 | 185 | nr | 216 | nr | 167 | 127 | 142 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 191 | 191 | 106/112 |
| m538 | 25 | 185 | 129 | 220 | 168 | 167 | 123 | 158 | 282 | 114 | 220 | 134 | 186 | 223 | 250 | 171 | 179 | 183 | 106/112 |
| LGL5197 14 | 25 | 185 | 129 | 220 | 160 | 167 | 127 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 246 | 171 | 179 | 187 | 106/112 |
| Cy4 | 25 | 185 | 129 | 220 | 164 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | 183 | 106/112 |
| Cy11 | 25 | 185 | 129 | 216 | 164 | 167 | 123 | 150 | 278 | 114 | 220 | 134 | 186 | 223 | 250 | nr | 183 | 183 | 106/112 |
| Cy23 | 25 | 185 | 129 | 216 | 164 | 170 | 123 | 154 | 282 | 114 | 224 | 134 | 186 | 223 | 250 | 179 | 187 | 183 | 106/112 |
| Cy41 | 25 | 185 | 129 | 216 | 164 | 167 | 119 | 162 | 286 | 118 | 220 | 134 | 186 | 223 | 250 | 183 | 187 | 187 | 106/112 |
| Cy50 | 25 | 185 | 129 | 220 | 164 | 167 | 123 | 154 | 278 | 114 | 220 | 134 | 186 | 223 | 258 | 179 | 183 | 183 | 106/112 |

Appendix 3. Complete microsatellite data for all YAP + chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS389I | DYS3891I | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7.2 | G09411 | Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1320 | 25 | 189 | 129 | 220 | 164 | 167 | 131 | 158 | 278 | 118 | 220 | 134 | 186 | nr | 246 | 175 | 179 | 187 | 106/112 |
| Co31 | 25 | 185 | 129 | 216 | 164 | 167 | 123 | 150 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 175 | 187 | 183 | 106/112 |
| Co6 | 25 | 185 | 129 | 212 | 164 | 167 | 123 | 150 | 274 | 114 | 220 | 134 | 186 | nr | 250 | 175 | 187 | 183 | 106/112 |
| Co23 | 25 | 185 | 129 | 216 | 164 | 167 | 123 | 150 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 175 | 187 | 183 | 106/112 |
| Co38 | 25 | 185 | 129 |  | 164 | 167 | 123 | 150 | 278 | 114 | 220 | 134 | 186 | 218 | 254 | 175 | 187 | 183 | 106/112 |
| Co9 | 25 | 185 | 129 |  | 164 | 167 | 123 | 150 | nr | nr | nr | nr | nr | nr | nr | nr | 183 | 183 | 106/112 |
| $\mathrm{Co4}$ | 25 | 185 | 129 | 216 | 164 |  | 123 | 150 | 274 | 114 | 220 | 134 | 186 | 218 | 254 | 175 | 187 | 183 | 106/112 |
| Co2 | 25 | 185 | 129 | 216 | 164 |  | 123 | 150 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 175 | 187 | 183 | 106/112 |
| NE74 | 25 | 189 | 135 | 216 | 168 | 173 | 123 | 150 | 270 | 114 | 220 | 134 | 186 | 223 | 250 | 171 | 183 | 187 | 106/112 |
| NW39 | 25 | nr | nr | 221? | 164 | nr | 131 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 195? | nr |
| Bav13 | 25 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 134 | nr | nr | nr | nr | nr | nr | 106/112 |
| Du90 | 25 | 185 | 129 | nr | 164 | 167 | check | nr | nr | 114 | 220 | 134 | 186 | 223 | 254 | nr | nr | 183 | nr |
| Sp13 | 25 | nr | nr | 216 | 172 | nr | 127 | 154 | 274 | 114 | 220 | 134 | 186 | 228 | 250 | 183 | 187 | 187 | 106/112 |
| Sp92 | 25 | nr | nr | 212 | 160 | nr | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | 187 | 106/112 |
| Sp110 | 25 | nr | nr | 212 | 160 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | nr | 187 | 187 | 106/112 |
| m486 | 25 | 185 | 129 | nr | nr | 167 | 123 | 154 | nr | 114 | 220 | 134 | 186 | 223 | 254 | nr | 179 | 183 | 106/112 |
| m58 | 25 | nr | nr | nr | nr | nr | 127 | 154 | 274 | 110 | 220 | 134 | 186 | 223 | 250 | 183 | 179 | 187 | 106/112 |
| LIB39 | 25 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 278 | 118 | 220 | 134 | 186 | 223 | 250 | 171 | 183 | 183 | 106/112 |
| A22 | 25 | 185 | 129 | 208 | 160 | 167 | 119 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 167 | 183 | 187 | 106/112 |
| T09 | 25 | nr | nr | 212 | 168 | 167 | 123 | 154 | nr | 114 | 220 | 134 | 186 | 223 | 250 | 171 | 183 | 183 | 106/112 |
| T25 | 25 | 185 | 129 | 216 | 164 | 167 | 127 | nr | nr | 114 | 220 | 134 | 186 | 218 | 254 | nr | nr | nr | 106/112 |
| T28 | 25 | 185 | 129 | 212 | 168 | 167 | 123 | 154 | nr | 114 | 220 | 134 | 186 | 223 | 254 | 171 | 187 | 183 | 106/112 |
| PRS15 | 25 | nr | nr | 216 | 160 | 167 | 123 | 158 | nr | 114 | 220 | 134 | 186 | 223 | 246 | 171 | 179 | 187 | 106/112 |
| PRS45 | 25 | nr | nr | 212 | 168 | nr | 123 | 154 | nr | 114 | 220 | 134 | 186 | 223 | 250 | 171 | 179 | 179 | 106/112 |
| PRS68 | 25 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 282 | 114 | 220 | 134 | 186 | 223 | 250 | 171 | 179 | 179 | 106/112 |
| PRS76 | 25 | 185 | 129 | 212 | 164 | 167 | 123 | 150 | 278 | 114 | 220 | 134 | 186 | 223 | 246 | 171 | 179 | 179 | 106/112 |
| PRS88 | 25 | 185 | 129 | 216 | 164 | 167 | 123 | 158 | 278 | 114 | 220 | 134 | 186 | 223 | 250 | 175 | 183 | 179 | 106/112 |
| MAK26 | 25 | nr | nr | 216 | 164 | nr | 123 | 150 | 278 | 114 | 220 | 134 | 186 | 223 | 250 | 171 | 187 | 183 | 106/112 |
| SYD32 | 25 | 185 | 129 | 220 | 164 | 167 | 127 | nr | nr | 114 | 220 | 131 | 186 | 223 | 250 | nr | nr | nr | 106/112 |
| SYD40 | 25 | 185 | 129 | 216 | 164 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 246 | 175 | nr | 187 | 106/112 |
| SDH53 | 25 | 185 | 129 | 208 | 164 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | 106/112 |
| SDH55 | 25 | 185 | 129 | 208 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 171 | 183 | 183 | 106/112 |
| SDH57 | 25 | 185 | 129 | 208 | 164 | 167 | 123 | 154 | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | 183 | 183 | 106/112 |
| UYG20 | 25 | 199 | nr | nr | nr | nr | nr | nr | nr | 118 | 220 | 134 | nr | nr | 246 | nr | nr | nr | nr |

Appendix 3. Complete microsatellite data for all YAP + chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS389I | DYS389II | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7. 2 | G09411 | Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| AGA28 | 25 | 187 | 129 | 208 | 160 | 167 | 123 | 154 | 286 | 114 | 220 | 134 | 186 | 223 | 254 | nr | nr | nr | nr |
| AGA154 | 25 | 187 | 129 | 208 | 160 | 167 | 123 | 154 | 278 | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | nr |
| AGA168 | 25 | 187 | 129 | 208 | 160 | 167 | 123 | 154 | 278 | 114 | 220 | 134 | 186 | nr | nr | 171 |  | 187 | 106/112 |
| MNA4 | 25 | 185 | 129 | 208 | 160 | 167 | 123 | 154 | 270 | 114 | 220 | 134 | 186 | 223 | 246 | 175 | 191 | 187 | 106/112 |
| MNA20 | 25 | 185 | 129 | 216 | 168 | 167 | 119 | 154 | 278 | 114 | 216 | 134 | 186 | 223 | 250 | 171 | 187 | 187 | 106/112 |
| YGX16 | 30 | nr | nr | nr | nr | nr | nr | nr | nr | 114 | 220 | 134 | 190 | 228 | 246 | nr | nr | nr | 106/112 |
| Tib3 | 30 | nr | nr | nr | nr | nr | 123 | nr | nr | 110 | 220 | 134 | 186 | 228 | 246 | nr | 183 | $n \mathrm{r}$ | 106/112 |
| Tib6 | 30 | nr | 129 | nr | nr | nr | 119 | nr | nr | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | nr | 106/112 |
| Tib9 | 30 | 193 | 129 | 224 | 164 | 164 | 119 | 150 | 270 | 114 | 220 | 134 | 186 | 223 | 246 | 175 | 179 | 187 | 106/112 |
| Tib15 | 30 | 189 | 129 | 220 | nr | 167 | 119 | nr | nr | 110 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | 106/112 |
| Tib32 | 30 | nr | nr | nr | nr | nr | 119 | nr | nr | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | nr | 106/112 |
| Tib34 | 30 | 193 | 129 | 216 | 164 | 167 | 119 | 154 | 270 | 110 | 220 | 134 | 186 | 223 | 242 | 171 | 179 | 187 | 106/112 |
| MAN27 | 30 | nr | nr | nr | nr | nr | nr | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | 106/112 |
| QIA29 | 30 | 193 | 129 | 216 | 168 | 170 | 119 | nr | nr | 114 | 220 | 134 | 186 | 228 | 246 | nr | nr | nr | nr |
| QIA32 | 30 | 193 | 129 | 220 | 164 | 164 | 119 | 150 | 270 | 114 | 220 | 134 | 186 | 223 | 246 | 171 | 183 | 187 | 106/112 |
| HUI10 | 30 | 189 | 129 | 212 | 164 | 167 | 123 | nr | nr | 110 | 220 | 134 | 186 | 223 | 242 | 171 | $n \mathrm{r}$ | nr | 106/112 |
| G19 | 31 | nr | nr | nr | nr | nr | nr | nr | nr | 110 | 220 | 134 | 198 | 213 | 250 | nr | $n \mathrm{n}$ | nr | 106/112 |
| G29 | 31 | 197 | nr | 208 | nr | nr | 123 | nr | nr | 122 | 220 | 134 | 198 | 213 | 246 | nr | nr | nr | 106/112 |
| A04 | 31 | 193 | 129 | 208 | 168 | 167 | 119 | 150 | 270 | 114 | 220 | 134 | nr | nr | 250 | 175 | 183 | 191 | 106/112 |
| A21 | 31 | nr | nr | 208 | 164 | 167 | 119 | 150 | nr | 114 | 220 | 134 | nr | nr | 250 | nr | 183 | 191 | 106/112 |
| A25 | 31 | nr | nr | 212 | 170 | nr | 127 | 150 | nr | nr | nr | 134 | nr | nr | nr | nr | 187 | 191 | 106/112 |
| A26 | 31 | 189 | 129 | 208 | 164 | 167 | 119 | 150 | nr | 114 | 220 | 134 | 202 | 223 | 246 | 171 | 183 | 191 | 106/112 |
| SAH7 | 31 | 201 | 129 | 208 | 160 | 170 | 123 | 138 | 262 | 114 | 220 | 134 | 194 | 223 | 254 | nr | nr | nr | nr |
| MNA9 | 31 | 193 | 129 | 208 | 164 | 167 | 123 | 158 | 274 | 110 | 220 | 134 | nr | nr | 254 | 171 | 191 | 195 | 106/112 |
| MNA23 | 31 | 193 | 129 | 208 | 164 | 167 | 123 | 150 | 266 | 110 | 220 | 134 | nr | nr | 254 | 171 | 191 | 191 | 106/112 |
| G36 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 |  | 114 | 220 | 134 | 186 | 223 | 246 | 167 | 183 | 187 | 106/112 |
| S194 | 32 | nr | nr | nr | nr | nr | nr | nr | nr | 110 | 216 | 128 | nr | nr | 246 | nr | nr | nr | 106/112 |
| NW27 | 32 | nr | nr | nr | nr | nr | nr | nr | nr | nr | 224 | 134 | nr | nr | 250 | nr | nr | nr | 106/112 |
| NE22 | 32 | nr | nr | nr | nr | nr | nr | nr | nr | nr | 220 | 134 | nr | nr | 250 | nr | nr | nr | 106/112 |
| NW14 | 32 | nr | nr | nr | nr | $n \mathrm{r}$ | nr | nr | nr | nr | 220 | 134 | 186 | 223 | nr | nr | nr | nr | 106? |
| NW41 | 32 | nr | nr | nr | nr | nr | nr | nr | nr | nr | 224 | 140 | 190 | 223 | 258 | nr | nr | nr | 106/112 |
| m44 | 32 | nr | nr | nr | 168 | 173 | 135 | nr | nr | check | 220 | 134 | 190 | 223 | 250 | nr | nr | nr | 106/112 |
| Cy3 | 32 | nr | nr | 212 | 164 | 167 | 123 | nr | nr | 114 | 220 | 137 | 186 | 223 | 250 | nr | nr | nr | 106/112 |
| Cy17 | 32 | 189 | 129 | 216 | 164 | 167 | 123 | nr | nr | 114 | 220 | 137 | 186 | 223 | 246 | nr | nr | nr | 106/112 |

Appendix 3. Complete microsatellite data for all YAP + chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS3891 | DYS389II | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7.2 | G09411 | Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| G6 | 32 | 189 | 129 | 220 | nr | 170 | 127 | nr | nr | 114 | 220 | 134 | 186 | 228 | 254 | nr | nr | nr | 106/112 |
| G25 | 32 | 185 | 129 | 212 | nr | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | nr | 106/112 |
| Bav25 | 32 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr |
| Du97 | 32 | 189 | 129 | 216 | 164 | 170 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | 106/112 |
| m211 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 228 | 250 | 167 | nr | 187 | 106/112 |
| Bug42 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 167 | 183 | 187 | 106/112 |
| 16F | 32 | 193 | nr | nr | nr | nr | 131 | 154 | nr | 114 | 220 | 134 | 186 | 223 | 246 |  | 179 | 187 | 106/112 |
| Tuk5 | 32 | 185 | 129 | 212 | nr | 167 | 123 |  | $n$ | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | nr | 106/112 |
| 651 | 32 | nr | nr | nr | nr | nr | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | nr | nr |
| BE2 | 32 | nr | nr | 216 | 164 | nr | 123 | 158 | 278 | 114 | 220 | 134 | 186 | 223 | 250 | 183 | 183 | 187 | 106/112 |
| BE16 | 32 | 185 | 129 | 220 | 164 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | nr | 250 | nr | nr | 191 | 106/112 |
| BE30 | 32 | nr | nr | nr | 164 | nr | 123 | 154 | nr | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | 187 | 106/112 |
| BE34 | 32 | nr | nr | nr | 164 | nr | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 254 | nr | nr | nr | nr |
| Cv66 | 32 | 185 | 129 | 216 | 168 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | nr | 183 | 187 | 106/112 |
| Cy13 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 254 | 183 | 183 | 187 | 106/112 |
| Cy15 | 32 | 181 | 123 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 183 | 183 | 187 | 106/112 |
| Cy22 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 190 | 223 | 250 | 183 | 183 | 187 | 106/112 |
| Су30 | 32 | 185 | 129 | 216 | 164 | 167 | 127 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 246 | 183 | 183 | 187 | 106/112 |
| Cy40 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 238 | 183 | 183 | 187 | 106/112 |
| G14 | 32 | 189 | 129 | 220 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 246 | 183 | 183 | 187 | 106/112 |
| G21 | 32 | 189 | 129 | 216 | 164 | 167 | 123 | 158 | 278 | 114 | 220 | 134 | 186 | 223 | 250 | 167 | 183 | 187 | 112? |
| G23 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 183 | 179 | 187 | nr |
| G33 | 32 | 185 | 129 | 220 | 164 | 167 | 123 | 158 | 278 | 114 | 220 | 134 | 186 | nr | 246 | nr | 183 | 187 | 106/112 |
| G41 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 254 | 183 | 183 | 187 | 106/112 |
| Bul12 | 32 | nr | nr | 216 | 168 | nr | 123 | nr | nr | 114 | 220 | 134 | 186 | nr | 246 | nr | nr | 183 | nr |
| Bul14 | 32 | 189 | 129 | 212 | 160 | 167 | 123 | 158 | 278 | 110 | 220 | 134 | 186 | 223 | 254 | nr | 187 | 187 | 106/112 |
| Bul27 | 32 | 185 | 129 | 216 | 168 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | nr | nr |
| 750 | 32 | nr | 129 | 216 | 164 | nr | 123 | 158 | 278 | 114 | 220 | 134 | 186 | 223 | 246 | 171 | 183 | 187 | 106/112 |
| 772 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 158 | 278 | 114 | 220 | 134 | 186 | 218 | 250 | 167 | 183 | 187 | 106/112 |
| 940 | 32 | 181 | 123 | 220 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 167 | 183 | 187 | 106/112 |
| 948 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | nr | 238 | 167 | 183 | 187 | 106/112 |
| 952 | 32 | 185 | 129 | 212 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | nr | 238 | 167 | 183 | 187 | 106/112 |
| 1324 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 167 | 187 | 187 | 106/112 |
| 1348 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 282 | nr | nr | 134 | nr | nr | nr | 167 | 183 | 187 | 106/112 |

