# Metabolite profiling of *Clostridium difficile* ribotypes using small molecular weight volatile organic compounds

S. Kuppusami,[1] M.R.J. Clokie,[2] T. Panayi[2], A.M. Ellis [1] and P.S. Monks[1*]

[1]Department of Chemistry, University of Leicester, Leicester, LE1 7RH, UK.

[2]Department of Immunity, Infection and Inflammation, University of Leicester, Leicester, LE1 7RH, UK.

*corresponding author – Prof. Paul S. Monks, Department of Chemistry, University of Leicester, Leicester, LE1 7RH, UK, e-mail P.S.Monks@le.ac.uk, Tel. +44-(0)116-252141, Fax. +44-(0)116-2523789.

**Abstract**

**Volatile organic compounds (VOCs) emitted by cultures of ten different *Clostridium difficile* ribotypes have been profiled using Proton Transfer Reaction-Time of Flight-Mass Spectrometry (PTR-ToF-MS). A total of 69 VOCs were identified and combinations of these VOCs were found to be characteristic for each of the ribotypes. The VOC patterns, with the aid of a statistical analysis, have been shown to be useful in distinguishing different ribotypes. A tentative assignment of different masses also shows that different ribotypes have markedly different emissions of methanol, p-cresol, dimethylamine and a range sulfur compounds (ethylene sulfide, dimethylsulfide and methyl thioacetate), which point to VOCs as potential indicators of different metabolic pathways in virulent and less-virulent strains. The results establish the potential of detecting emitted VOC metabolites to differentiate between closely related *C. difficile* ribotypes and in the longer term provide metabolic insight into virulence.**

Keywords: Volatile organic compounds, Clostridium difficile, ribotype

## 1. Introduction

*Clostridium difficile* is a spore-forming bacterium that causes infectious diarrhoea. The symptoms range from a mild disturbance to severe illness with ulceration and bleeding from the colon and perforation of the intestine, which can be fatal (Bartlett 1990; Knoop et al. 1993). Patients treated with broad spectrum antibiotics are at the greatest risk of contracting *C. difficile* infection. In addition to antibiotic exposure, elderly patients with compromised immune systems, who have recently undergone surgery and who have long periods of time in healthcare settings, are also at higher risk (Settle and Kerr 2011). If *C. difficile* infection is not diagnosed promptly and treated appropriately, it can progress to colitis and possible death (Settle and Kerr 2011). There are many *C. difficile* ribotypes (genotypes) that frequently cause infection or outbreaks in clinical settings (Wilcox et al. 2012). The pathogenic ribotypes, such as the PCR ribotype 027, are termed as "hypervirulent" and often cause large outbreaks of infection (Burns and Minton 2011; Clements et al. 2010). At present, enzyme immunoassay (EIA) tests or tests based on PCR are used to identify faecal toxins or toxin genes produced by *C. difficile* (Grein et al. 2014). These tests are rapid, but there are problems with both sensitivity and specificity (Probert 2011). Furthermore, the problematic ribotypes are an evolving target and it is difficult for the existing tests to respond in appropriate timeframes in order to detect these changes in epidemiology. Molecular PCR-based tests depend on *a priori* knowledge of the gene sequences of strains being detected and thus new or different strains may be missed. In contrast, a test that is dependent on metabolic VOCs would be free from such constraints.

Microorganisms produce metabolic VOCs for various reasons, such as (1) during growth (Bunge et al. 2008), (2) as infochemicals for inter- and intra-organism communication (Kai et al. 2009; Perl et al. 2011), (3) for cell-to-cell communication signals (Bunge et al. 2008), or (4) as growth-promoting or inhibiting agents to their own populations, or to other species (Bunge et al. 2008; Kai et al. 2009). There is growing interest in the detection and identification of bacteria by measuring their release of metabolic volatile organic compounds (VOCs) (Tait et al. 2014). A large variety of VOCs are produced by bacteria, including fatty acids, aliphatic alcohols, ketones, dimethyl polysulfides, alkenes, nitrogen-containing compounds and volatile sulfur-containing compounds (Bunge et al. 2008; Scotter et al. 2006; Thorn et al. 2011). Presently, the biological functions of many bacterial volatiles are not understood in detail.

Metabolite VOCs of *Clostridium* have been measured from microbial headspace *in vitro*. One of the earliest papers that described the production of volatiles by *Clostridium spp* demonstrated the

release of dimethyl disulfide, various short chain acids, 2,3-butanediol, isopentanol and acetoin (Stotzky and Schenck 1976). Pons *et al.* used volatile amines released by various *Clostridium* species, such as dimethylamine, trimethylamine, isobutylamine and 3-methylbutylamine, as markers to differentiate between *Clostridium* species (Pons et al. 1985). In the 1980s, several studies concentrated on the rapid identification of *C. difficile* and reported that this was possible using *p*-cresol and caproic acid as markers (Berg et al. 1985; Nunez-Montiel et al. 1983; Phillips and Rogers 1981). Nunez-Montiel *et al.* noted that no other *Clostridial* bacteria other than the *difficile* sub-species, or indeed other microorganisms tested in their analysis, produce p-cresol and caproic acid when inoculated in norleucine-tyrosine broth, making these two compounds potential markers for *C. difficile* (Nunez-Montiel et al. 1983). The use of VOCs to distinguish between *C. difficile* and other bacteria such as *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhimurium*, and *Pseudomonas aeruginosa*, was attempted by Bruins et al. (Bruins et al. 2009). This study concluded that it was possible to identify and discriminate *C. difficile* strains from other bacteria using VOC analysis. However, the VOCs released were greatly dependent on factors such as growth media, growth phase, growth conditions (pH, temperature, humidity, oxygen content) and species investigated.