Appendix 3. Complete microsatellite data for all YAP+ chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS389I | DYS389II | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7.2 | G09411 | Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 2616 | 32 | nr | nr | nr | nr | nr | 123 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr |
| 2618 | 32 | nr | nr | nr | nr | nr | 123 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr |
| 2635 | 32 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 134 | nr | nr | $n \mathrm{n}$ | nr | nr | nr | nr |
| Slov19 | 32 | nr | nr | 212 | 164 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | $n \mathrm{r}$ | nr |
| Slov24 | 32 | nr | nr | 216 | 168 | nr | 123 | 150 | 270 | 114 | 220 | 134 | 186 | 223 | 250 | nr | 183 | 187 | 106/112 |
| Slov28 | 32 | nr | nr | 216 | 164 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | 191 | 112? |
| Slov83 | 32 | nr | nr | 216 | 168 | nr | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | nr |
| Vb3 | 32 | nr | 129 | 220 | 164 |  | 123 | 154 | 274 | 114 | 220 | 134 | nr | nr | nr | nr | 183 | 187 | 106/112 |
| Vb7 | 32 | 191 | 129 | nr | nr | nr | nr | nr | nr | 114 | 220 | 134 | nr | nr | 254 | $n \mathrm{n}$ | nr | nr | 106/112 |
| Vg3 | 32 | 185 | 129 | 220 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 167 | 183 | 187 | 106/112 |
| NE19 | 32 | 185 | 129 | 212 | 164 | 167 | 123 | 154 | 274 | nr | nr | nr | nr | nr | nr | 167 | 183 | 187 | 106/112 |
| NW89 | 32 | 185 | 132 | 216 | 164 | 167 | 123 | 154 | 270 | 114 | 220 | 134 | 186 | 223 | 250 | 171 | 183 | 187 | 106/112 |
| NW45 | 32 | nr | nr | 216 | 164 | check | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 167 | 187 | 187 | 106/112 |
| S150 | 32 | 185 | 129 | 222 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 171 | 183 | 187 | 106/112 |
| S168 | 32 |  |  | 216 | 164 |  | 123 | 154 | nr | 114 | 220 | 134 | 186 | 228 | 250 | nr | 179 | 187 | 106/112 |
| NW65 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 278 | nr | nr | 134 | nr | nr | nr | 167 | 183 | 187 | 106/112 |
| NW75 | 32 | nr | nr | 216 | 164 | nr | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | nr |
| Bav88 | 32 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 134 | nr | nr | nr | nr | nr | nr | nr |
| Bav91 | 32 | nr | nr | nr | nr | nr | 123 | 154 | nr | nr | nr | 134 | nr | nr | nr | nr | 183 | nr | 106/112 |
| 1A6 | 32 | nr | nr | nr | 168 | nr | 127 | 158 | nr | 114 | 220 | 134 | 186 | nr | 254 | nr | 183 | nr | 106/112 |
| 1A12 | 32 | 181 | 129 | 220 | 164 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | nr | 246 | nr | nr | $n \mathrm{n}$ | nr |
| 1D11 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 158 | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | 187 | 191 | nr |
| 2B12 | 32 | 181 | 123 | 216 | 164 | 167 | 127 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 179 | 183 | 187 | 106/112 |
| 2C1 | 32 | 189 | 129 | nr | nr | 170 | nr | 154 | nr | 114 | 220 | 134 | 186 | nr | 254 | nr | 187 | 191 | nr |
| EA4 | 32 | 185 | 129 | 220 | 164 | 167 | 127 | 154 | 274 | 114 | 220 | 134 | nr | nr | 246 | nr | 183 | 187 | 106/112 |
| Bug60 | 32 | nr | nr | 216 | 164 | 167 | 127 | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | 191 | 106/112 |
| Bug63 | 32 | nr | rr | nr | 164 | nr | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | 183 | nr |
| Bug65 | 32 | nr | nr | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | nr | 183 | 187 | 106/112 |
| Sp16 | 32 | nr | nr | 212 | 164 | nr | 123 | 150 | 270 | 114 | 220 | 134 | 186 | 223 | 250 | 182 | 183 | 183 | 106/112 |
| Sp42 | 32 | 185 | 132 | 208 | 156 | 167 | 119 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 183 | 183 | 187 | 106/112 |
| Sp51 | 32 | 189 | 129 | nr | nr | 167 | nr | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | nr | 187 | 187 | 106/112 |
| Sp64 | 32 | 185 | 123 | 216 | 164 | nr | nr | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | nr | 183 | 191 | 106/112 |
| ROMA61 | 32 | 185 | 129 | 216 | 168 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | 187 | 187 | nr |
| ROVR207 | 32 | nr | nr | 212 | 164 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 254 | 182 | 183 | 183 | nr |

Appendix 3. Complete microsatellite data for all YAP+ chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS3891 | DYS3891I | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7.1 | YA7. 2 | G09411 | 1 Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| ROVR225 | 32 | nr | nr | 216 | 164 | nr | 119 | 154 | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | 183 | 187 | 106/112 |
| IRE743 | 32 | nr | nr | 212 | 170 | nr | 129 | nr | nr | nr | nr | 134 | nr | nr | nr |  | nr | nr | 106/112 |
| T05 | 32 | nr | nr | nr | 164 | nr | 123 | nr | nr | 114 | 220 | 134 | 186 | 218 | 250 | nr | nr | 183 | nr |
| T07 | 32 | nr | nr | nr | 164 | nr | 123 | 154 | nr | 114 | 220 | 134 | 186 | 223 | 250 | 167 | 187 | 187 | 106/112 |
| T17 | 32 | 185 | 129 | nr | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | nr | nr | 246 | 167 | 183 | 187 | 106/112 |
| T21 | 32 | 185 | 129 | 212 | 168 | 167 | 123 | 154 | 274 | nr | nr | 134 | nr | nr | nr | 167 | 183 | 187 | 106/112 |
| T22 | 32 | 185 | 129 | 216 | 168 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | nr | nr | 250 | 167 | 183 | 187 | 106/112 |
| T23 | 32 | 185 | 129 | 216 | 168 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | nr |
| T38 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 246 | 167 | 183 | 187 | 106/112 |
| T40 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | nr | 114 | 220 | 134 | 186 | 223 | 250 | 167 | 187 | 187 | 106/112 |
| R26 | 32 | 197 | 129 | 204 | 164 | 167 | 131 | nr | 274 | nr | nr | 140 | nr | nr | nr | nr | 183 | 187 | 106/112 |
| p187/23 | 32 | nr | nr | 220 | 164 | 167 | 123 | 154 | 274 | 114 | 216 | 128 | 186 | 223 | 250 | 175 | 183 | 191 | 106/112 |
| SYD15 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 167 | 187 | 187 | 106/112 |
| SYD24 | 32 | 185 | 129 | 220 | 164 | 167 | 123 | 158 | 278 | 114 | 220 | 134 | 186 | 223 | 250 | 171 | 179 | 187 | 106/112 |
| UYG15 | 32 | nr | nr | nr | nr | nr | nr | 154 | nr | nr | nr | 134 | nr | nr | nr | nr | nr | nr | 106/112 |
| HUI15 | 32 | nr | nr | nr | nr | nr | nr | nr | nr | 118 | 228 | 140 | nr | nr | 250 | nr | nr | nr | 106/112 |
| ARA14 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 138 | 258 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| BER2 | 32 | 193 | 129 | 220 | 164 | 167 | 123 | 142 | 262 | 114 | 220 | nr | nr | nr | 246 | nr | nr | nr | nr |
| AGA29 | 32 | 193 | 129 | 216 | 164 | 167 | 123 | 142 | 266 | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | nr | nr |
| AGA35 | 32 | 185 | 132 | 220 | 164 | 167 | 123 | 142 | 266 | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | nr |
| AGA159 | 32 | 185 | 129. | 220 | 164 | $167{ }^{\circ}$ | 123 | 138 | 258 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA160 | 32 | 185 | 129 | 216 | 164 | 167 | 123 | 142 | 266 | nr | nr | 128 | nr | nr | nr | nr | nr | nr | nr |
| RAB113 | 32 | 185 | 129 | 212 | 164 | 167 | 111 | 150 | 270 | 114 | 220 | 134 | 186 | 223 | 242 | 171 | 187 | 191 | 106/112 |
| RAB115 | 32 | 189 | 129 | 212 | 164 | 167 | 123 | 150 | 270 | nr | 220 | 134 | nr | nr | 242 | 171 | 187 | 191 | 106/112 |
| RAB136 | 32 | 185 | 129 | 212 | 164 | 167 | 123 | 150 | 270 | 114 | 220 | 134 | 186 | 223 | 242 | 171 | 183 | 191 | 106/112 |
| RAB139 | 32 | 185 | 129 | 212 | 164 | 167 | 123 | 150 | 270 | 114 | 220 | 134 | 186 | 223 | 242 | 171 | 187 | 191 | 106/112 |
| Du79 | 33 | 189 | 129 | 216 | 160 | 167 | 123 | nr | nr | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | 106/112 |
| m125 | 33 | 185 | 129 | 220 | 160 | 167 | 127 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 246 | 175 | 187 | 187 | 106/112 |
| G49 | 33 | 185 | 129 | 216 | nr | 158 | 123 | nr | nr | 122 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | 106/112 |
| G8 | 33 | 185 | 129 | 216 | 164 | 167 | 123 | 158 | 278 | 114 | 220 | 134 | 186 | 223 | 242 | nr | 187 | 191 | nr |
| G44 | 33 | 185 | 129 | 220 | 164 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 183 | 187 | 187 | 106/112 |
| Sar14 | 33 | nr | 129 | nr | nr | nr | nr | nr | nr | 114 | 216 | 134 | 190 | nr | 242 | nr | nr | nr | 106/112 |
| Sp30 | 33 | 185 | 129 | 220 | 160 | 167 | 123 | 154 | 270 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 187 | 187 | 106/112 |
| Sp31 | 33 | 181 | 123 | 208 | 156 | nr | 119 | 154 | 270 | 114 | 220 | 134 | 186 | 223 | 242 | nr | 187 | 187 | 106/112 |

Appendix 3. Complete microsatellite data for all YAP + chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS389I | DYS38911, | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7. 2 | G09411 | 1 Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Sp39 | 33 | 185 | 129 | 212 | 160 | 170 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 183 | 187 | 187 | 106/112 |
| Sp115 | 33 | nr | nr | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | nr | 183 | 187 | 106/112 |
| Mad189 | 33 | nr | nr | 208 | 160 | nr | 131 |  | nr | nr | 220 | 134 | nr | nr | 246 | nr | nr | nr | 106/112 |
| LIB7 | 33 | nr | nr | 216 | 160 | nr | 123 | 158 | 270 | nr | 220 | 134 | nr | nr | 242 | nr | 187 | 187 | 106/112 |
| SAH2 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH3 | 33 | 185 | 129 | 212 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH5 | 33 | 185 | 129 | 220 | 160 | 167 | 123 | . 146 | 266 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH6 | 33 | 185 | 129 | 220 | 160 | 167 | 123 | 146 | 266 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH8 | 33 | 185 | 129 | 220 | 160. | 170 | 123 | 146 | 266 | nr | nr | 134 | nr | nr | 246 | nr | nr | nr | nr |
| SAH10 | 33 | 185 | 129 | 220 | 160 | 170 | 123 | 146 | 266 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH11 | 33 | 185 | 129 | 220 | 160 | 167 | 123 . | 146 | 266 | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | nr | nr |
| SAH12 | 33 | 185 | 129 | 220 | 160 | 170 | 123 | 146 | 266 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH13 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH14 | 33 | 185 | 129 | 220 | 160 | 167 | 123 | 146 | 266 | 114 | 220 | 134 | nr | nr | nr | nr | nr | nr | nr |
| SAH20 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 142 | 258 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH22 | 33 | 185 | 129 | 220 | 160 | 170 | 123 | 146 | 266 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH23 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH31 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH35 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH40 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | nr |
| SAH44 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 266 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH48 | 33 | 185 | 129 | 220 | 160 | 170 | 123 | 146 | 266 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH49 | 33 | 185 | 129 | 220 | 160 | 170 | 123 | 146 | 266 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH52 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH59 | 33 | 185 | 129 | 220 | 160 | 167 | 123 | 146 | 266 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| SAH60 | 33 | 185 | 129 | 220 | 160 | 167 | 123 | 146 | 266 | 114 | 220 | 134 | nr | nr | 242 | nr | nr | nr | nr |
| ARA3 | 33 | 185 | 129 | 216 | 164 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | nr | nr |
| ARA5 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| ARA6 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | - 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| ARA7 | 33 | 185 | 132 | 212 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| ARA8 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 142 | 262 | 114 | 220 | 134 | 186 | 223 | 246 | nr | nr | nr | nr |
| BER3 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 142 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| INT2 | 33 | 185 | 129 | 220 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | nr | 242 | nr | nr | nr | nr |
| INT3 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |

Appendix 3. Complete microsatellite data for all YAP+ chromosomes analysed

| Sample | Hap | DYS19 | DYS388: | DYS390 | DYS391 | DYS392 | DYS393 | DYS3891 | DYS38911 | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7.1 | YA7.2 | G09411 | Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| INT6 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| INT7 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA9 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA18 | 33 | 185 | 132 | 216 | 160 | 167 | 123 | 142 | 258 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA19 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA23 | 33 | 185 | 129 | 220 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA24 | 33 | 189 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | nr | nr | 134 | nr | nr | nr | nr | nr | nr | nr |
| AGA25 | 33 | 185 | 129 | 212 | 160 | 167 | 123 | 142 | 258 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA26 | 33 | 185 | 129 | 212 | 156 | 167 | 123 | 146 | 262 | nr | nr | 140 | nr | nr | nr | nr | nr | nr | nr |
| AGA31 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 142 | 258 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA32 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA33 | 33 | 189 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA34 | 33 | 189 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA37 | 33 | 189 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA39 | 33 | 185 | 129 | 212 | 160 | 167 | 123 | 150 | 266 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA40 | 33 | 197 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA150 | 33 | 185 | 129 | 212 | 160 | 167 | 123 | 154 | 270 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA151 | 33 | 185 | 129 | 216 | 160 | 167 | 127 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA152 | 33 | 185 | 129 | 216 | 164 | 167 | 123 | 146 | 262 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr |
| AGA155 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA157 | 33 | 185 | 129 | 216 | 168 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA161 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA163 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA165 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | nr | nr | 242 | nr | nr | $n \mathrm{r}$ | nr |
| AGA166 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 142 | 258 | 114 | 220 | 134 | nr | nr | nr | nr | nr | nr | nr |
| AGA167 | 33 | 185 | 129 | 212 | 160 | 167 | 123 | 142 | 258 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| AGA169 | 33 | 185 | 129 | 212 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | 167 | 187 | 179 | 106/112 |
| RAB014 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | nr | nr | 134 | nr | nr | nr | 175 | 187 | 187 | 106/112 |
| RAB017 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 179 | 187 | 187 | 106/112 |
| RAB019 | 33 | 185 | 129 | 212 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | nr | 187 | 187 | 106/112 |
| RAB038 | 33 | 185 | 129 | 216 | 160 | 170 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 187 | 187 | 106/112 |
| RAB039 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | nr | nr | 134 | nr | nr | nr | nr | 187 | 187 | 106/112 |
| RAB040 | 33 | 185 | 132 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | nr | nr | 242 | 175 | 187 | 187 | 106/112 |
| RAB050 | 33 | 185 | 129 | 212 | 168 | 167 | 123 | 158 | 274 | nr | nr | 134 | nr | nr | nr | nr | 187 | 187 | 106/112 |