*C. difficile* continues to be a major healthcare problem and new strains continually emerge and circulate (Wiegand et al. 2012). Although there has been an increased effort in sequencing the genome of *C. difficile* strains, less work has been carried out on downstream analysis of the sequence data. For this study we selected ten ribotypes of *C. difficile* which were isolated from local sources but which represent clinically relevant strains of various degrees of prevalence. The ribotype 027 is a known hypervirulent strain that is associated with many outbreaks (Labbe et al. 2008). R014/020 is historically the most prevalent ribotype in Europe and it still accounts for ≈15% of all cases (Cheknis et al. 2009; Freeman et al. 2010). The ribotypes 002 and 078 are both abundant in Europe, with 078 being increasingly a cause for concern by its association with a high prevalence of infection (Bauer et al. 2011; Stabler et al. 2012). In contrast, ribotypes 013, 05, 107, 026, 087 are less prevalent. The environmental strain 076 was also used as a comparison (Hargreaves et al. 2013). The motivation behind our investigation was twofold; the first was a fundamental exploration of multiple VOC "fingerprints" as a metabolome for *C. difficile* ribotypes, the second was to assess the potential in outline of the application of emitted VOCs as a rapid and non-invasive method to diagnose *C. difficile* infections without the need for sample preparation. As a first step towards these goals, we analysed the profiles of gas-phase VOC metabolites for ten distinct ribotypes of *C. difficile*. We also attempted to metabolically profile *C. difficile* at the ribotype by

analysing volatile organic compounds released by the bacteria using PTR-ToF-MS, and we present the findings here.

## 2. Methods

### *2.1 Growth and maintenance of microorganisms*

The following *C. difficile* ribotypes were used in this study: R027, R014/R020, R002, R013, R005, R107, R026, R087, R078, and R076. All ribotypes, except R076, were clinical samples isolated from stool specimens collected from patients infected with *C. difficile* at the UHL Leicester hospitals (in descending order of number of cases detected). All cultures were isolated in the Department of Infection, Immunity, and Inflammation at the University of Leicester using standard procedures (Nale et al. 2012). Ribotypes were determined according to standard protocols (Indra et al. 2008) with reference to the strains categorized as part of the *C. difficile* ribotyping network (Fawley and Wilcox, pers comm.). All ribotypes are clinically relevant and commonly found in the UK (Stabler et al., 2012). The bacteria were cultured in agar plates containing Brain Heart Infusion (BHI) agar supplemented with 7% horse blood and incubated at 37 °C anaerobically for 48 hours when required.

### *2.2 Bacterial culture experimental design*

All bacterial culture headspace analysis was performed under anaerobic conditions and 37°C, the optimum conditions needed to maintain *C. difficile* growth. A custom-made glass container which can accommodate a single culture plate was used. The upper section of the container has two outlets, one connected to the PTR-ToF-MS instrument and the other acts as an inlet for gas flow into the sample container. A culture plate with the lid removed was placed into the glass container with the upper and lower sections of the container secured to form an air-tight chamber. The glass chamber was heated to 37 °C by wrapping the chamber with a thermal blanket. Prior to the addition of any bacteria-laden culture plate, anaerobic gas (80% $N_2$, 10% $H_2$ and 10% $CO_2$) (BOC, UK) was supplied into the glass container and a background measurement was taken using PTR-ToF-MS. An agar plate inoculated with *C. difficile* was then placed into the glass container and analysed for 10 minutes with a flow of 150 ml $min^{-1}$ of the effluent fed to the PTR-ToF-MS instrument. After each headspace analysis a corresponding uninoculated blood agar plate supplemented with 7% horse blood (blank) was similarly analysed as a culture medium control. Five distinct samples of each *C. difficile* ribotype were analysed on different days in a period spread over two months in order to demonstrate the consistency of VOC signatures emitted by the cultures over a longer time-period.

The samples used in the analyses presented in this paper where all from the same growth-phase between 48 and 72 hours of incubation. Furthermore, a test group (blind test) consisting of four unknown ribotypes with five samples each, were analysed in the same manner as per other cultures previously analysed. This analysis took place one month after the initial analysis ended.

*2.3 Proton Transfer Reaction-Time of Flight-Mass Spectrometry (PTR-ToF-MS)*

VOC headspace samples were analysed using PTR-ToF-MS (Blake et al. 2009; Blake et al. 2004). Details of the instrument used in this work and its performance have been described previously (Barber et al. 2012; White et al. 2013; Wyche et al. 2007). PTR-MS exploits proton transfer, in this case from $H_3O^+$, to transfer protons to neutral volatile organic compounds, with subsequent detection by mass spectrometry. The instrument used consists of a radioactive ion source which ionizes water molecules to form $H_3O^+$. Proton transfer occurs if the proton affinity of the neutral analyte molecule is higher than the proton affinity of $H_2O$. Major inorganic components of ambient air such as $N_2$, $O_2$, and $CO_2$, have lower proton affinities than water and so are not observed. The instrument used in this work has a time-of-flight (ToF) mass analyser, which measures a full range of ion masses simultaneously. Detection sensitivities below the parts per billion by volume (ppbV) level are easily achievable. The technique directly analyzes gas mixtures without sample pre-concentration and measures in real-time. As PTR-MS is a relatively soft-ionisation technique most signal peaks can be assumed to arise from protonated molecular ions.

PTR-ToF-MS operating conditions were as follows: drift tube voltage, 2190 V; drift tube pressure, 6 mbar; drift tube temperature, 23°C; *E/N*, 90/190 Townsends (Td) (where *E* is the electric field strength in the drift tube and *N* is the gas number density in the drift tube; inlet flow, 150 ml $min^{-1}$.

The C. difficile samples were relatively wet samples grown on solid growth medium. The average ratio of m/z 19/37 were around 14 with typical absolute value of hydronium was 715,000 and for monohydrate cluster ions at 51,000.

*2.4 Statistical Analysis*

For all of the culture samples the raw PTR-ToF-MS data collected were normalized to $10^6$ counts per second (cps) of the sum of $H_3O^+$ and $H_3O^+(H_2O)$ (*m/z* 19 and *m/z* 37, respectively) signals. To avoid carrying over artifacts between individual samples, data recorded for the first 5 minutes were not

included in the data analysis. The next 5 minutes of data recorded were used for statistical analysis. First, the significance of the VOCs measured in the bacterial cultures were compared to the measured headspace of the glass chamber containing the uninoculated blood agar plate using a two-sided Mann-Whitney test (using the Minitab software). This test was used to assess whether the VOCs detected from each of the ribotype cultures were significantly different from the uninoculated agar, where a *p*-value < 0.05 was considered as statistically significant. The VOCs emitted from all ribotypes selected that satisfy the Mann-Whitney test and that had signals more than 100 normalised counts per second (ncps) after subtracting blank medium measurements, were then used in multivariate analysis. Specifically, the data were subjected to principal component analysis (PCA), partial least square discriminant analysis (PLSDA) and cluster analysis. PCA is a statistical tool used to visualize patterns of classification from the data sets analysed. It simplifies high dimensionality data sets by converting observations into principal components that emphasise variances in the data (Varmuza and Filzmoser 2009). A leave-one-out cross validation PCA method was applied. A dendrogram was produced using hierarchical cluster analysis with Mahalanobis distance coefficients, which is visualizing tool for classifying *C. difficile* ribotypes. The aim was to show similarities and differences found in the ribotypes investigated. PLSDA, a multivariate regression method, was used for ribotype classification by leave-one-out cross-validation. PLSDA assesses the relationship between *m/z* measurements (descriptor matrix) and sample class (response matrix) to predict whether a ribotype belonged to its own class or other classes of ribotypes investigated. PCA, PLSDA and cluster analysis was performed using PLS Toolbox (Eigenvector Research Inc., USA) operated in Matlab.

## 3. Results and Discussion

### *3.1 Clostridium difficile ribotype analysis*

Figure 1 compares individual mass spectra accumulated over one minute (after excluding the initial 5 minutes at the start of the analysis and blank subtracted) obtained from the headspace of the analysed cultures of *C. difficile* ribotypes. The measurements reveal complex mass spectra for six of the ribotypes: R027, R014/R020, R002, R013, R005 and R107. In contrast less complex mass spectra were found for R026, R078, R087 and R076. Few peaks are observed in the 200 to 300 amu mass region after subtraction of blank measurements. Mann-Whitney significance tests showed that only masses between 15 and 200 amu were significantly different from the blank medium measurements. The *C. difficile* ribotype cultures produced signals at different relative abundances, some ranging over more than six orders of magnitude. From the Mann-Whitney tests 69 signals in

the mass range between 15 and 200 were significantly different ($p < 0.05$) compared to those of the blank medium samples.

PCA was used to generate a visual representation of the discrimination between the ribotypes by their metabolite profiles using the 69 mass peaks identified as being significant. The first two principal components accounted for 68.97% of the variance. Figure 2 shows that most of the ribotypes can be separated from each other, with the exception of ribotypes R014/R020, R002 and R013, which almost overlap each other. Ribotypes R107 and R005 were clustered very closely but there is a clear separation between them. All of the other ribotypes can be clearly distinguished from each other.