Appendix 3. Complete microsatellite data for all YAP+ chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392, | DYS393 | DYS389I | DYS3891I | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7. 2 | G09411 | Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| RAB059 | 33 | 189 | 129 | 216 | 164 | 167 | 123 | 158 | 274 | nr | nr | 134 | nr | nr | nr | 175 | 187 | 187 | 106/112 |
| RAB067 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 278 | nr | nr | 134 | nr | nr | nr | 171 | 187 | 187 | 106/112 |
| RAB070 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 162 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 179 | check | 183 | 106/112 |
| RAB075 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 187 | 187 | 106/112 |
| RAB78 | 33 | 189 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 246 | nr | 187 | 187 | 106/112 |
| RAB104 | 33 | 185 | 129 | 216 | 164 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 246 | 175 | 187 | 187 | 106/112 |
| RAB117 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 187 | 187 | 106/112 |
| RAB118 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 183 | 187 | 106/112 |
| RAB119 | 33 | 185 | 129 | 216 | 164 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 187 | 187 | 106/112 |
| RAB126 | 33 | 185 | 129 | 216 | 168 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 246 | 175 | 187 | 187 | 106/112 |
| RAB129 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 246 | 175 | 187 | 187 | 106/112 |
| RAB130 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 150 | 270 | 114 | 220 | 134 | 186 | 223 | 242 | 167 | 179 | 179 | 106/112 |
| RAB131 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 171 | 187 | 187 | 106/112 |
| RAB133 | 33 | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 246 | 175 | 187 | 187 | 106/112 |
| RAB134 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 191 | 187 | 106/112 |
| RAB137 | 33 | 185 | 129 | 212 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 228 | 242 | 175 | 187 | 187 | 106/112 |
| RAB203 | 33 | 185 | 129 | 220 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 187 | 187 | 106/112 |
| RAB204 | 33 | 185 | 129 | 212 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 187 | 187 | 106/112 |
| MOJ1 | 33 | 185 | 129 | 212 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | nr | nr | 246 | 175 | 187 | 187 | 106/112 |
| MOJ2 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 187 | 187 | 106/112 |
| MOJ5 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | nr | 238 | 175 | 187 | 187 | 106/112 |
| MOJ4 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 154 | 270 | nr | nr | 134 | nr | nr | nr | 167 | 179 | 187 | 106/112 |
| MOJ6 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 187 | 187 | 106/112 |
| MOJ7 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 162 | 278 | nr | nr | 140 | nr | nr | nr | 175 | 179 | 187 | 106/112 |
| MOJ8 | 33 | 185 | 129 | 216 | 164 | 167 | 123 | nr | 278 | nr | nr | 134 | nr | nr | nr | nr | 187 | 195 | 106/112 |
| MOJ11 | 33 | 189 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | nr | 242 | 175 | 183 | 187 | 106/112 |
| MOJ12 | 33 | 185 | 129 | 216 | 160 | 167 | 127 | 158 | 274 | 110 | 220 | 134 | nr | nr | 242 | 175 | 187 | 187 | 106/112 |
| MRA11 | 33 | 185 | 123 | 216 | 160 | 161 | 123 | 158 | 274 | 122? | 220 | 140 | nr | nr | 246 | 167 | 179 | 187 | 106/112 |
| MNA3 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | nr | nr | 114 | 220 | 134 | nr | nr | nr | nr | 187 | 187 | 106/112 |
| MNA6 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 270 | 114 | 220 | 134 | 186 | nr | 246 | 175 | 187 | 187 | 106/112 |
| MNA8 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | nr | 242 | 175 |  | 187 | 106/112 |
| MNA28 | 33 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr |
| MRA13 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | nr | 242 | 175 | 187 | 187 | 106/112 |
| MOU4 | 33 | 185 | 132 | 216 | 160 | 167 | 123 | 154 | 270 | nr | nr | 134 | nr | nr | nr | 175 | 187 | 187 | 106/112 |

Appendix 3. Complete microsatellite data for all YAP+ chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS3891 | DYS38911 | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7.2 | G09411 | Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| MOU5 | 33 | 189 | 129 | 220 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 171 | 187 | 187 | 106/112 |
| MSM1 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 171 | 187 | 187 | 106/112 |
| MSM2 | 33 | 189 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 191 | 187 | 106/112 |
| MSM4 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | 186 | 223 | 238 | 175 | 187 | 187 | 106/112 |
| MSM6 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | nr | nr | nr | nr | nr | nr | 175 | 187 | 183 | 106/112 |
| MSM8 | 33 | 185 | 132 | 216 | 160 | 167 | 123 | 158 | 270 | 114 | 220 | 134 | 186 | 223 | 242 | 175 | 187 | 187 | 106/112 |
| MSM9 | 33 | 193 | 138 | 224 | 168 | 173 | 131 | nr | nr | nr | nr | 134 | nr | nr | nr | nr | nr | 195 | 106/112 |
| MSM13 | 33 | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | nr | nr | nr | 171 | 187 | 183 | 106/112 |
| MER12 | 33 | 185 | 132 | 216 | 164 | 170 | 127 | 158 | 274 | nr | nr | nr | nr | nr | nr | 179 | 187 | 187 | 106/112 |
| MER13 | 33 | 185 | 129 | 216 | 164 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | nr | nr | nr | 175 | 187 | 187 | 106/112 |
| MER17 | 33 | nr | nr | nr | nr | nr | nr | nr | nr | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| MNA15 | 33 | nr | nr | nr | nr | nr | nr | nr | nr | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| MS42 | 34 | nr | 129 | nr | 164 | nr | 123 | nr | nr | 114 | 220 | 134 | 186 | nr | 250 | nr | nr | nr | 106/112 |
| 439 | 34 | 201 | 129 | 216 | 168 | 167 | 123 | 158 | 282 | 114 | 220 | 134 | 186 | 223 | 246 | 183 | 183 | 187 | 106/112 |
| 457 | 34 | 201 | 129 | 224 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 254 | 179 | 183 | 183 | 106/112 |
| 458 | 34 | 201 | 129 | 224 | 164 | 167 | 123 | 158 | 278 | 118 | 220 | 134 | 186 | 223 | 250 | 183 | 191 | 187 | 106/112 |
| 832 | 34 | 205 | 135 | nr | nr | 173 | 127 | nr | nr | 114 | 220 | 134 | 186 | nr | nr | nr | nr | 191 | 112? |
| 883 | 34 | 201 | 129 | 220 | 164 | 167 | 123 | 158 | nr | 114 | 220 | 134 | 186 | 223 | 250 | 183 | 187 | 187 | 106/112 |
| 931 | 34 | 201 | 129 | 224 | 164 | 167 | 123 | 154 | 274 | 114 | 216 | 134 | nr | nr | 246 | 182 | 183 | 183 | 106/112 |
| MSM11 | 25? | 185 | 129 | 216 | 160 | 167 | 123 | 158 | 274 | 114 | 220 | 134 | nr | nr | 242 | 175 | 191 | 187 | 106/112 |
| MER14 | 25? | nr | nr | nr | nr | nr | nr | nr | 274 | nr | nr | nr | nr | nr | nr | 175 | 183 | 187 | 106/112 |
| HUI24 | 32? | nr | nr | nr | nr | nr | nr | nr | nr | nr | 216 | 134 | nr | nr | 250 | nr | nr | nr | 106/112 |
| AGA14 | 32? | 185 | 129 | 212 | 164 | 167 | 127 | 150 | 270 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | nr | nr |
| 447 | 34 | nr | nr | 220 | 172 | nr | 123 | 158 | 278 | 114 | 220 | 134 | 186 | 223 | 254 | 183 | 187 | 187 | 106/112 |
| m77 | ? | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 134 | nr | nr | nr | nr | nr | nr | 106/112 |
| RN38 | ? | 189 | 129 | 220 | 164 | 167 | 123 | 150 | nr | 114 | 220 | 134 | 186 | 223 | 246 | nr | 187 | 187 | 106/112 |
| JW85 | ? | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 134 | nr | nr | nr | nr | nr | nr | nr |
| Bav81 | ? | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 134 | nr | nr | nr | nr | nr | nr | 106? |
| m714 | ? | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 128 | $n \mathrm{r}$ | nr | nr | nr | nr | nr | 106/112 |
| 17F | ? | 185 | nr | nr | nr | nr | 123 | nr | nr | nr | 216 | 128 | nr | nr | 242 | nr | nr | nr | 106/112 |
| m538 | ? | nr | nr | nr | 164 | nr | 123 | nr | 282 | $n \mathrm{r}$ | 220 | 134 | 186 | 223 | 250 | nr | 179 | 183 | 106/112 |
| A14 | ? | 189 | 129 | nr | 168 | 167 | 123 | 154 | 270 | 114 | 220 | 134 | nr | nr | 246 | nr | 183 | 187 | 106/112 |
| MSM7 | ? | 185 | 129 | 216 | 160 | 167 | 123 | nr | nr | nr | 220 | 134 | 186 | 223 | 242 | 175 | 187 | 187 | 106/112 |
| MER18 | ? | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 134 | nr | nr | nr | nr |  | nr | 106/112 |

Appendix 3. Complete microsatellite data for all YAP+ chromosomes analysed

| Sample | Hap | DYS19 | DYS388 | DYS390 | DYS391 | DYS392 | DYS393 | DYS3891 | DYS389II | DYS434 | DYS435 | DYS436 | DYS437 | DYS438 | DYS439 | YA7. 1 | YA7. 2 | G09411 | Amel |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| INT4 | ? | 185 | 129 | 216 | 160 | 167 | 123 | 146 | 262 | 114 | 220 | 134 | 186 | 223 | 242 | nr | nr | $n \mathrm{r}$ | nr |
| MOJ3 | ? | 181 | 123 | 212 | 156 | 161 | 119 | 150 | 270 | nr | nr | nr | nr | nr | nr | 167 | 179 | 179 | 106/112 |
| AGA41 | ? | 197 | 129 | 204 | 164 | 167 | 127 | 138 | 258 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr |
| MOU3 | ? | 193 | 129 | 204 | 164 | 167 | 119 | 154 | 278 | 114 | 220 | 134 | 186 | nr | 250 | 171 | 187 | 187 | 106/112 |
| ALB55 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 224 | 134 | nr | nr | 254 | nr | nr | nr | 106/112 |
| Bul2 | nr | nr | nr | nr | nr | nr | nr | nr | nr | 114 | 220 | 134 | 186 | 223 | 250 | nr | nr | nr | 106/112 |
| OK77 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 106/112 |
| SELK129 | nr | nr | 129 | 216 | 164 | 167 | 123 | 154 |  | 114 | 220 | 134 | 186 | nr | 246 | 167 | 183 | nr | 106/112 |
| 773 | nr | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 167 | 183 | 187 | 106/112 |
| 1357 | nr | 181 | 123 | nr | nr | nr | nr | nr | nr | nr | nr | 134 | nr | nr | nr | nr | nr | nr | nr |
| Sar17 | nr | nr | nr | nr | nr | nr | nr | nr | nr | 114 | 220 | 134 | nr | nr | 250 | nr | nr | nr | nr |
| 6601-CEPH | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr |
| Bug44 | nr | 185 | 129 | 216 | 164 | 167 | 123 | 154 | 274 | 114 | 220 | 134 | 186 | 223 | 250 | 182 | 183 | 187 | 106/112 |
| Sp32d | nr | nr | nr | nr | nr | nr | 123 | 154 | 270 | 106 | 220 | 134 | 186 | 218 | 246 | nr | 187 | 187 | 106/112 |
| Mad111 | nr | 185 | 129 | nr | 164 | nr | 123 | nr | nr | nr | 220 | 134 | nr | nr | 246 | nr | nr | nr | 106/112 |
| m117 | nr | nr | nr | nr | 168 | nr | 127 | nr | nr | 122 | nr | 140 | nr | 233 | 254 | nr | 183 | 191 | inr |
| R39 | nr | 193 | 129 | 204 | 164 | 167 | 123 | 150 | 274 | 114 | 216 | 128 | nr | 223 | 242 | 179 | 187 | 187 | 106/112 |
| QIA27 | nr | nr | nr | nr | nr | nr | nr | nr | nr | nr | 216 | 134 | nr | nr | 246 | nr | nr | nr | 106/112 |
| MNA41 | nr | nr | nr | nr | nr | nr | nr | nr | nr | 114 | 220 | 134 | 186 | 228 | 258 | nr | nr | nr | nr |

# Y-Chromosomal Diversity in Europe Is Clinal and Influenced Primarily by Geography, Rather than by Language 

Zoë H. Rosser, ${ }^{1}$ Tatiana Zerjal, ${ }^{2}$ Matthew E. Hurles, ${ }^{1, *}$ Maarja Adojaan, ${ }^{5}$ Dragan Alavantic, ${ }^{6}$ António Amorim, ${ }^{7}$ William Amos, ${ }^{8}$ Manuel Armenteros, ${ }^{9}$ Eduardo Arroyo, ${ }^{10}$ Guido Barbujani, ${ }^{11}$ Gunhild Beckman, ${ }^{12}$ Lars Beckman, ${ }^{12}$ Jaume Bertranpetit, ${ }^{13}$ Elena Bosch, ${ }^{13,+}$ Daniel G. Bradley, ${ }^{14}$ Gaute Brede, ${ }^{15}$ Gillian Cooper, ${ }^{8}$ Helena B. S. M. Côrte-Real, ${ }^{16}$ Peter de Knijff, ${ }^{17}$ Ronny Decorte, ${ }^{18}$ Yuri E. Dubrova, ${ }^{1}$ Oleg Evgrafov, ${ }^{19}$ Anja Gilissen, ${ }^{18}$ Sanja Glisic, ${ }^{6}$ Mukaddes Gölge, ${ }^{20}$ Emmeline W. Hill, ${ }^{14}$ Anna Jeziorowska, ${ }^{21}$ Luba Kalaydjieva, ${ }^{22}$ Manfred Kayser, ${ }^{23, *}$ Toomas Kivisild, ${ }^{3}$ Sergey A. Kravchenko, ${ }^{24}$ Astrida Krumina, ${ }^{25}$ Vaidutis Kučinskas, ${ }^{26}$ João Lavinha, ${ }^{16}$ Ludmila A. Livshits, ${ }^{24}$ Patrizia Malaspina, ${ }^{27}$ Syrrou Maria, ${ }^{28}$ Ken McElreavey, ${ }^{29}$ Thomas A. Meitinger, ${ }^{30}$ Aavo-Valdur Mikelsaar, ${ }^{4}$ R. John Mitchell, ${ }^{31}$ Khedoudja Nafa, ${ }^{32}$ Jayne Nicholson, ${ }^{3}$ Søren Nørby, ${ }^{33}$ Arpita Pandya, ${ }^{2}$ Jüri Parik, ${ }^{5}$ Philippos C. Patsalis, ${ }^{28}$ Luísa Pereira, ${ }^{7}$ Borut Peterlin, ${ }^{34}$ Gerli Pielberg, ${ }^{5}$ Maria João Prata, Carlo Previderé, ${ }^{35}$ Lutz Roewer, ${ }^{23}$ Siiri Rootsi, ${ }^{5}$ D. C. Rubinsztein, ${ }^{36}$ Juliette Saillard, ${ }^{33}$ Fabrício R. Santos, ${ }^{2,5}$ Gheorghe Stefanescu, ${ }^{37}$ Bryan C. Sykes, ${ }^{32}$ Aslihan Tolun, ${ }^{38}$ Richard Villems, ${ }^{5}$ Chris Tyler-Smith, ${ }^{2}$ and Mark A. Jobling ${ }^{1}$


#### Abstract

Clinal patterns of autosomal genetic diversity within Europe have been interpreted in previous studies in terms of a Neolithic demic diffusion model for the spread of agriculture; in contrast, studies using mtDNA have traced many founding lineages to the Paleolithic and have not shown strongly clinal variation. We have used 11 human Ychromosomal biallelic polymorphisms, defining 10 haplogroups, to analyze a sample of $3,616 \mathrm{Y}$ chromosomes belonging to 47 European and circum-European populations. Patterns of geographic differentiation are highly nonrandom, and, when they are assessed using spatial autocorrelation analysis, they show significant clines for five of six haplogroups analyzed. Clines for two haplogroups, representing $45 \%$ of the chromosomes, are continentwide and consistent with the demic diffusion hypothesis. Clines for three other haplogroups each have different foci and are more regionally restricted and are likely to reflect distinct population movements, including one from north of the Black Sea. Principal-components analysis suggests that populations are related primarily on the basis of geography, rather than on the basis of linguistic affinity. This is confirmed in Mantel tests, which show a strong and highly significant partial correlation between genetics and geography but a low, nonsignificant partial correlation between genetics and language. Genetic-barrier analysis also indicates the primacy of geography in the shaping of patterns of variation. These patterns retain a strong signal of expansion from the Near East but also suggest that the demographic history of Europe has been complex and influenced by other major population movements, as well as by linguistic and geographic heterogeneities and the effects of drift.