In order to move this to a more quantitative description a partial least squares discriminant analysis was performed. PLSDA evaluates the effectiveness of a multi-variate statistical model in terms of sensitivity and specificity. Sensitivity evaluates how good the model is at classifying a ribotype into its correct class. Specificity indicates how likely a ribotype is not classified into a wrong class. The sensitivity and specificity values of the classification by PLSDA are shown in Table 1. The PLSDA model predicts all ribotypes except R027 classification with 100% sensitivity. R027 was classified with 80% sensitivity because of larger variances in VOC abundances found in the five samples analysed. All ribotypes had specificity values of over 70% and R027, R005 and R076 showed 100% specificity. As some of the ribotypes have similar VOC emission patterns, a particular ribotype, for example R013, could be wrongly identified as R014/R020 or R002, thus reducing its specificity value.

In order to explore the similarities and differences between the VOC emissions from ribotypes a dendrogram was produced using cluster analysis (see Figure 3). The resulting dendrogram (Figure 3) shows that each ribotype was grouped together and overall ribotype relatedness was observed at <30 on the variance weighted distance between clusters centres (according to the Mahalanobis distance measure used for analysis). According to the dendrogram in Figure 3, R013, R014/R020 and R002 show the largest differences in terms of VOC emissions from the other ribotypes. R107 and R005 ribotypes are closely related to each other. Ribotypes R027, R026, R078, and R087 are more similar when compared with the other ribotypes. This form of analysis provides evidence for how closely these ribotypes are related to each other and may ultimately help to explain why there are differences of metabolic activity in the ribotypes, although this is beyond the scope of the current study. It could also reveal additional metabolic signatures associated with the hypervirulence in strains such as ribotypes R027 and R078.

*3.2 Clostridium difficile ribotype 'blind test'*

A 'blind test' was undertaken to validate the use of VOCs as a marker for *C. difficile* ribotypes. The data from the first analysis in Figure 2 were used as a training set to build the statistical model and then four unknown ribotypes, of which there were five VOC samples collected from each, were used as test set. Figure 4 shows the resulting PCA plot produced by the model. A PLSDA analysis was carried out and the model correctly predicted R107 as being one of the unknowns. The three other ribotypes were more difficult to identify. One was positioned very close to R027 and was tentatively identified as the R027 ribotype. Another ribotype was closely positioned to R087 and R078 and the final one had characteristics similar to R026 and R076. The samples were later confirmed the ribotypes as R027, R107, R087 and R026.

Even though the 'blind' R107 was clustered near R005, the relative intensities of VOCs released by R107 had clear differences that made it possible to easily differentiate it from R005. The R027 data from the 'blind' test clustered near the R027 of the test set, and the relative intensities of VOCs produced closely resembled the intensities of test set R027. The unknown R026 sample was between R026 and R076 in the test set, making it impossible to firmly identify it's ribotype. The same situation applied to R087, as it could only be said to be either R087 or R078 in the test set. So overall the 'blind test' results show that the metabolite VOCs produced by the ribotypes were consistent and in some cases allowed the clear identification of the ribotype. Even in the least well performing cases, where the ribotype could not be identified with 100% certainty, they could be classified into a group of strongly related ribotypes from the PTR-ToF-MS measurements.

*3.3 Metabolite Identification*

Metabolite identification by PTR-ToF-MS is based on the detection and attribution of protonated product ions (neutral analyte molecule plus 1 amu). As PTR-ToF-MS is a direct-MS method it cannot distinguish nominally isobaric species nor structural isomers (Wyche et al. 2005). Consequently, it is not easy to identify the specific VOC(s) responsible for a specific peak in the mass spectrum and it is impossible when two or more compounds have the same (nominal) mass. Identification of compounds can further be complicated by fragmentation and clustering of product ions.

Inspection of Figure 2 shows the masses used as loadings or vectors in the PCA analysis. The masses (or to a first approximation VOC compounds) discriminate the different ribotypes. Masses 30, 46, 61,

64, 70, 74, 88, 94 and 114 had a higher influence on the clustering of R005 and R107 compared to other ribotypes. The peaks at *m/z* 27, 33, 34, 47, 91, 92, 117, and 118 determined the grouping pattern of R027, R078, R087 and R026. R076, a non-pathogenic ribotype, had a combination of masses (*m/z* 27, 29, 33, 40, 41, 47, 51, 58, 103 and 119) that influenced the clustering of this ribotype. The position of these masses away from those seen for other ribotypes could imply that the VOCs responsible may be produced in pathogenic ribotypes at lower concentrations, or indeed may not be produced at all when compared to non-pathogenic R076.

The presence of a characteristic volatile metabolite or a combination of metabolites is attributable to specific metabolic pathways that are active in bacteria (Kai et al. 2009). The compounds released could vary in concentration depending on the growth medium and growth conditions. The chemical identity of these volatiles will allow insight into the underlying metabolic pathways that are active during bacterial growth. However, the number of compounds underlying our statistical analysis was large and therefore we did not attempt to identify them all. To do this further work would be required, such as a GC-MS analysis. Instead here we provide a tentative identification of some of the more significant peaks seen in the mass spectra.

Figure 5 shows the signal intensities of peaks arising from several m/z values for the various ribotypes and for which we provide a tentative compound assignment. One of the most prominent peaks was observed at *m/z* 33 and it seems likely that this derives from protontated methanol. Methanol was observed to be highest in R026, R078, R087 and R076 and was detected at significantly lower levels in R027, R014/R020 and R013. No methanol signals were found for R107 and R005. Little is known about methanol production by *C. difficile*, although Garner et al. reported detection of the compound in 27% of stool samples of patients with *C. difficile* infection compared to 40% and 36% in healthy donors in a cohort and longitudinal study, respectively (Garner et al. 2007).

Another significant peak at *m/z* 46 can be assigned to dimethylamine. Evidence for this is twofold. First, the even mass is consistent with a nitrogen-containing molecule. Second, a previous study has detected dimethylamine in the headspace of *C. difficile* (Pons et al. 1985). In the current study significant quantities of dimethylamine were only produced by R005 and R107, as can be seen in Figure 5. It was observed to be emitted by other ribotypes in the raw data but the intensity in the blank medium was much higher than the detected intensity from the ribotypes.

Earlier studies have reported *p*-cresol emission from *C. difficile* (Berg et al. 1985; Nunez-Montiel et al. 1983; Phillips and Rogers 1981). Another study, which concentrated on the production of end products from the metabolism of aromatic acids of phenylalanine, tyrosine and tryptophan by growing *Clostridia* cultures, found that only *C. difficile* emitted p-cresol and this compound was not detected in 22 other *Clostridia* bacteria (Elsden et al. 1976). More recently, Dawson et al (Dawson et al. 2011) have shown R027 produce more p-cresol that R012 ribotype. Consequently, a signal at *m*/*z* 109 was tentatively assigned to *p*-cresol and was detected in R014/R020, R002 and R013. Again a substantial signal at *m*/*z* 109 from the blank medium made it difficult to identify *p*-cresol from the other ribotype emissions. Further, it is worth noting that *C. difficile* doesn't produce p-cresol in BHIS liquid broth (Dawson et al. 2011).

Signal at *m/z* 61 was assigned to ethylene sulfide and was found in five of the ribotypes, registering particularly high intensities for ribotypes R107 and R005. Ethylene sulfide was previously reported to be released by *Clostridium* species (Rimbault et al. 1986). Signal at *m/z* 63 was tentatively assigned to dimethyl sulfide and was measured with high intensities for R014/R020, R013 and R002. It was found to be about 2 to 3 orders of magnitude lower in R027, R026, R087 and R076. This finding is supported by a previous study that detected dimethyl sulfide released by *Clostridium* species, (Stotzky and Schenck 1976) although it should be noted that Garner et al. did not detect dimethyl sulfide in any stool samples infected with *C. difficile* (Garner et al. 2007). This may suggest that dimethyl sulfide is not emitted by *C. difficile* when growth conditions are changed. Another tentative assignment is that signal at *m/z* 91 derives from methyl thioacetate. This was detected in R027 with a signal intensity over three times larger than for other ribotypes, and indeed the compound was not detected at all in the non-pathogenic ribotype R076. (S)-methylthioacetate was reported to be emitted by *Clostridium* species in headspace microbial using gas chromatography (Rimbault et al. 1986).

**4. Conclusions**

The work described here has two potential impacts. First, it suggests that the detection of emitted VOCs by PTR-ToF-MS may have utility as a rapid means of identifying *C. difficile* infection. Second, the VOCs may be markers for different active metabolic pathways in specific ribotypes.

The rapid detection and identification of *C. difficile* is a primary concern in healthcare facilities and clinical microbiology laboratories. Rapid and accurate diagnoses are important to reduce *C. difficile* infections, as well as to provide the right treatment to infected patients. Delayed treatment and

inappropriate antibiotic regimens not only cause high morbidity and mortality, but also adds costs to the healthcare system through lost bed days (Probert et al. 2004). Current detection methods generally take two to five days to identify *C. difficile*. This study demonstrates that PTR-ToF-MS analysis is capable of detecting VOCs of *C. difficile* metabolites in the headspace of cultures within minutes. Many of the ten *C. difficile* ribotypes were successfully distinguished from one another using the profiles of detected metabolites from the different ribotypes. The results presented here provide a foundation for a *C. difficile* biomarker library that could one day serve as an information base and diagnostic tool in identifying *C. difficile* infections. Our approach may lead to a rapid clinical diagnostic test based on the VOCs released from faecal samples of patients infected with *C.difficile*. There is no doubt there are challenges in sampling and attributing *C. difficile* VOCs from faecal samples.

The presence or non-presence of VOCs in both pathogenic and non-pathogenic ribotypes can also provide information on the different active metabolic pathways existing in various *C. difficile* ribotypes. *Clostridium difficile* is a diverse and variable species that can be sub-classified into hundreds of ribotypes, and also presents significant genetic diversity (Stabler et al, 2012). There is still uncertainly in relating ribotype, or even genotype to strain virulence. The different intensities of the metabolic VOCs released by the *C. difficile* ribotypes may explain the biological functions that occur in each ribotype and how this would affect infectious behaviour. For example, PCR ribotype R027 has found to be the cause of many hospital-related *C. difficile* infections and outbreaks worldwide, so why is R027 far more aggressive than the other ribotypes? Although a genetic basis is predicted for the success of the strain in that it encodes an extra toxin known and has a mutation in a gene that encodes a protein that represses toxin production (Warny et al. 2005), the phenotype of this strain has not been fully characterized. It is non-trivial to go from VOC emissions to answering this type question but data of this nature may ultimately provide useful clues.