## Introduction

The earliest accepted date for the occupation of Europe by anatomically modern humans is $\sim 40,000$ years before the present (YBP) (Boyd and Silk 1997). Population size during the Paleolithic was probably stable and small, limited by the resources available from a hunting-gathering economy (Landers 1992). The development of ag-

Received July 10, 2000; accepted for publication September 25, 2000; electronically published November 9, 2000.

Address for correspondence and reprints: Dr. Mark A. Jobling, Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, United Kingdom. Email: maj4@leicester.ac.uk
© 2000 by The American Society of Human Genetics. All rights reserved. 0002-9297/2000/6706-0018\$02.00
riculture (the Neolithic transition) was important, because the abundance of food supplies allowed populations to expand (Hassan 1973).
The origins of agriculture have become the focus of

[^0]attempts to interpret the genetic landscape of modern Europe. The fact that agriculture arose in the Near East $\sim 10,000$ YBP (evinced by the dating of archaeological sites) is not disputed; the argument has arisen over the mechanism of its subsequent dispersal. In the demic diffusion model (Ammerman and Cavalli-Sforza 1984), the spread is thought to be due to a movement of people and would therefore have substantially changed the genetic composition of European populations; the contrasting, cultural diffusion model (Dennell 1983; Zvelebil and Zvelebil 1988) holds that the ideas and technologies were transferred without substantial population movement and thus suggests that current patterns of genetic diversity should have their roots in the Paleolithic.

These opposing hypotheses are undoubtedly overly
ersidad Complutense, Madrid; ${ }^{11}$ Dipartimento di Biologia, Universita di Ferrara, Ferrara, Italy; ${ }^{12}$ Umeå University, Department of Medical Genetics, Umeå, Sweden; ${ }^{13}$ Unitat de Biologia Evolutiva, Facultat de Ciecies de la Salut I de la Vida, Universitat Pompeu Fabra, Barcelona; ${ }^{14}$ Department of Genetics, Trinity College, Dublin; ${ }^{15}$ University of Oslo, Centre for Biotechnology, Oslo; ${ }^{16}$ Instituto Nacional de Saúde Dr. Ricardo Jorge, Lisbon; ${ }^{17}$ Forensic Laboratory for DNA Research, MGC-Department of Human and Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands; ${ }^{18}$ Laboratory for Forensic Genetics and Molecular Archaeology, Center for Human Genetics, K.U. Leuven, Leuven, Belgium; ${ }^{19}$ Research Centre for Medical Genetics, Russian Academy of Medical Sciences, UFA Science Centre, Department of Biochemistry and Cytochemistry, Moscow; ${ }^{20}$ Department of Physiology, University of Kiel, Kiel; ${ }^{21}$ Department of Medical Genetics, Institute of Endocrinology, Medical University of Lódz, Lódz, Poland; ${ }^{22}$ Department of Human Biology, Edith Cowan University, Joondalup Campus, and Western Australian Institute for Medical Research, Royal Perth Hospital, Perth; ${ }^{23}$ Genetic Research Laboratory, Institute of Legal Medicine, Medical Faculty (Charité), HumboldtUniversity Berlin, Berlin; ${ }^{24}$ Institute of Molecular Biology and Genetics, National Academy of Science of Ukraine, Kiev; ${ }^{25}$ Department of Medical Biology and Genetics, Medical Academy of Latvia, Riga; ${ }^{26}$ Center of Human Genetics, University of Vilnius, Vilnius, Lithuania; ${ }^{27}$ Department of Biology, University of Rome "Tor Vergata," Rome; ${ }^{28}$ The Cyprus Institute of Neurology and Genetics, Nicosia; ${ }^{29}$ Unité d'Immunogénétique Humaine, Institut Pasteur, Paris; ${ }^{30}$ Department of Medical Genetics, Kinderpoliklinik, Munich; ${ }^{31}$ La Trobe University, School of Genetics and Human Variation, Bundoora, Australia ${ }^{32}$ Department of Human Genetics, Memorial Sloan-Kettering Cancer Center, New York; ${ }^{33}$ Laboratory of Biological Anthropology, Institute of Forensic Medicine, University of Copenhagen, Copenhagen; ${ }^{34}$ Division of Medical Genetics, Department of Obstetrics and Gynaecology, Ljubljana, Slovenia; ${ }^{15}$ Dipartimento di Medicina Legale e Sanita Pubblica, Pavia, Italy; ${ }^{36}$ I.C. Biologice, Iasi, Romania; and ${ }^{37}$ Bogazici University, Department of Molecular Biology and Genetics, Istanbul

Present affiliation: McDonald Institute for Archaeological Research, University of Cambridge, Cambridge.
${ }^{\dagger}$ Present affiliation: Department of Genetics, University of Leicester, Leicester, United Kingdom.
$\ddagger$ Present affiliation: Max Planck Institute for Evolutionary Anthropology, Department of Evolutionary Genetics, Leipzig.
' Present affiliation: Departamento de Biologia Geral, Instituto Ciências Biológicas/Universidade Federal de Minas Gerais, Minas Gerais, Brazil.
simplistic but have been widely adopted as models in genetic studies (Sokal et al. 1991; Cavalli-Sforza et al. 1993; Barbujani et al. 1994; Piazza et al. 1995; Semino et al. 1996; Chikhi et al. 1998a, 1998b; Richards and Sykes 1998; Simoni et al. 2000a), since they predict patterns of diversity that should be easily recognizablein particular, demic diffusion is expected to result in clines with foci in the Near East. Principal components (PC) analysis of classical gene-frequency data reveals clines within Europe, and the first principal component, which indeed has a Near Eastern focus, has been taken to support the demic diffusion hypothesis (Menozzi et al. 1978; Cavalli-Sforza et al. 1993). A similar pattern has been observed in spatial autocorrelation analysis of DNA-based polymorphisms, including microsatellites, which have identified geographic patterns compatible with a substantial directional demographic expansion affecting much of the continent (Chikhi et al. 1998a). However, although these patterns in the genetic data are impressive and suggest major east-west population movements, their time depths are not known, and associating them with particular demographic events is usually speculative. They could be just as well due to the original peopling of Europe during the Upper Paleolithic as to the Neolithic transition. In this regard, some support for the latter does come from the finding of significant partial correlations between classical marker frequencies and the relative dates for the origin of agriculture in different locations (Sokal et al. 1991).

By contrast, analysis of diversity in European mtDNA reveals a relatively homogeneous landscape (Comas et al. 1997), with clines detectable only in the south (Simoni et al. 2000a). However, this is a contentious area, and conclusions may depend on the depth of analysisfor example, which sublineages are studied. An eastwest gradient of pairwise differences has been discerned and claimed to be compatible with expansion from the Middle East (Comas et al. 1997). However, attempts to identify and date founding lineages (Richards et al. 1996) have suggested that Paleolithic lineages may persist in Europe to a degree that is inconsistent with the demic diffusion hypothesis, although an ancient origin of certain alleles or haplogroups (HGs) is certainly compatible with a later spread of those alleles within a geographic region (Langaney et al. 1992; Templeton 1993).

Language can provide additional evidence about past demography (Renfrew 1989), although direct information about past languages on the basis of writing is limited to the past 5,000 years, and inferences before that time are controversial (Renfrew 2000). Europe is remarkable for its linguistic homogeneity, languages of the Indo-European (IE) family being spoken by most populations from India to Ireland (Renfrew 1989). In one persuasive view, demic diffusion from the Near East provides a common explanation for the spread of both
agriculture and IE languages (Renfrew 1987). Other ideas have been put forward, however; one, which has been adopted by some geneticists because of its apparent compatibility with the pattern seen in the third principal component of variation of classical gene frequencies (Cavalli-Sforza et al. 1994), is that the IE language was spread by the movement, from north of the Caspian Sea, of the Kurgan people, pastoral nomads who domesticated the horse (Gimbutas 1970). An alternative view has it that the spread of IE language preceded the origins of agriculture and was due to the reexpansion of hunter-gatherers after the end of the Last Glacial Maximum (Adams and Otte 2000).
Despite the hegemony of IE languages, there is diversity within them, and some members of other language families also exist; one example, Basque, clearly represents a survival from an earlier era. Various methods for the detection of genetic barriers in autosomal gene frequencies within Europe (Barbujani 1991) show that most of these barriers correlate with linguistic boundaries, and it may be that language and geographic proximity are equally good predictors of genetic affinity (Barbujani 1997). However, some examples of non-IE languages reflect not persistence but recent acquisition through "elite dominance": for example, the Hungarians acquired their Uralic language from the invading Magyars only $\sim 1,100$ YBP (Cavalli-Sforza et al. 1994), and the Altaic language of the Turks was acquired as a result of the Turkic invasions during the 11th-15th centuries (Renfrew 1989). This process of language acquisition by elite dominance is not expected to be accompanied by a high degree of genetic admixture, and, if this is so, populations such as the Hungarians and Turks are unlikely to be separated from surrounding populations by genetic barriers.

Use of the Y chromosome to investigate human population histories (Jobling and Tyler-Smith 1995) is increasing as convenient polymorphic markers become available. However, the effective population size of this chromosome is one-quarter that of any autosome, and this means that it is particularly influenced by drift. Effective population size may be further reduced through the variance in the number of sons that a father has and perhaps by selective sweeps (Jobling and TylerSmith 2000). Conclusions about populations on the basis of this single locus must therefore be made with caution. One useful property of the Y chromosome is its high degree of geographic differentiation, compared with other parts of the genome, which has been explained by drift and a greater effective migration of women than of men, through the phenomenon of patrilocality (Seielstad et al. 1998), in which women are more likely to move from their birthplace after marriage than are men. The Y chromosome may therefore be a sensitive system for detecting the population movements
that have shaped European genetic diversity; there again, it may be so susceptible to drift that ancient patterns have been obscured.
Published data on European Y-chromosome diversity are not extensive; markers have been of limited informativeness, and the distribution of population samples has often been unsatisfactory. By use of two "classical" Y-chromosome markers-the complex and highly polymorphic 49f/TaqI system (Ngo et al. 1986; Lucotte and Loirat 1999) and the biallelic marker 12 f 2 (Casanova et al. 1985)-patterns of diversity have been demonstrated that have been claimed to be clinal and to support the demic diffusion model (Semino et al. 1996). Subsequent analysis using Y-chromosome-specific microsatellites (Quintana-Murci et al. 1999) and a combination of microsatellites and two biallelic markers (Malaspina et al. 1998) showed similar east-west gradients. 49 f has been exploited more fully to analyze the correlation between Y-chromosome diversity, mtDNA diversity, and language in a global sample, and it has been suggested that the Y chromosome shows the stronger correlation with language (Poloni et al. 1997).

Recent progress in the development of Y-chromosome polymorphic markers that can be assayed by use of PCR now allows us to explore these issues in greater detail. In this study, we use 11 such markers to assay the diversity of Y-chromosomal lineages in a large sample of men from 47 populations distributed over most of Europe.

## Subjects and Methods

## Subjects

Y chromosomes from 3,616 men from 47 populations (table 1) were included in this study; the majority were classified by birthplace of the paternal grandfather. DNA samples were from collections of the authors, and informed consent was obtained. A total of 311 samples from the Baltic region are from the study by T. Zerjal, L. Beckman, G. Beckman, A.-V. Mikelsaar, A. Krumina, V. Kučinskas, M. E. Hurles, and C. Tyler Smith (unpublished data). The 257 Irish Y chromosomes included 221 chromosomes studied elsewhere (Hill et al. 2000), which were typed here with three additional markers. The 129 North African samples were those studied elsewhere by Bosch et al. (1999); chromosomes with the M9 G allele and 92R7 C allele were additionally typed with LLY22g (see below). The 172 East Anglian samples were studied elsewhere by Cooper et al. (1996).

## Biallelic Markers

A total of 11 biallelic markers were used in this study (fig. 1). These were chosen on the basis of previous work by us and by others (Santos and Tyler-Smith 1996; Sem-

Table 1
HG Frequency Data in 47 Populations

| Population (no.) | Location | Language family (Subfamity) | No. (\%) of Individuals with HG |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | 1 | 2 | 3 | 4 | 7 | 8 | 9 | 12 | 16 | 21 | 22 | 26 |
| Icelandic (28) | $64^{\circ} 1 \mathrm{~N}, 21^{\circ} 6 \mathrm{~W}$ | IE (Germanic) | 13 (46) | 9 (32) | 6 (21) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Saami (48) | $68^{\circ} \mathrm{N}, 22^{\circ} \mathrm{E}$ | Uralic (Finno-Ugric) | 3 (6) | 15 (31) | 10 (21) | 0 | 0 | 0 | 0 | 0 | 20 (42) | 0 | 0 | 0 |
| Northern Swedish (48) | $63^{\circ} 7 \mathrm{~N}, 20^{\circ} 3^{\prime} \mathrm{E}$ | IE (Germanic) | 11 (23) | 22 (48) | 9 (19) | 0 | 0 | 0 | 1 (2) | 0 | 4 (8) | 1 (2) | 0 | 0 |
| Gotlander (64) | $57^{\circ} 5 \mathrm{~N}, 18^{\circ} 5^{\prime} \mathrm{E}$ | IE (Germanic) | 11 (17) | 38 (59) | 10 (16) | 0 | 0 | 0 | , | 0 | 4 (6) | 0 | 0 | 1 (2) |
| Norwegian (52) | $59^{\circ} 9 \mathrm{~N}, 10^{\circ} 8^{\prime} \mathrm{E}$ | IE (Germanic) | 15 (29) | 17 (33) | 16 (31) | 0 | 0 | 0 | 1 (2) | 0 | 2 (4) | 1 (2) | 0 | 0 |
| Danish (56) | $55^{\circ} 7 \mathrm{~N}, 12^{\circ} 6^{\prime} \mathrm{E}$ | IE (Germanic) | 28 (50) | 18 (32) | 4 (7) | 0 | 0 | 0 | 4 (7) | 0 | 1 (2) | 1 (2) | 0 | 0 |
| Finnish (57) | $60^{\circ} 1 \mathrm{~N}, 25^{\circ} \mathrm{E}$ | Uralic (Finno-Ugric) | 1 (2) | 13 (23) | 6 (10) | 0 | 0 | 0 | 0 | 1 (2) | 35 (61) | 1 (2) | 0 | 0 |
| Estonian (207) | $59^{\circ} 4 \mathrm{~N}, 24^{\circ} 7^{\prime} \mathrm{E}$ | Uralic (Finno-Ugric) | 18 (9) | 30 (14) | 56 (27) | 0 | 0 | 0 | 2 (1) | 8 (4) | 76 (37) | 6 (3) | 0 | 11 (5) |
| Latvian (34) | $56^{\circ} 9 \mathrm{~N}, 24^{\circ} 1^{\prime} \mathrm{E}$ | IE (Balto-Slavic) | 5 (15) | 4 (12) | 14 (41) | 0 | 0 | 0 | 0 | 0 | 11 (32) | 0 | 0 | 0 |
| Lithuanian (38) | $54^{\circ} 7 \mathrm{~N}, 25^{\circ} 3^{\prime} \mathrm{E}$ | IE (Balto-Slavic) | 2 (5) | 5 (13) | 13 (34) | 0 | 0 | 0 | 0 | 0 | 18 (47) | 0 | 0 | 0 |
| Russian (122) | $55^{\circ} 8 \mathrm{~N}, 37^{\circ} 7^{\prime} \mathrm{E}$ | IE (Balto-Slavic) | 8 (7) | 21 (17) | 57 (47) | 0 | 0 | 0 | 5 (4) | 5 (4) | 17 (14) | 8 (7) | 0 | 1 (1) |
| Belarusian (41) | $53^{\circ} 9 \mathrm{~N}, 27^{\circ} 5^{\prime} \mathrm{E}$ | IE (Balto-Slavic) | 4 (10) | 14 (34) | 16 (39) | 0 | 0 | 0 | 1 (2) | 0 | 1 (2) | 4 (10) | 0 | 1 (2) |
| Ukrainian (27) | $50^{\circ} 4 \mathrm{~N}, 30^{\circ} 5^{\prime} \mathrm{E}$ | IE (Balto-Slavic) | 1 (4) | 13 (48) | 8 (30) | 0 | 0 | 0 | 0 | 8 | 4 (11) | 1 (4) | 0 | 0 |
| Mari (48) | $56^{\circ} 5^{\prime} \mathrm{N}, 48^{\circ} \mathrm{E}$ | Uralic (Finno-Ugric) | 5 (10) | 2 (4) | 14 (29) | 0 | 0 | 0 | 3 (6) | 8 (17) | 16 (33) | 0 | 0 | 0 |
| Chuvash (17) | $55^{\circ} 5 \mathrm{~N}, 47^{\circ} \mathrm{E}$ | Altaic (Turkic) | 2 (12) | 4 (24) | 3 (18) | 0 | 0 | 0 | 1 (6) | 0 | 3 (18) | 1 (6) | 0 | 3 (18) |
| Georgian (64) | $41^{\circ} 5 \mathrm{~N}, 44^{\circ} 5^{\prime} \mathrm{E}$ | Caucasian (Southern Caucasian) | 12 (19) | 31 (48) | 4 (6) | 0 | 0 | 0 | 15 (23) | 0 | - | 1 (2) | 0 | 1 (2) |
| Ossetian (47) | $43^{\circ} 1 \mathrm{~N}, 44^{\circ} 5^{\prime} \mathrm{E}$ | IE (Indo-Iranian) | 20 (43) | 5 (11) | 1 (2) | 0 | 0 | 0 | 16 (34) | 0 | 0 | 3 (6) | 0 | 2 (4) |
| Armenian (89) | $40^{\circ} 2 \mathrm{~N}, 44^{\circ} 5^{\prime} \mathrm{E}$ | IE (Armenian) | 22 (25) | 28 (31) | 5 (6) | 0 | 0 | 0 | 26 (29) | 0 | 3 (3) | 3 (3) | 0 | 2 (2) |
| Turkish (167) | $41^{\circ} \mathrm{N}, 29^{\circ} \mathrm{E}$ | Altaic (Turkic) | 34 (20) | 41 (25) | 8 (5) | 0 | 0 | 0 | 55 (33) | 2 (1) | 2 (1) | 17 (10) | 0 | 8 (5) |
| Cypriot (45) | $35^{\circ} 3 \mathrm{~N}, 33^{\circ} 4^{4} \mathrm{E}$ | IE (Greek) | 4 (9) | 10 (22) | 1 (2) | 0 | 0 | 0 | 15 (33) | 1 (2) | 0 | 12 (27) | 0 | 2 (4) |
| Greek (36) | $38^{\circ} \mathrm{N}, 23^{\circ} 7^{\prime \prime} \mathrm{E}$ | IE (Greek) | 4 (11) | 8 (22) | 3 (8) | 0 | 0 | 0 | 10 (28) | 0 | 0 | 10 (28) | 0 | 1 (3) |
| Bulgarian (24) | $42^{\circ} 7 \mathrm{~N}, 23^{\circ} 3^{\prime} \mathrm{E}$ | IE (Balto-Slavic) | 4 (17) | 10 (42) | 3 (12) | 0 | 0 | 0 | 3 (12) | 0 | 0 | 4 (17) | 0 | 0 |
| Czech (53) | $50^{\circ} 2 \mathrm{~N}, 14^{\circ} 5^{\prime} \mathrm{E}$ | IE (Balto-Slavic) | 10 (19) | 10 (19) | 20 (38) | 0 | 0 | 0 | 6 (11) | 3 (6) | 0 | 4 (8) | 0 | 0 |
| Slovakian (70) | $48^{\circ} 1 \mathrm{~N}, 17^{\circ} 1 \mathrm{E}$ | IE (Balto-Slavic) | 12 (17) | 12 (17) | 33 (47) | 0 | 0 | 0 | 2 (3) | 1 (1) | 2 (3) | 7 (10) | 0 | 1 (1) |
| Romanian (45) | $44^{\circ} 4 \mathrm{~N}, 26^{\circ} 1{ }^{1} \mathrm{E}$ | IE (Italic) | 8 (18) | 12 (27) | 9 (20) | 0 | 0 | 0 | 11 (24) |  | , | 3 (7) | 1 (2) | 1 (2) |
| Yugoslavian (100) | $44^{\circ} 8 \mathrm{~N}, 20^{\circ}{ }^{\circ} \mathrm{E}$ E | IE (Balto-Slavic) | 11 (11) | 49 (49) | 16 (16) | 0 | 0 | 0 | 8 (8) | 2 (2) | 0 | 13 (13) | , | 1 (1) |
| Slovenian (70) | $46^{\circ} 1 \mathrm{~N}, 14^{\circ} 5^{\prime} \mathrm{E}$ | IE (Balto-Slavic) | 15 (21) | 19 (27) | 26 (37) | 0 | 0 | 0 | 4 (6) | 0 | 0 | 5 (7) | 1 (1) | 0 |
| Hungarian (36) | $47^{\circ} 5 \mathrm{~N}, 19^{\circ} 1^{\prime} \mathrm{E}$ | Uralic (Finno-Ugric) | 11 (30) | 10 (28) | 8 (22) | 0 | 0 | 0 | 1 (3) | 0 | 0 | 6 (17) | 0 | 0 |
| Polish (112) | $51^{\circ} 7 \mathrm{~N}, 19^{\circ} 5 \mathrm{E}$ | IE (Balto-Slavic) | 20 (18) | 19 (17) | 61 (54) | 0 | 0 | 0 | 4 (4) | 1 (1) | 5 (4) | 2 (2) |  | 0 |
| Italian (99) | $41^{\circ} 9 \mathrm{~N}, 12^{\circ} 5^{\prime} \mathrm{E}$ | IE (Italic) | 44 (44) | 14 (14) | 2 (2) | 0 | 0 | 0 | 20 (20) | 0 | 0 | 13 (13) |  | 6 (6) |
| Sardinian (10) | $39^{\circ} 2 \mathrm{~N}, 9^{\circ} 1{ }^{\text {ce }}$ | IE (Italic) | 3 (30) | 4 (40) | 0 | 0 | 0 | 1 (10) | 0 | 0 | 0 | 2 (20) | 0 | 0 |
| Bavarian (80) | $48^{\circ} 1 \mathrm{~N}, 11^{\circ} 6^{\prime} \mathrm{E}$ | IE (Germanic) | 38 (48) | 18 (23) | 12 (15) | 0 | 0 | 0 | 4 (5) | 0 | 0 | 6 (8) | 2 (3) | 0 |
| German (30) | $52^{\circ} 5 \mathrm{~N}, 13^{\circ} 4 \mathrm{E}$ | IE (Germanic) | 12 (40) | 6 (20) | 9 (30) | 0 | 0 | 0 | 1 (3) | 0 | 1 (3) | 0 | 0 | 1 (3) |
| Dutch (84) | $52^{\circ} 3 \mathrm{~N}, 4^{\circ 9} 9^{\circ} \mathrm{E}$ | IE (Germanic) | 36 (43) | 27 (32) | 11 (13) | 0 | 0 |  | 6 (7) | 0 | 0 | 3 (8) | 1 (1) | 0 |
| French (40) | $48^{\circ} 9 \mathrm{~N}, 2^{\circ} 3^{\prime} \mathrm{E}$ | IE (Italic) | 20 (50) | 10 (25) | 2 (5) | 0 | 0 | 1 (3) | 2 (5) | 0 | 0 | 3 (8) | 2 (5) | 0 |
| Belgian (92) | $50^{\circ} 8 \mathrm{~N}, 4^{\circ} 3^{\prime} \mathrm{E}$ | IE (Germanic) | 58 (63) | 21 (23) | 4 (4) | 0 | 0 | 0 | 5 (5) | 0 | 0 | 2 (2) | 1 (1) | 1 (1) |
| Western Scortish (120) | $57^{\circ} 2 \mathrm{~N}, 6^{\circ} 2 \mathrm{~W}$ | IE (Celtic) | 87 (72) | 23 (19) | 8 (7) | 0 | 0 | 0 | 0 | 0 | 0 | 2 (2) | 0 | 0 |
| Scottish (43) | $56^{\circ} \mathrm{N}, 3^{\circ} 2 \mathrm{~W}$ | IE (Celtic) | 34 (79) | 5 (12) | 3 (7) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 (2) |
| Cornish (51) | $50^{\circ} 3 \mathrm{~N}, 4^{\circ} 4 \mathrm{~W}$ | IE (Celtic) | 42 (82) | 9 (18) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| East Anglian (172) | $52^{\circ} 6 \mathrm{~N}, 1^{\circ} 3^{3} \mathrm{E}$ | IE (Germanic) | 97 (56) | 52 (30) | 15 (9) | 0 | 0 | 0 | 1 (1) | 0 | 0 | 5 (3) | 1 (1) | 1 (1) |
| Irish (257) | $53^{\circ} 3 \mathrm{~N}, 6^{\circ} 3^{3} \mathrm{~W}$ | IE (Celtic) | 207 (81) | 39 (15) | 2 (1) | 0 | 0 | 0 | 2 (1) | 0 | 1 (.5) | 6 (2) | 0 | 0 |
| Basque (26) | $43^{\circ} 3 \mathrm{~N}, 2^{\circ 9}$ W | Basque (Basque) | 19 (73) | 2 (8) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 (19) | 0 |
| Spanish (126) | $40^{\circ} 4 \mathrm{~N}, 3^{\circ} 7 \mathrm{~W}$ | IE (Italic) | 86 (68) | 17 (13) | 3 (2) | 0 | 0 | 0 | 4 (3) | 0 | 0 | 12 (10) | 3 (2) | 1 (1) |
| Southern Portuguese (57) | $38^{\circ} 7 \mathrm{~N}, 9^{\circ} 1^{1} \mathrm{~W}$ | IE (Italic) | 32 (56) | 8 (14) | 1 (2) | 0 | 0 | 0 | 5 (9) | 0 | 0 | 10 (17) | 0 | 1 (2) |
| Northern Portuguese (328) | $41^{\circ} 2 \mathrm{~N}, 8^{\circ} 6^{\prime} \mathrm{W}$ | IE (Italic) | 203 (62) | 54 (16) | 0 | 0 | 0 | 0 | 21 (6) | 0 | 0 | 35 (11) | 6 (2) | 9 (3) |
| Algerian (27) | $36^{\circ} 5 \mathrm{~N}, 3^{\circ} \mathrm{E}$ | Afro-Asiatic (Semitic) | 0 | 1 (4) | 0 | 0 | 0 | 1 (4) | 11 (41) | 0 | 0 | 14 (52) | 0 | 0 |
| Northern African (129) | $35^{\circ} 5 \mathrm{~N}, 5^{\circ} 7 \mathrm{~W}$ | Afro-Asiatic (Berber and Semitic) | $5(4)$ | 4 (3) |  | $\frac{0}{0}$ | 0 |  | $\frac{15}{29(12)}$ |  |  | $\frac{99(77)}{3269}$ |  |  |
| Total ( 3,616 ) |  |  | $\overline{1,337(37)}$ | 803(22) | $512(14)$ | $\overline{0}$ | $\overline{0}$ | $9(0.3)$ | 291 (8) | 32 (1) | $22616)$ | 326 (9) | $23(0.7)$ | 57(2) |



Figure 1 Maximum-parsimony network of Y-chromosomal biallelic HGs. Circles and squares represent compound haplotypes, or HGs; numbers within them are their arbitrarily assigned names; and arrows or lines between them represent the defining biallelic mutations. The order of occurrence of the 92R7 and DYS257 mutations is not known, because the intermediate HG has not been found; arrows for these polymorphisms are shown adjacent to each other. Where ancestral state is known, arrows point to the derived state. HGs analyzed in this study are indicated by circles; arrows or boxes between them give the nature of the mutation ( 0 , ancestral; 1 , derived), and, where appropriate, the restriction enzyme used and the allele cleaved in PCR-RFLP analysis. For HGs not analyzed (squares), information on geographic association is provided by shading. The correspondence of some of these HGs with the haplotype nomenclature of Karafet et al. (1999) and Hammer et al. (2000), whose work is referred to in the text, is as follows: HGs $1+22$, haplotype 1C; HG 3, haplotype 1D; HG 4, haplotype 3G; HG 7, haplotypes $1 A+2$; HG 8, haplotype 5 ; HGs $12+26$, haplotype $1 \mathrm{U} ; \mathrm{HG} 16$, haplotype 1I; HG 21, haplotypes $3 \mathrm{~A}+4$; and HG 9, haplotype "Med."
ino et al. 1996; Underhill et al. 1997; Zerjal et al. 1997; Hammer et al. 1998; Hurles et al. 1999), indicating that the HGs that they define are likely to be found within European populations. There are several nomenclature systems currently in use for Y-chromosomal lineages, and, since we refer to the data of Karafet et al. (1999) and Hammer et al. (2000) in the text, we give some correspondences in the legend to figure 1 . HG 7 is specific to sub-Saharan African populations (Karafet et al. 1999) but is typed here by default, since it is defined by the ancestral state of the recurrent SRY-1532 polymorphism (fig. 1). Maximum-parsimony analysis of haplotypes defined by these markers generates a unique tree (figs. 1 and 2) in which DYS257 (Hammer et al. 1998) and 92R7 (Mathias et al. 1994) are phylogenetically equivalent (Jobling et al. 1998; Z. H. Rosser, M. E. Hurles, and M. A. Jobling, unpublished data). For this part
of the phylogeny, 92 R 7 was typed routinely, and DYS257 was typed when necessary to confirm results. Nine of the markers have been described elsewhere: YAP (Hammer 1994) was typed according to the method of Hammer and Horai (1995), SRY-1532 (Whitfield et al. 1995) according to Kwok et al. (1996), SRY-2627 according to Veitia et al. (1997), 92R7 (Mathias et al. 1994) according to Hurles et al. (1999), DYS257 according to Hammer et al. (1998), M9 (Underhill et al. 1997) according to Hurles et al. (1998), sY81 according to Seielstad et al. (1994), Tat according to Zerjal et al. (1997), and SRY-8299 (Whitfield et al. 1995) according to Santos et al. (1999).
12 f 2 (Casanova et al. 1985) was typed using a newly developed PCR assay. This polymorphism was originally suggested to be an $\sim 2-\mathrm{kb}$ insertion/deletion, but our analysis suggests that its molecular basis is more com-