**Table 1.** Sensitivity and specificity values of the *C. difficile* ribotype classification by PLSDA

| Cross Validation (leave-one-out) | R027 | R014/ R020 | R002 | R013 | R005 | R107 | R026 | R087 | R078 | R076 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sensitivity % | 80.0 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 80.0 | 100 |
| Specificity % | 100 | 73.3 | 86.7 | 93.3 | 100 | 88.9 | 88.9 | 77.8 | 91.1 | 100 |

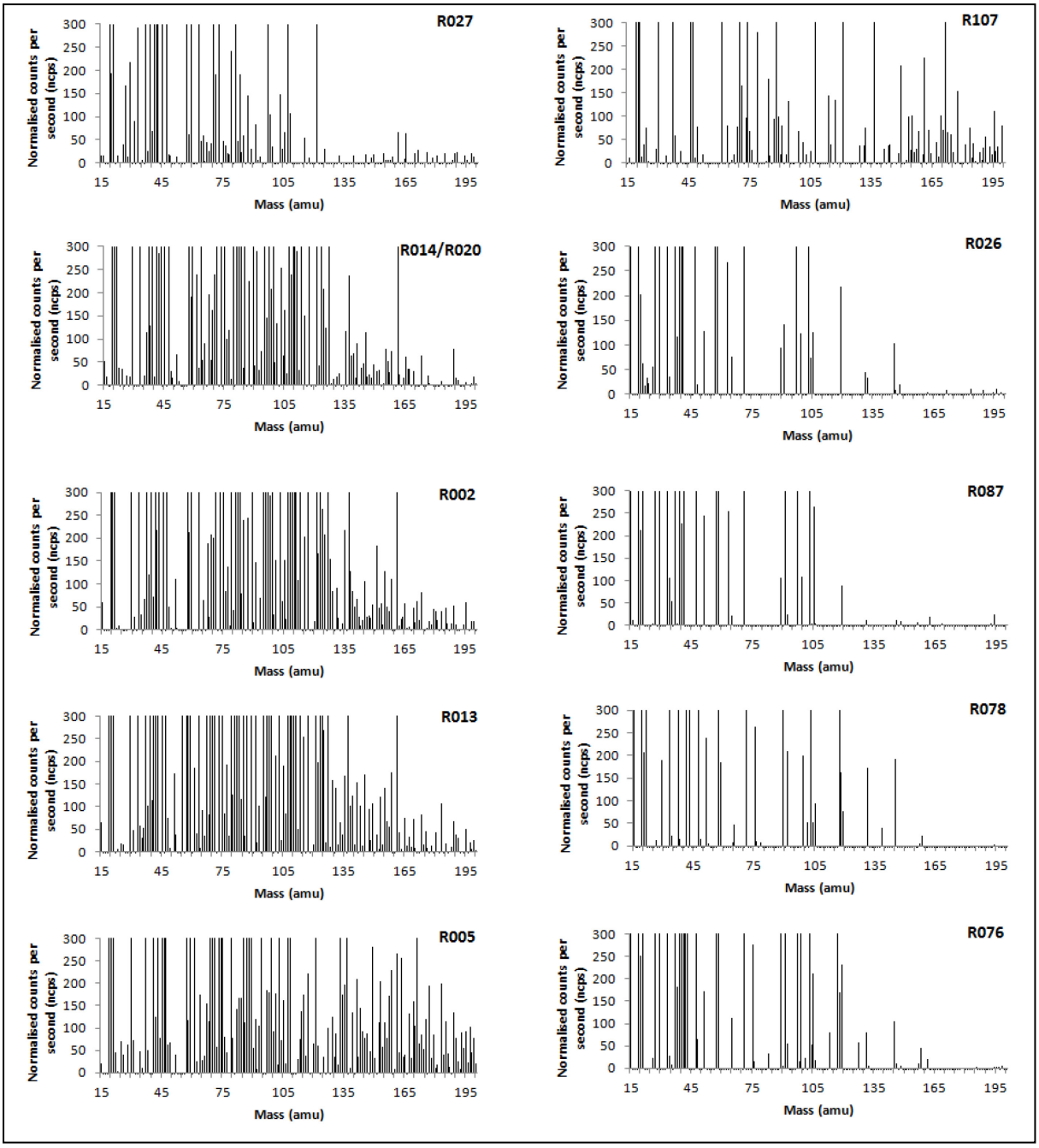


**Figure 1**. Mass spectra (after subtraction of blank) from the VOC headspace analysis of different ribotypes of *C. difficile* cultures. Note that in all of the spectra several peaks have count rates well above the maximum shown but have been displayed in this way in order to shown some of the more subtle details in the profiles of less abundant VOCs.