$n=3616$

Figure 2 HG profile of the entire sample set. HG diversity within the complete sample set of $3,616 \mathrm{Y}$ chromosomes, summarized on a simplified version of the network shown in figure 1. The area of each black circle is proportional to the frequency of the HG. Small unblackened circles indicate unobserved HGs (4 and 7). The position of the HG closest to the root (HG7) is indicated.
plex than this. The PCR assay generates a 500 -bp product from chromosomes carrying the Taq $\mathrm{I} / 10-\mathrm{kb}$ allele, but this product is absent from TaqI/8-kb-allele chromosomes (HG 9). An 820-bp amplicon from the SRY region, present in all chromosomes, is amplified as a control. Analysis of the 12 f 2 region gives no information about ancestral state, but we assume that presence of the $500-\mathrm{bp}$ amplicon is ancestral. Primer sequences for the 12 f 2 amplicon are 12 f 2 D ( $5^{\prime}$-CTG ACT GAT CAA AAT GCT TAC AGA TC- $3^{\prime}$ ) and 12 f 2 F ( $5^{\prime}$-TCT TCT AGA ATT TCT TCA CAG AAT TG-3'), and those for the $S R Y$ control amplicon are $3^{\prime} S R Y 15$ ( $5^{\prime}$-CTT GAT TTT CTG CTA GAA CAA G- $3^{\prime}$ ) and $3^{\prime}$ SRY16 ( $5^{\prime}$-TGT CGT TAC ATA AAT GGG CAC- $3^{\prime}$ ). PCR conditions were $33-35$ cycles of $94^{\circ} \mathrm{C}$ for $30 \mathrm{~s}, 59^{\circ} \mathrm{C}$ for 30 s , and $72^{\circ} \mathrm{C}$ for 45 s . An alternative assay, generating shorter amplicons, was used with degraded DNAs. The primers 12 f 2 D (see above) and 12 f 2 G ( $5^{\prime}$-GGA TCC CTT CCT TAC ACC TTA TAC-3') produce an 88 -bp product from TaqI/10-kb-allele chromosomes (and no product from TaqI/8-kb-allele chromosomes), which is coamplified with the Tat 112-bp amplicon (Zerjal et al. 1997) as a control, under the following conditions: 33-36 cycles of $94^{\circ} \mathrm{C}$ for $30 \mathrm{~s}, 59^{\circ} \mathrm{C}$ for 30 s , and $72^{\circ} \mathrm{C}$ for 30 s . All chromosomes known, from previous hybridization analysis, to carry TaqI/8-kb alleles lacked the 12 f 2 test amplicons in both of these assays. However, some YAP + chromosomes belonging to HG 4 also lack the $12 \mathrm{f} 2 \mathrm{am}-$ plicons, suggesting that the polymorphism may be recurrent (Blanco et al. 2000).
The LLY22g HindIII polymorphism was typed by a PCR-RFLP assay that will be described elsewhere (E.

Righetti and C. Tyler-Smith, unpublished data). The deep-rooting markers $S R Y-1532$, M9, YAP, and 92R7 were typed on all samples. For many samples, all other markers were also typed. However, in some cases, remaining markers were typed hierarchically-for instance, SRY-8299 and sY81 were, in some cases, typed only on chromosomes classified as YAP+.

## Experimental Procedures

Haplotyping was carried out in Leicester; Oxford (both laboratories); Barcelona; Belgrade; Dublin; Leuven, Belgium; Lisbon; Porto, Portugal; Rome; and Tartu, Estonia. Procedures were based on those described by Hurles et al. (1998). To verify typing methodologies, a set of 12 quality-control DNAs was satisfactorily typed blindly by all participating laboratories.

## Statistical Analysis

Spatial autocorrelation analysis was done by AIDA (Bertorelle and Barbujani 1995), for the entire data set, and SAAP (Sokal and Oden 1978), for individual HGs. PC analysis of covariances was carried out according to the method of Harpending and Jenkins (1973).
Mantel (1967) tests, done by ARLEQUIN version 2.0 (Schneider et al. 2000), were used to determine whether language or geography has the stronger impact on genetic differentiation. Genetic distances (as a pairwise $F_{\text {ST }}$ matrix) were computed within ARLEQUIN, and geographic distances were calculated from latitude and longitude by use of great-circle distances, in a program written in Interactive Data Language 5.1 (Research Systems Inc.) by M. E. Hurles. Within IE languages, linguistic distances were adapted from Dyen et al. (1992), who used the lexicostatistical method of Swadesh (1952) on comparisons of 200 -word lists: percentage similarities were first converted to dissimilarities, and these numbers then assigned as nonpercentage distances between languages (ranging from 9 [Czech to Slovak] to 88 [Armenian to Irish]). All IE languages within the data set were represented, with the exception of Scottish, which was assigned a distance of 10 from Irish; we also tested the effects of other values, in the range 5-20. The Belgian sample was divided into its two linguistic groups-those speaking French ( 56 individuals) and those speaking Dutch (36). An arbitrary and conservative, larger value, 200, was then assigned as a distance between language families. As was done by Poloni et al. (1997), Mantel tests were also performed using different inter-languagefamily distances, of 400 and 1,000 . Two of the non-IE language families, Altaic and Uralic, are represented by more than one language within our data set. On the basis of a consideration of the classification by Ruhlen (1991) and of the inter-IE-language distances of Dyen et al. (1992), plausible distances were assigned within these families, and the effect of altering these values over
a range was tested. Within Uralic, values were as follows: Finnish to Estonian, 25 (altered value range $10-30$ ); Finnish-Estonian to Saami, 30 (20-40); Finnish-Esto-nian-Saami to Mari, 40 (30-70); and Hungarian to all other Uralic languages, 80 (40-90). Values for Chuvash and Turkish (Altaic) were 40 (20-60).
To locate zones of abrupt genetic change, or genetic boundaries, and to assess their significance, we used the program ORINOCO, written in Interactive Data Language 5.1 (Research Systems) by M.E. Hurles (Hurles 1999), which adapts a method known as "wombling" (Barbujani et al. 1989), initially developed for the analysis of allele frequencies. First, an inverse-distancesquared weighted algorithm was used to interpolate the frequencies for each of the eight observed HGs at each grid point within a $100 \times 100$ array (with account taken of the curvature of the earth and with correspondence to a grid point every $0.36^{\circ}$ latitude and $0.72^{\circ}$ longitude). The derivatives of these eight interpolated surfaces were then calculated at every node of the grid, and the magnitudes of the derivatives were summed, thus giving a measurement of the slope of the combined surfacesthat is, the overall rate of Y -chromosomal genetic change in 10,000 rectangles covering Europe. The significance of these gradients was considered in two ways, both of which take into account isolation by distance within the landscape (Barbujani et al. 1989). First, a simple significance threshold was applied, with only the top $5 \%$ of values. Second, a Monte-Carlo algorithm was used to permute the HG data 1,000 times, and summed derivatives were calculated for each permutation. This algorithm maintains the observed sample sizes and positions and therefore controls for the conflated effects, in the generation of false positives, of sampling and heterogeneity in distances between sample sites. Grid points obtained with the original HG data were then retained only if the values of their summed derivatives were $>95 \%$ of the values obtained from the permuted data. Grid points could then be plotted on a map, color coded to indicate the strength of the barrier, to show the positions of significant barriers, and were also displayed on Delaunay triangulation connections (Brassel and Reif 1979) between sample sites. The Algerian and northernAfrican samples were excluded from the barrier analysis, since their high degree of difference from all other samples (as shown in PC analysis) represents a strong genetic barrier that would bias the detection of barriers elsewhere.

## Results

Y chromosomes from 3,487 males belonging to 47 populations (fig. 3A) were haplotyped using biallelic markers and were classified into HGs (table 1); data on 129 northern-African Y chromosomes (Bosch et al. 1999)
were also included (see the Subjects and Methods section), giving a total of 3,616 . The resulting frequency data for the entire sample are summarized in figure 2. Two HGs, 7 and 4, are absent, which is consistent with published information: HG 7 has been discussed above (see the Subjects and Methods section), and HG 4 is restricted to eastern and central Asia (Karafet et al. 1999).

No single population has a frequency distribution resembling that of the overall sample (fig. 2), emphasizing the strong geographic differentiation of Y-chromosomal variation in Europe. This is evident in the HG frequency data in figure 3: distributions of HGs are highly nonrandom, with, for example, a concentration of HG 1 chromosomes in the west, HG 9 chromosomes in the southeast, HG 16 chromosomes in the northeast, and HG 3 chromosomes in central and eastern Europe.

## Clinal Distribution of $Y$-Chromosomal Lineages

To examine the geographic differentiation of these HGs more quantitatively, we used spatial autocorrelation analyses (Sokal and Oden 1978). These methods give a measure of the average level of genetic similarity, between populations within particular geographic distance classes, that can be represented as correlograms (fig. 4), and they allow clinal variation, reflecting population movement or natural selection, to be distinguished both from isolation by distance, reflecting shortrange dispersals and drift, and from nonsignificance. We first used AIDA (Bertorelle and Barbujani 1995), which takes into account molecular distances between HGs and provides autocorrelation indices (Moran's II) for the entire data set, including the rare HGs. The pattern (fig. $4 A$ ) is strongly clinal, recognized as a change from positive to negative autocorrelation indices with increasing distance class. The SAAP analysis (fig. $4 B-G$ ), omitting low-frequency HGs (HGs 8, 12, 22, and 26), confirms this clinal pattern and reveals information about individual lineages. The distributions of all of the HGs examined, with the exception of HG 2 , are strongly clinal (fig. 4), confirming the visual impression given by figure 3. In two cases (HGs 3 and 16), values become positive or zero in the longest-distance class (a "depression"), indicating the regional-rather than continentwideinfluence of these clines.

HG 2 is the most ancestral lineage that we find within Europe, and it lies at a starlike node within the tree; chromosomes within this HG are essentially undefined and are likely to consist of a set of discrete sublineages that themselves probably have greater geographic coherence. Consistent with this, HG 2 chromosomes are widely distributed across the whole landscape and constitute the only high-frequency lineage that does not show clinal variation (figs. $3 B$ and 4C). Because of this


Figure 3 Distribution of populations sampled and geographic distribution of Y-chromosomal HG diversity. A, Abbreviated population names. alg = Algerian; arm = Armenian; bas = Basque; bav = Bavarian; bgm = Belgian; brs = Belarusian; bul = Bulgarian; chu = Chuvash; cyp $=$ Cypriot; cze $=$ Czech; dk $=$ Danish; dut $=$ Dutch; ene $=$ East Anglian; enw $=$ Cornish; est $=$ Estonian; fin $=$ Finnish; fra $=$ French; geo $=$ Georgian; ger $=$ German; gk $=$ Greek; got $=$ Gotlander; hun $=$ Hungarian; ice $=$ Icelandic; irl $=$ Irish; ita $=$ Italian; lat $=$ Latvian; lit = Lithuanian; mar $=$ Mari; naf $=$ northern African; nor $=$ Norwegian, oss $=$ Ossetian; pol $=$ Polish; pon = northern Portuguese; pos $=$ southern Portuguese; rom $=$ Romanian; rus $=$ Russian; saa $=$ Saami; sar $=$ Sardinian; scm $=$ Scottish; scw $=$ western Scottish; slk $=$ Slovakian; sln $=$ Slovenian; spa $=$ Spanish; swe $=$ northern Swedish; tur $=$ Turkish; ukr $=$ Ukrainian; yug $=$ Yugoslavian. For a list of linguistic affiliations, see table 1. B-F, HG diversity within each of 47 populations, summarized on a map of Europe. The area of each pie chart is proportional to the sample size, up to a number of $\geqslant 100$; sizes are indicated schematically within $B$. The area of each black or gray sector is proportional to the frequency of the corresponding HG.


Figure 4 Spatial autocorrelation analyses. A, Correlogram, calculated using AIDA, for the entire data set. Overall significance is given. $B-G$, Correlograms, calculated using SAAP, for the six most frequent HGs. The significance of each point is indicated by its symbol, and the overall significance of each correlogram is also given. LDD = long-distance differentiation. In all correlograms, the X -axes show distance classes (km).
uninformativeness, HG 2 will not be further considered here. HG 26 occurs at low frequency (fig. 3B); like HG 2 , it lies at a deep internal node within the tree and probably contains unidentified coherent sublineages.

We find two other HGs at low frequency-HGs 8 and 22. HG 8 is common in sub-Saharan Africa (Karafet et al. 1999) and is present in our northern-African samples at $\sim 5 \%$ (fig. $3 E$ ). Only two European examples exist,
in Sardinia and France, which may represent recent admixture.

HG 22 chromosomes (fig. 3C) reach appreciable frequencies only in the French ( $5 \%$ ) and Basques (19\%). This HG has been analyzed in detail in a study elsewhere (Hurles et al. 1999), which suggested that it has a recent Iberian origin and that non-Iberian examples represent migrants. The distribution here is consistent with this analysis.

## A Major Cline Consistent with the Demic Diffusion Model

HGs 1 and 9 show complementary clines on the continental scale, from the southeast of Europe to the northwest (figs. $3 C$ and $D$ and $4 B$ and $E$ ): indeed, when the Irish sample is further subdivided on the basis of geographic information contained within surnames (Hill et al. 2000), HG 1 reaches near-fixation ( $98.5 \%$ ) in the west of Ireland. HG 9 reaches its highest frequencies ( $\sim 33 \%$ ) in the Caucasus and in Anatolia (fig. 3D), where it is thought that agriculture originated (Cavalli-Sforza et al. 1994). The strong clinal pattern of these two HGs, which together account for almost half ( $45 \%$ ) of the chromosomes in our study, resembles the first principal component of genetic variation of classical loci and is consistent with the demic diffusion hypothesis. However, distributions of the remaining HGs are very different from these and cannot be interpreted as a simple reflection of population movement from the Near East.

## A Northeast/Southwest Cline Signaling an Expansion from North of the Black Sea

The distribution of HG 3 chromosomes is also strongly clinal (fig. 4D), but with a very different axis (fig. 3D) and more on a regional scale, and is likely to reflect population-historical events distinct from those responsible for the distributions of HGs 1 and 9. It reaches its highest frequencies in central-eastern Europe, comprising approximately half of the chromosomes in the Russian, Polish, and Slovakian samples; frequencies in the southeast and southwest are low. This distribution resembles the third principal component of variation of classical gene frequencies, which has been interpreted by some geneticists (Cavalli-Sforza et al. 1994) as marking the movement, from north of the Caspian Sea, of the Kurgan people, dated to $\sim 7,000$ YBP.

## A North-South Cline: A Northern-African Influence?

Within Europe, HG 21 chromosomes are concentrated in the south (fig. $3 E$ ). Their frequency in the two northern-African samples is very high ( $52 \%$ and $77 \%$ ), and their frequencies in the Greek and Cypriot samples are also high ( $\sim 27 \%$ ), which might reflect a barrier to gene flow between Africa and Europe, as is also shown
by the analysis of autosomal protein markers (Simoni et al. 1999) and microsatellites (Bosch et al. 2000). In other southern-European populations, such as those in Spain, Portugal, Sardinia, Italy, Turkey, and Yugoslavia, frequencies are in the range of $10 \%-20 \%$. The decline in frequencies to the north is rather uniform. This regional cline (fig. 4G) has similiarities to that detected in the second principal component of classical gene frequencies (Cavalli-Sforza et al. 1994), which has been interpreted on a climatic basis.

## A Lineage Concentrated in the Northeast: A Contribution of Uralic Speakers?

HG 16 is at high frequency in the north, east of the Baltic Sea (fig. 3F), a distribution consistent with that noticed previously in a global survey (Zerjal et al. 1997). Its pattern is again clinal but regional (fig. 4F). HG 12, ancestral to HG 16, is at low frequency in the sample overall. However, its distribution overlaps that of HG 16 , with no examples in the western half of the continent, and is concentrated more in the south (fig. 3F). It is most frequent $(17 \%)$ in the Mari, who may be the population of origin of the Tat mutation, which defines HG 16 (T. Zerjal and C. Tyler-Smith, unpublished data).

With the exception of the Hungarians, who acquired their Uralic language through elite dominance by the Magyars during recent times (Cavalli-Sforza et al. 1994), all Uralic-speaking populations tested (Finnish, Estonians, Saami, and Mari) show a high frequency of HG 16. However, two nearby populations, the Lithuanians and Latvians, also show HG 16 at high frequency but speak languages of the IE family-for this lineage at least, the association appears to be geographic rather than linguistic. In the following section, we use methods that summarize variation among all lineages, to examine this issue in more detail.

## Ceography and Language as Causes of Cenetic Differentiation

Population comparisons through PC analysis.-PC analysis is a method that allows the graphic display, in a few dimensions, of the maximum amount of variance within a multivariate data set, with minimum loss of information. Figure 5 shows the results of a PC analysis of the Y-chromosome HG data, in which populations are labeled according to linguistic affiliation. PC1-PC3 summarize $71.4 \%$ of the variance.
The major division is between the two populations from northern Africa and the others. This is unsurprising, given their high frequencies of HGs 21 and 9 and their near absence of HG 1, and indicates that the Mediterranean, even at its narrowest point, has represented a barrier to gene flow, as has been suggested previously by autosomal DNA analysis. The Mediterranean pop-


Figure $5 \quad \mathrm{PC}$ analysis of Y-chromosomal HG diversity. A, PC2 plotted against PC1. B, PC3 plotted against PC2. The percentage of variance explained by each component is given on the axes. Linguistic affiliation for each population is indicated symbolically; the Belgian sample is part Dutch-/part French-speaking and has a hybrid symbol. Abbreviations are as in figure 3.
ulations of Greece and Cyprus occupy an intermediate position between the northern Africans and the rest.
Basques speak a non-IE language unrelated to any other language (Ruhlen 1991) and thus represent the most striking example of a linguistic isolate in Europe.

This isolation seems to be reflected in the PC analysis, in which they are separated from other populations (fig. $5 A$ ); however, this may be due to high frequency of a young lineage (HG 22; Hurles et al. 1999), rare elsewhere, rather than to persistence of ancient ones. Their
closest neighbors in the PC analysis are not the geographically close populations of Iberia but those of the Atlantic fringe, most of which speak Celtic-IE languages. In this context, the Cornish sample ("enw" in Figs. 3 and 5) is grouped not with the eastern English sample (ene) but with the Scottish and Irish-a reflection of geography or of the original Celtic language of this region (Ruhlen 1991) or both.
Among Uralic-speaking populations, this analysis confirms the impression given by figure $3 F$ : with the exception of the Hungarians, who lie close to IE language speakers, these populations are grouped together with the Finns separated from the rest in PC3 (fig. 5B). Also within this group are the Lithuanians and Latvians, supporting the idea that this is primarily a geographic association.
The overall impression from figure 5 is that geographic proximity may be a better predictor of Y-chromosomal genetic affinity than is language: as well as the examples discussed above, the Italic-IE language-speaking Romanians are distant from other Italic language speakers, and the Turks lie between the geographically neighoring but linguistically distant Armenians and Greeks.

## Correlating Geography, Language, and Cenetics through Mantel Testing

Mantel (1967) tests provide an objective way of assessing the relative importance of different factors in the shaping of genetic diversity. In this method, correlation coefficients between pairs of factors (from genetics, geography, and language) can be calculated, together with significance values; partial correlation coefficients are then calculated between genetics and geography and between genetics and language, with the third factor kept constant to control for the strength of the correlation between geography and language. The populations from northern Africa are linguistically remote and geographically peripheral, and the PC analysis has shown their genetic differentiation. We therefore excluded them from the Mantel analysis, to examine effects within Europe itself. Genetics and geography (table 2) are strongly and significantly correlated ( $P<.001$ ), and the correlation between genetics and language is less strong but still significant ( $P=.014$ ). The partial correlation of genetics and geography, with language kept constant, is again strong and significant ( $P<.001$ ) ; in contrast, the partial correlation of genetics and language is low and nonsignificant ( $P=.095$ ). We examined the effect of changing the values that we had assigned to distances within Uralic and within Altaic and between Irish and Scottish (see the Subjects and Methods section), and this had a negligible influence on our results. Increasing the distance assigned between language families had the effect of reducing still further the partial correlation between

Table 2
Correlation and Partial Correlation Coefficients between Genetic, Geographic, and Linguistic Distance

| Distance Considered | Correlation <br> Coefficient | $P^{a}$ |
| :--- | :---: | :---: |
| Genetics and geography | .387 | $<.001$ |
| Genetics and language | .198 | $<.01$ |
| Genetics and geography, language held constant | .349 | $<.001$ |
| Genetics and language, geography held constant | .088 | NS |

- NS = not significant.
genetics and language, as well as its significance. This analysis confirms the primacy of geography, rather than language, in the shaping of Y-chromosomal genetic diversity within Europe.


## Location of Y-Chromosomal Genetic Barriers within Europe

Although the analysis above indicates a lack of largescale correlation between language and genetics, it does not address local genetic differentiation, which may reflect local effects of language. Genetic-barrier analysis, which locates the zones of sharpest genetic change within a landscape, provides a way to do this.

Figure 6 shows the results of a genetic-barrier analysis of the Y-chromosome HG data for 45 populations, for the top $5 \%$ of barriers and a $95 \%$ significance filter (see the Subjects and Methods section). Within western Europe, minor barriers separate the Basques from some neighboring populations, the western from the eastern English, and the Dutch from the Belgians. In the east, there are two major barriers, one between the Uralicspeaking Mari and Altaic-speaking Chuvash and one between the Georgians and the Ossetians, who speak languages belonging to different families and who are also separated by the Caucasus Mountains. Most of the major barriers lie in the middle of the European landscape, running from Italy in the south to the Baltic Sea in the north, including one barrier around the island population of Gotland.

To what extent are linguistic differences contributing to Y-chromosomal barriers within Europe? Since 37 different languages are spoken among our 45 sample sites, we expect most genetic barriers to fall between populations speaking different languages. However, if language differences do constitute barriers to gene flow, then we might expect that the degree of linguistic difference between a pair of populations should correlate with the chance of a genetic barrier occurring-that is, the greatest proportion of genetic barriers should fall between populations speaking languages from different families, a lesser proportion between those speaking languages from different subfamilies, and the least between those speaking languages within a subfamily. There are


Figure 6 Significant Y-chromosomal genetic barriers within Europe. A, Output from the ORINOCO program. Positions of genetic barriers showing $95 \%$ significance after permutation (see the Subjects and Methods section) are indicated by blue through red areas on the black background, with sample sites indicated by stars. A three-dimensional animation of the actual output from the program can be viewed at the Molecular Genetics Laboratory of the McDonald Institute for Archaeological Research Web site. B, Schematic version of the output shown in $A$, with the positions of barriers indicated as thick lines on Delaunay connections (thin lines) between sample sites.

122 Delaunay connections in figure $6 B, 48$ of which are crossed by a genetic barrier. We count the proportion of connections that are crossed by a genetic barrier in each of the three classes, between language families, between subfamilies, and within subfamilies; these values are $46.2 \%$ (18/39), $40.5 \%$ (15/37), and $32.6 \%$ (15/46), respectively. Although the ranking of these three values is that expected under the hypothesis, differences between them are not significant ( $P>.1$, three-way $\chi^{2}$ test). This suggests that language may not be the primary force contributing to genetic barriers here. However, this analysis does not take into account the fact that two nonIE languages, Hungarian and Turkish, have been acquired recently: the PC analysis and the relative absence of Y-chromosomal genetic barriers around these populations supports the idea that elite dominance was not accompanied by extensive genetic admixture. If we remove these two populations and repeat the above analysis, differences between the proportions increase (to $50.0 \%$ [13/26], $43.2 \%$ [19/44], and 31.9\% [15/47], respectively) but remain not significant ( $P>.1$ ).

## Discussion

We have described the most detailed survey to date of human Y-chromosomal diversity within Europe. Samples were distributed over most of the continent, including its western and eastern fringes; inclusion of these regions, omitted from some other studies, has allowed both the detection of influences from the east and clines extending to the extreme west, for example. However, some regions remain poorly sampled, and, if the possible effects of local differentiation are to be studied, moreextensive sampling is needed. At the eastern edge of Europe lie the steppes, which stretch uninterrupted to China. Analogous studies of Asian Y chromosomes are under way and will place the European data within a broader context (W. Bao, S. Zhu, M. E. Hurles, T. Zerjal, M. A. Jobling, J. Xu, Q. Shu, R. Du, H. Yang, and C. Tyler-Smith, unpublished data).
We used 11 biallelic markers in this study, but there is still a need for more. For instance, HG 2, constituting $22 \%$ of the total sample and as much as $49 \%$ in the sample from Yugoslavia, is poorly defined and therefore constitutes a potential source of error in our analyses, since equal weight is given both to this and to welldefined HGs. The pace of new marker discovery is increasing (Underhill et al. 1997; Shen et al. 2000), and soon the resources will be available to adequately define all major European lineages.
Consistent with global surveys (Underhill et al. 1997; Karafet et al. 1999), this continental study confirms the high degree of geographic differentiation of Y-chromosomal lineages. This differentiation makes the $Y$ chromosome a sensitive indicator of either admixture,
as demonstrated in studies of Polynesia (Hurles et al. 1998), South America (Bianchi et al. 1997), and Uruguay (Bravi et al. 1997), for example, or an absence of admixture, as has been shown in Jewish populations in Europe and northern Africa (Hammer et al. 2000). Knowledge about admixture is of particular importance in the choice of populations for studies that use linkagedisequilibrium analysis (McKeigue 1997) in both simple and complex disorders.

## Clines of Y-Chromosomal HGs

The effects of drift on human Y-chromosome diversity are likely to be great. It is striking, therefore, to observe clear clinal variation in five of the six major lineages within Europe-this suggests that drift has not erased the patterns of variation established by past population movement. Natural selection on Y chromosomes (Jobling and Tyler-Smith 2000) provides an alternative explanation for such clines; possible effects of geographically variable factors (such as temperature) on fertility within specific lineages have yet to be investigated, but, in the absence of evidence to the contrary, we assume that the variation that we are assaying is selectively neutral and can therefore be interpreted in terms of population history.

The contrast between the clinal variation of Y-chromosomal lineages and the lack of clines in mtDNA data (Simoni et al. 2000a) is marked, although the latter is still a matter of debate (Simoni et al. 2000b; Torroni et al. 2000). It seems consistent with studies of global genetic diversity (Seielstad et al. 1998), which have ascribed such differences to patrilocality. However, direct evidence about mating practices in European prehistory is lacking-indeed, populations in some regions, such as northern Iberia, may have practiced matrilocality (Collins 1986).

Clines for HGs 1 and 9, encompassing $45 \%$ of the chromosomes-and doing so on a continental scaleshow a pattern similar to that seen both in the first principal component of classical gene-frequency data and in the autocorrelation analysis of six Y-chromosomal microsatellites (Casalotti et al. 1999). A simplistic interpretation is that HG 9 chromosomes were carried in a major demographic expansion of agricultural migrants from the Near East and that HG 1 chromosomes were a preexisting predominant European lineage. Estimates of the ages of these lineages, from coalescent analysis, are not inconsistent with this scenario: the mutation defining HG 1 has been dated at $\sim 23,000$ YBP (Karafet et al. 1999), and that defining HG 9 has been dated at $14,800 \pm 9,700$ YBP (Hammer et al. 2000).

Demic diffusion-and, indeed, any major directional gene-flow process-is generally expected to generate clines for only a fraction of the alleles at one locus (Sokal
et al. 1989, 1997). Although two HGs show clines compatible with expansion from the Near East, three further lineages show different clinal patterns, indicating distinct population movements: southward and westward from north of the Black Sea (HG 3), from eastern Europe or northern Asia westward to the Baltic Sea (HG 16), and from south to north (HG 21). These clines are more regionally localized than those for HGs 1 and 9 , pointing to phenomena affecting only part of the continent. It is tempting to assign known or surmised population-historical movements to these genetic gradients, but this should be done with caution.
The distribution of HG 3 chromosomes resembles the third principal component of variation of classical gene frequencies. There are several possible interpretations of this pattern. One explanation (Cavalli-Sforza et al. 1994) is that it marks the Kurgan expansion from north of the Caspian Sea, dated to $\sim 7,000$ YBP. However, alternative explanations-such as the spread of pastoralism, or east-to-west movements of people such as the Scythians, Mongols, and Huns-seem equally likely (Renfrew 2000). Globally, HG 3 chromosomes are absent from Africa and the Americas, but their distribution is wide within Asia as well as in Europe (Zerjal et al. 1999), consistent with their association with a recent and major expansion within Eurasia. Microsatellite diversity analysis (Zerjal et al. 1999) used the mutationrate estimates of Heyer et al. (1997) to date the most recent common ancestor of a set of European and Asian HG 3 chromosomes to 3,800 YBP ( $95 \%$ confidence interval [CI] 1,600-13,000 YBP); the use of more-recent mutation-rate estimates (Kayser et al. 2000) would yield a date of 2,550 YBP ( $95 \%$ CI $1,650-4,260$ YBP). Coalescent analysis has dated the SRY-1532 mutation defining HG 3 to $\sim 7,500$ YBP (Karafet et al. 1999). If these dates are to be relied on, they seem to suggest that the expansion of HG 3 chromosomes was due to population movements later than those of the Kurgan people.
Currently, dates cannot be attached to the clines, and the modern distributions of lineages are the outcome of many millennia of population movement. Assigning plausible dates to demographic movements is important, and here the Y chromosome can potentially contribute. Finer-scale definitions of monophyletic lineages within Europe, by use of new markers, and the analysis of these, by use of microsatellites, offers the possibility that timescales for the major demographic events can be inferred.

## Language, Ceography, and $\gamma$-Chromosomal Diversity

The Mantel tests demonstrate that patterns of Y-chromosomal genetic variation do not correlate as well with language as with geography. However, it should be borne in mind that geography and language together explain
only $16.8 \%$ of the genetic variance (data not shown); therefore, other forces, such as founder effects and genetic drift, have also been important in determining the current patterns of spatial variation. Our findings seem at odds with those of Poloni et al. (1997), who showed that most of the population differentiation of Y-chromosome haplotypes was due to language. However, there are important differences between the two studies. The samples of Poloni et al. (1997) were global, rather than from a single continent, and showed a correspondingly greater linguistic and genetic diversity. The populations that we have studied are located within a single continent, and most speak languages belonging to one language family, IE; indeed, much of the genetic patterning that we now see may have its roots in the spread of that language family (Renfrew 1987). The effect of increasing genetic, geographic, and linguistic diversity in the input to the Mantel tests can be seen by including the northern-African samples (data not shown), which are both geographically and linguistically distant from most other populations. This increases the partial correlations between genetics and geography and between genetics and language and also increases the significance of the latter to $P=.024$, which, however, is still lower than the significance of the genetics-geography partial correlation ( $P<.001$ ).
The results of genetic-barrier analysis (fig. 6) need to be interpreted with caution when, as in this case, sample distribution is uneven; the method is likely to be sensitive to the introduction of new populations, especially between existing sample sites that are far apart. However, the analysis has suggested that there is little correlation between genetic barriers and levels of linguistic separation, even when elite dominance is taken into account by removing the Hungarians and Turks from the analysis. Although cultural factors other than language (such as politics and religion) might also be associated with genetic barriers, we have examined language because it has the greatest time depth. However, this is still likely to be less than the age of geographic barriers, the relative importance of which cannot easily be analyzed. Twentyfive of 48 Delaunay connections crossed by genetic barriers also coincide with geographic barriers (under a conservative definition that considers only large stretches of water and the two major mountain ranges, the Alps and the Caucasus), which seems to emphasize the greater importance of geographic factors in subdividing populations, resulting in large differences in Y-chromosomal HG frequencies.