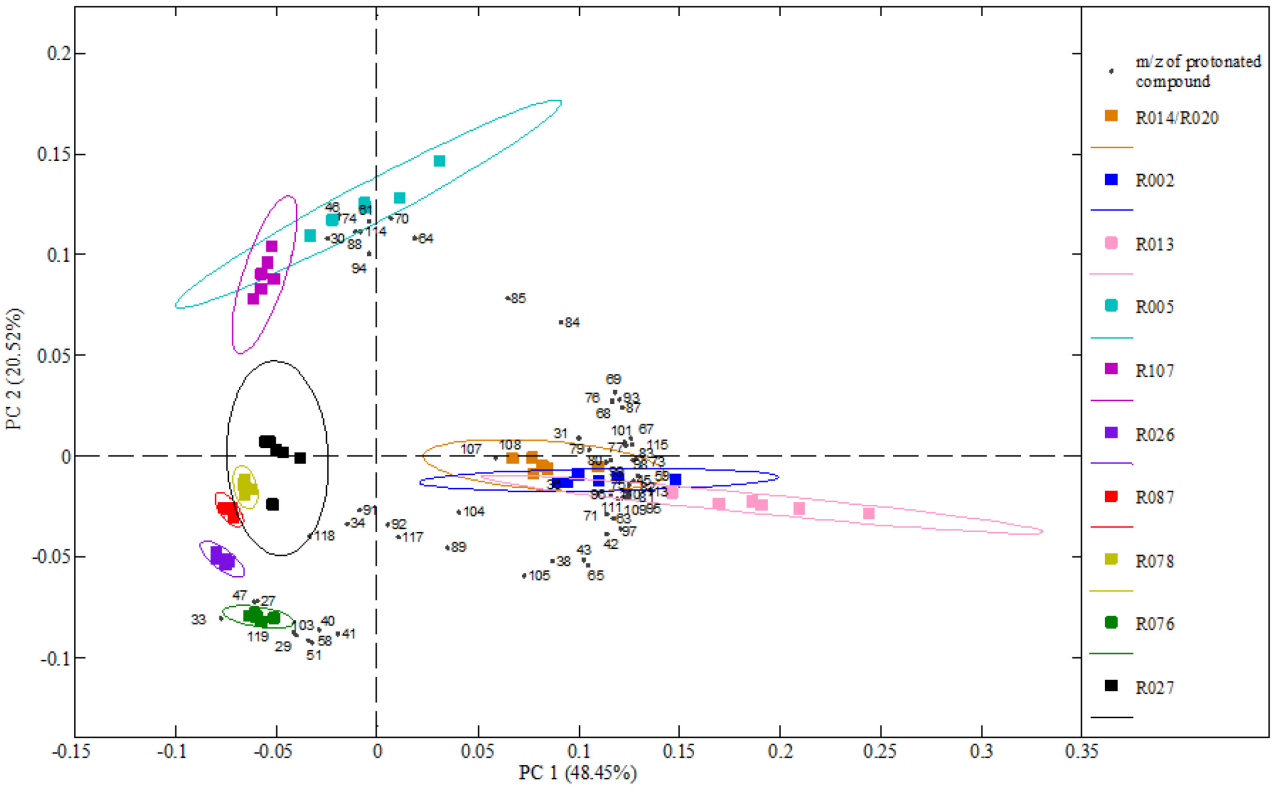


**Figure 2**. A PCA biplot for the different *C. difficile* ribotypes. The first principal component has been plotted against the second principal component. The PCA was performed using peaks at 69 distinct *m/z* values. The oval lines surrounding each class depict a 95% confidence level.