In synthesis, it seems that many kinds of barriers are probably recent, on an evolutionary timescale (see Renfrew 1987); after they have been established, fluctuations of allele frequencies have become partly or largely independent in the populations separated by those barriers. Therefore, it is perhaps not surprising to find little
correlation between the degree of language differentiation at a language boundary and the amount of genetic change observed across that boundary. As has been shown in the analysis of protein polymorphisms (Sokal et al. 1990), linguistic differences tend to cause some degree of population subdivision, regardless of whether such differences are between language families, between languages of the same family, or even between dialects of the same language.
Although we have dichotomized the forces of geography and language, in reality they work together; spatially coincident weak geographic and linguistic barriers may together form strong barriers to gene flow. Some of the strongest genetic barriers observed, in central Europe, coincide with neither strong linguistic nor strong geographic barriers. Linguistic and geographic heterogeneities and the effects of drift, on a background retaining a strong signal of expansion from the Near East and of other migrations, have combined to shape the genetic landscape of Europe.

## Acknowledgments

We thank the DNA donors for making this study possible, and we thank Laurent Excoffier for assistance. Z.H.R. was supported by a BBSRC Studentship, T.Z. by a Wellcome Trust Bioarchaeology Studentship, M.E.H. by an MRC Studentship, F.R.S. by the Leverhulme Trust, and L.P. by Ph.D. grant PRAXIS XXI/BD/13632/97 from Fundação para a Ciência e a Tecnologia. D.C.R. is a Glaxo Wellcome Research Fellow. C.T.-S. is supported by the CRC, and M.A.J. is a Wellcome Trust Senior Fellow in Basic Biomedical Science, supported by grant 057559. Iberian sample collection was partially funded by multidisciplinary project grant PR182/96 6745 from Complutense University.

## Electronic-Database Information

The URL for data in this article is as follows:
Molecular Genetics Laboratory of the McDonald Institute for Archaeological Research, http://www-mcdonald.arch.cam .ac.uk/Genetics/home.html

## References

Adams J, Otte M (1999) Did Indo-European languages spread before farming? Curr Anthropol 40:73-77
Ammerman AJ, Cavalli-Sforza LL (1984) Neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton, NJ
Barbujani $G$ (1991) What do languages tell us about human microevolution? Trends Ecol Evol 6:151-156
(1997) DNA variation and language affinities. Am J Hum Genet 61:1011-1014
Barbujani G, Oden NL, Sokal RR (1989) Detecting regions of abrupt change in maps of biological variables. Syst Zool 38: 376-389

Barbujani G, Pilastro A, de Domenico S, Renfrew C (1994) Genetic variation in North Africa and Eurasia: neolithic demic diffusion vs. paleolithic colonisation. Am J Phys Anthropol 95:137-154
Bertorelle G, Barbujani G (1995) Analysis of DNA diversity by spatial autocorrelation. Genetics 140:811-819
Bianchi NO, Bailliet G, Bravi CM, Carnese RF, Rothhammer F, Martínez-Marignac VL, Pena SDJ (1997) Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers. Am J Phys Anthropol 102:79-89
Blanco P, Shlumukova M, Sargent CA, Jobling MA, Affara N, Hurles ME (2000) Divergent outcomes of intra-chromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. J Med Genet 37:752-758
Bosch E, Calafell F, Pérez-Lezaun A, Clarimón J, Comas D, Mateu E, Martínez-Arias R, Morera B, Brakez Z, Akhayat O, Sefiani A, Hariti G, Cambon-Thomsen A, Bertranpetit J (2000) Genetic structure of north-west Africa revealed by STR analysis. Eur J Hum Genet 8:360-366
Bosch E, Calafell F, Santos FR, Pérez-Lezaun A, Comas D, Benchemsi N, Tyler-Smith C, Bertranpetit J (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am J Hum Genet 65:1623-1638
Boyd R, Silk JB (1997) How humans evolved. WW Norton, New York
Brassel KE, Reif D (1979) A procedure to generate Thiessen polygons. Geogr Anal 11:289-303
Bravi CM, Sans M, Bailliet G, Martinez-Marignac VL, Portas M, Barreto I, Bonilla C, Bianchi NO (1997) Characterization of mitochondrial DNA and Y-chromosome haplotypes in a Uruguayan population of African ancestry. Hum Biol 69: 641-652
Casalotti R, Simoni L, Belledi M, Barbujani G (1999) Y-chromosome polymorphisms and the origins of the European gene pool. Proc R Soc Lond B Biol Sci 266:1959-1965
Casanova M, Leroy P, Boucekkine C, Weissenbach J, Bishop C, Fellous M, Purrello M, Fiori G, Siniscalco M (1985) A human Y-linked DNA polymorphism and its potential for estimating genetic and evolutionary distance. Science 230: 1403-1406
Cavalli-Sforza LL, Menozzi P, Piazza A (1993) Demic expansions and human evolution. Science 259:639-646
(1994) The history and geography of human genes. Princeton University Press, Princeton, NJ
Chikhi L, Destro-Bisol G, Bertorelle G, Pascali V, Barbujani G (1998a) Clines of nuclear DNA markers suggest a largely neolithic ancestry of the European gene pool. Proc Natl Acad Sci USA 95:9053-9058
Chikhi L, Destro-Bisol G, Pascali V, Baravelli V, Dobosz M, Barbujani G (1998b) Clinal variation in the nuclear DNA of Europeans. Hum Biol 70:643-657
Collins R (1986) The Basques. Blackwell, Oxford
Comas D, Calafell F, Mateu E, Pérez-Lezaun A, Bosch E, Bertranpetit J (1997) Mitochondrial DNA variation and the origin of the Europeans. Hum Genet 99:443-449
Cooper G, Amos W, Hoffman D, Rubinsztein DC (1996) Network analysis of human Y microsatellite haplotypes. Hum Mol Genet 5:1759-1766

Dennell R (1983) European economic prehistory: a new approach. Academic Press, London
Dyen I, Kruskal JB, Black P (1992) An Indoeuropean classification: a lexicostatistical experiment. Trans Am Philos Soc 82:1-132
Gimbutas M (1970) Proto-Indo-European culture: the Kurgan culture during the fifth, fourth and third millennia B.C. In: Cardona G, Hoenigswald HM, Senn A (eds) Indo-European and Indo-Europeans. University of Pennsylvania Press, Philadelphia, pp 155-195
Hammer MF (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11:749-761
Hammer MF, Horai S (1995) Y-chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951-962
Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, Zegura SL (1998) Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. Mol Biol Evol 15: 427-441
Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, Ostrer H, Bonné-Tamir B (2000) Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. Proc Natl Acad Sci USA 97:6769-6774
Harpending H, Jenkins T (1973) Genetic distance among Southern African populations. In: Crawford MH, Workman PL (eds) Methods and theories of anthropological genetics. University of New Mexico Press, Albuquerque, pp 177-199
Hassan FA (1973) On mechanisms of population growth during the neolithic. Curr Anthropol 14:535-542
Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6:799-803
Hill EW, Jobling MA, Bradley DG (2000) Y chromosomes and Irish origins. Nature 404:351-352
Hurles ME (1999) Mutation and variability of the human Y chromosome genetics. University of Leicester, Leicester
Hurles ME, Irven C, Nicholson J, Taylor PG, Santos FR, Loughlin J, Jobling MA, Sykes BC (1998) European Ychromosomal lineages in Polynesia: a contrast to the population structure revealed by mitochondrial DNA. Am J Hum Genet 63:1793-1806
Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Pérez-Lezaun A, Bosch E, Shlumukova M, CambonThomsen A, McElreavey K, López de Munain A, Röhl A, Wilson IJ, Singh L, Pandya A, Santos FR, Tyler-Smith C, Jobling MA (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Ychromosomal DNA polymorphism. Am J Hum Genet 65: 1437-1448
Jobling MA, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. Trends Genet 11:449456

- (2000) New uses for new haplotypes: the human Y chromosome, disease, and selection. Trends Genet 16:356-362
Jobling MA, Williams G, Schiebel K, Pandya A, McElreavey K, Salas L, Rappold GA, Affara NA, Tyler-Smith C (1998)

A selective difference between human Y-chromosomal DNA haplotypes. Curr Biol 8:1391-1394
Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, deKnijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. Am J Hum Genet 64:817-831
Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human $Y$ chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66:1580-1588
Kwok C, Tyler-Smith C, Medonca BB, Hughes I, Berkovitz GD, Goodfellow PN, Hawkins JR (1996) Mutation analysis of $2 \mathrm{~kb} 5^{\prime}$ to SRY in XY females and XX intersex subjects. J Med Genet 33:465-468
Landers J (1992) Reconstructing ancient populations. In: Jones S, Martin R, Pilbeam D (eds) The Cambridge encyclopedia of human evolution. Cambridge University Press, Cambridge, pp 402-405
Langaney A, Roessli D, van Blyenburgh NH, Dard P (1992) Do most human populations descend from phylogenetic trees? Hum Evol 7:47-61
Lucotte G, Loirat F (1999) Y-chromosome DNA haplotype 15 in Europe. Hum Biol 71:431-437
Malaspina P, Cruciani F, Ciminelli BM, Terrenato L, Santolamazza P, Alonso A, Banyko J, Brdicka R, Garcia O, Gaudiano C, Guanti G, Kidd KK, Lavinha J, Avila M, Mandich P, Moral P, Qamar R, Mehdi SQ, Ragusa A, Sefanescu G, Caraghin M, Tyler-Smith C, Scozzari R, Novelletto A (1998) Network analyses of Y-chromosomal types in Europe, northern Africa, and western Asia reveal specific patterns of geographic distribution. Am J Hum Genet 63:847-860
Mantel NA (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209-220
Mathias N, Bayés M, Tyler-Smith C (1994) Highly informative compound haplotypes for the human Y chromosome. Hum Mol Genet 3:115-123
McKeigue PM (1997) Mapping genes underlying ethnic differences in disease risk by linkage disequilibrium in recently admixed populations. Am J Hum Genet 60:188-196
Menozzi P, Piazza A, Cavalli-Sforza LL (1978) Synthetic maps of human gene frequencies in Europeans. Science 201: 786-792
Ngo KY, Vergnaud G, Johnsson C, Lucotte G, Weissenbach J (1986) A DNA probe detecting multiple haplotypes of the human Y chromosome. Am J Hum Genet 38:407-418
Piazza A, Rendine S, Minch E, Menozzi P, Mountain J, CavalliSforza LL (1995) Genetics and the origin of European languages. Proc Natl Acad Sci USA 92:5836-5840
Poloni ES, Semino O, Passarino G, Santachiara-Benerecetti AS, Dupanloup L, Langaney A, Excoffier L (1997) Human genetic affinities for Y-chromosome P49a,f/TaqI haplotypes show strong correspondence with linguistics. Am J Hum Genet 61:1015-1035
Quintana-Murci L, Semino O, Minch E, Passarino G, Brega A, Santachiara-Benerecetti AS (1999) Further characteristics of proto-European Y chromosomes. Eur J Hum Genet 7:603-608

Renfrew C (1987) Archaeology and language: the puzzle of Indo-European origins. Jonathan Cape, London
(1989) The origins of Indo-European languages. Sci Am 261:106-114
(2000) At the edge of knowability: towards a prehistory of languages. Camb Archaeol J 10:7-34
Richards M, Côrte-Real H, Forster P, Macaulay V, WilkinsonHerbots H, Demaine A, Papiha S, Hedges R, Bandelt H-J, Sykes B (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59: 185-203
Richards M, Sykes B (1998) Evidence for Paleolithic and Neolithic gene flow in Europe. Am J Hum Genet 62:491-492
Ruhlen M (1991) A guide to the world's languages. Edward Arnold, London
Santos FR, Carvalho-Silva DR, Pena SDJ (1999) PCR-based DNA profiling of human Y chromosomes. In: Epplen JT, Lubjuhn T (eds) Methods and tools in biosciences and medicine. Birkhaüser Verlag, Basel, pp 133-152
Santos FR, Tyler-Smith C (1996) Reading the human Y chromosome: the emerging DNA markers and human genetic history. Braz J Genet 19:665-670
Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN ver 2.0: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva
Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, Cavalli-Sforza LL (1994) Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. Hum Mol Genet 3:2159-2161
Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278-280
Semino O, Passarino G, Brega A, Fellous M, Santachiara-Benerecetti AS (1996) A view of the Neolithic demic diffusion in Europe through two Y chromosome-specific markers. Am J Hum Genet 59:964-968
Shen P, Wang F, Underhill PA, Franco C, Yang W-H, Roxas A, Sung R, Lin AA, Hyman RW, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (2000) Population genetic implications from sequence variation in four $Y$ chromosome genes. Proc Natl Acad Sci USA 97:7354-7359
Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G (2000a) Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet 66:262-278
(20006) Reconstruction of prehistory on the basis of genetic data. Am J Hum Genet 66:1177-1179
Simoni L, Gueresi P, Pettener D, Barbujani G (1999) Patterns of gene flow inferred from genetic distances in the Mediterranean region. Hum Biol 71:399-415
Sokal RR, Harding RM, Oden NL (1989) Spatial patterns of human gene frequencies in Europe. Am J Phys Anthropol 80:267-294

Sokal RR, Oden NL (1978) Spatial autocorrelation in biology. Biol J Linn Soc 10:199-249
Sokal RR, Oden NL, Legendre P, Fortin MJ, Kim J, Thomson BA, Vaudor A, Harding RM, Barbujani G (1990) Genetics and language in European populations. Am Nat 135: 157-175
Sokal RR, Oden NL, Thomson BA (1997) A simulation study of microevolutionary inferences by spatial autocorrelation analysis. Biol J Linn Soc 60:73-93
Sokal RR, Oden NL, Wilson C (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. Nature 351:143-145
Swadesh M (1952) Lexico-statistic dating of prehistoric ethnic contacts: with special reference to North American Indians and Eskimos. Proc Am Philos Soc 96:452-463
Templeton AR (1993) The "Eve" hypothesis: a genetic critique and reanalysis. Am Anthropol 95:51-72
Torroni A, Richards M, Macaulay V, Forster P, Villems R, Nørby S, Savontaus M-L, Huoponen K, Scozzari R, Bandelt H-J (2000) mtDNA haplogroups and frequency patterns in Europe. Am J Hum Genet 66:1173-1177
Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chramosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996-1005
Veitia R, Ion A, Barbaux S, Jobling MA, Souleyreau N, Ennis K, Ostrer H, Tosi M, Meo T, Chibani J, Fellous M, McElreavey K (1997) Mutations and sequence variants in the testis-determining region of the Y chromosome in individuals with a 46,XY female phenotype. Hum Genet 99:648-652
Whitfield LS, Sulston JE, Goodfellow PN (1995) Sequence variation of the human Y chromosome. Nature 378:379-380
Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhövel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C (1997) Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60: 1174-1183
Zerjal T, Pandya A, Santos FR, Adhikari R, Tarazona E, Kayser M, Evgrafov O, Singh L, Thangaraj K, Destro-Bisol G, Thomas MG, Qamar R, Mehdi Q, Rosser ZH, Hurles ME, Jobling MA, Tyler-Smith C (1999) The use of Y-chromosomal DNA variation to investigate population history: recent male spread in Asia and Europe. In: Papiha SS, Deka R, Chakraborty R (eds) Genomic diversity: applications in human population genetics. Plenum Press, New York, pp 91-102
Zvelebil M, Zvelebil KV (1988) Agricultural transition and Indo-European dispersal. Antiquity 62:574-583


[^0]:    ${ }^{1}$ Department of Genetics, University of Leicester, Leicester; ${ }^{2} \mathrm{CRC}$ Chromosome Molecular Biology Group, Department of Biochemistry, and ${ }^{3}$ Institute of Molecular Medicine, University of Oxford, Oxford; ${ }^{4}$ Institute of General and Molecular Pathology, University of Tartu and ${ }^{5}$ Estonian Biocentre, Tartu, Estonia; ${ }^{6}$ Laboratory for Radiobiology and Molecular Genetics, Institute of Nuclear Sciences "Vinca," Belgrade; ${ }^{7} 5$ IPATIMUP and Faculdade de Ciências, Universidade do Porto, Porto, Portugal; ${ }^{8}$ Department of Zoology, University of Cambridge, Cambridge; ' ${ }^{9}$ Centro de Investigacion y Criminalistica, Laboratorio de ADN, Policia Judicial, Guardia Civil, and ${ }^{10}$ Laboratorio de Biología Forense, Departamento de Toxicología y Legislación Sanitaria, Univ-