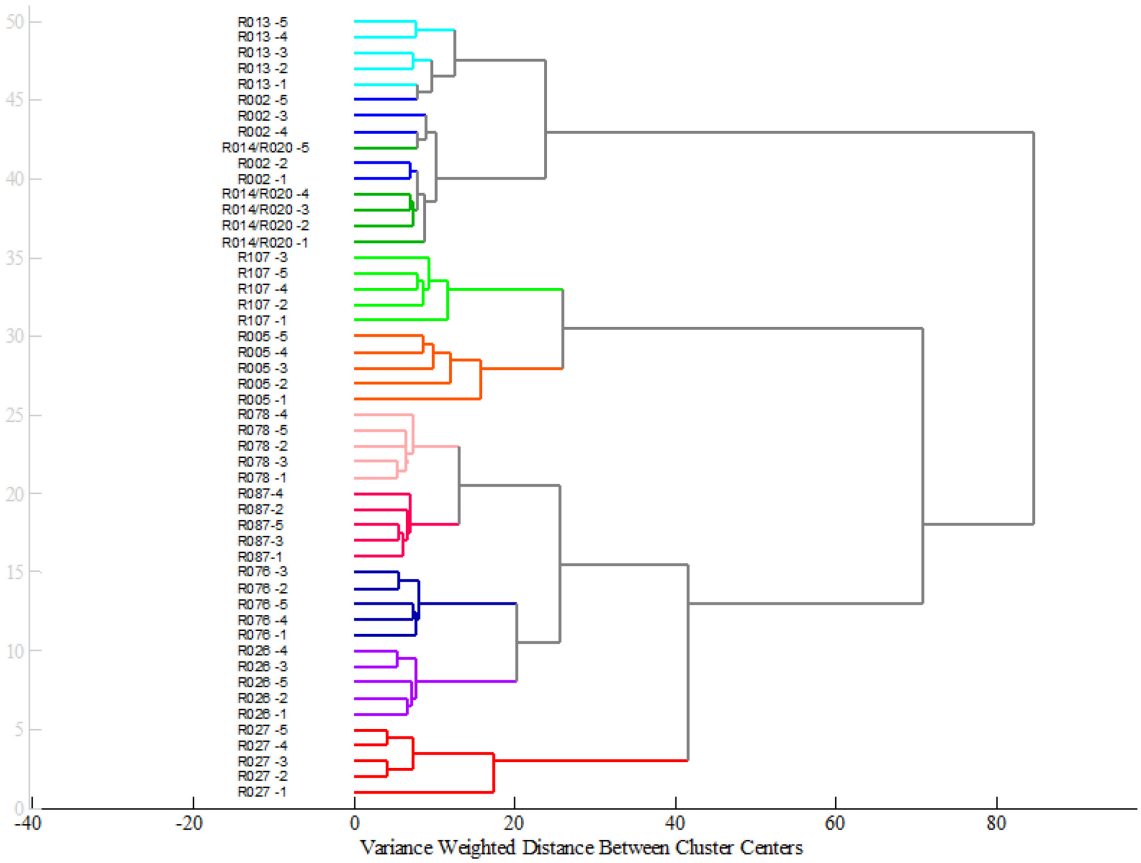


**Figure 3**. Dendrogram of *C. difficile* ribotypes produced by cluster analysis (using the Mahalanobis distance) according to the 69 mass peaks selected from a Mann-Whitney test.

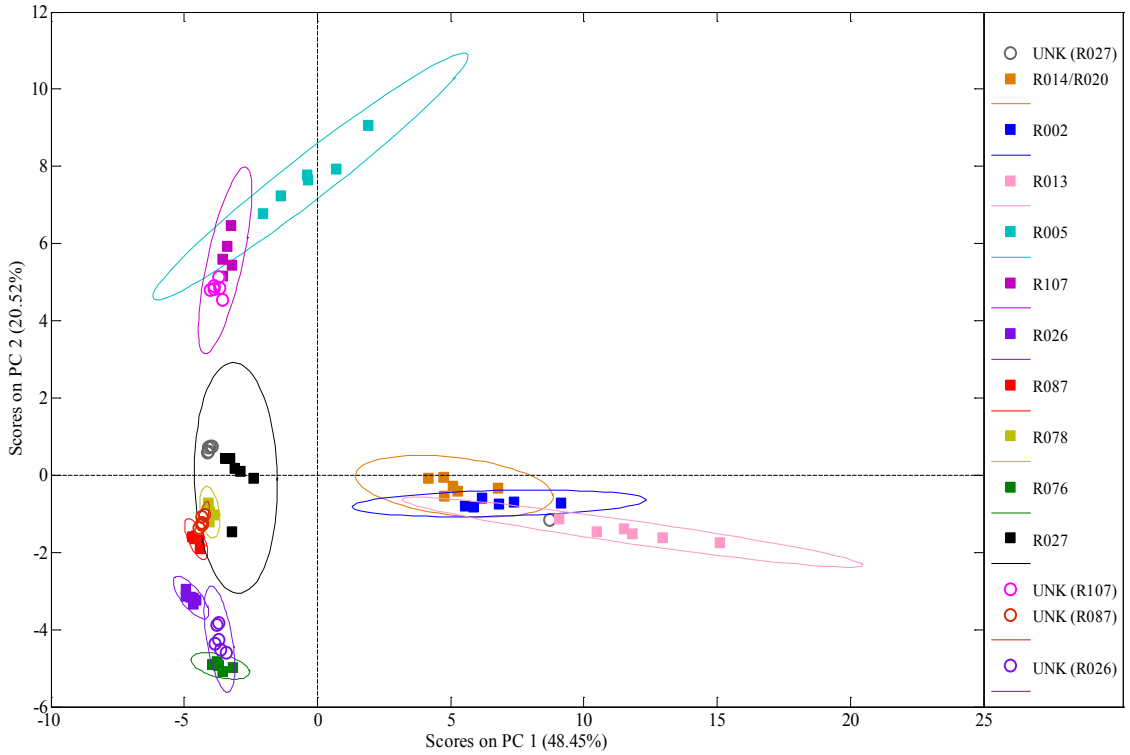


**Figure 4**. A PCA biplot for the blind test of *C. difficile* ribotypes. The first principal component has been plotted against the second principal component. The previous ribotype analysis was loaded as a validation and the 'blind test' group as a test group. Cross validation using the leave-one-out procedure was used. *UNK = unknown.

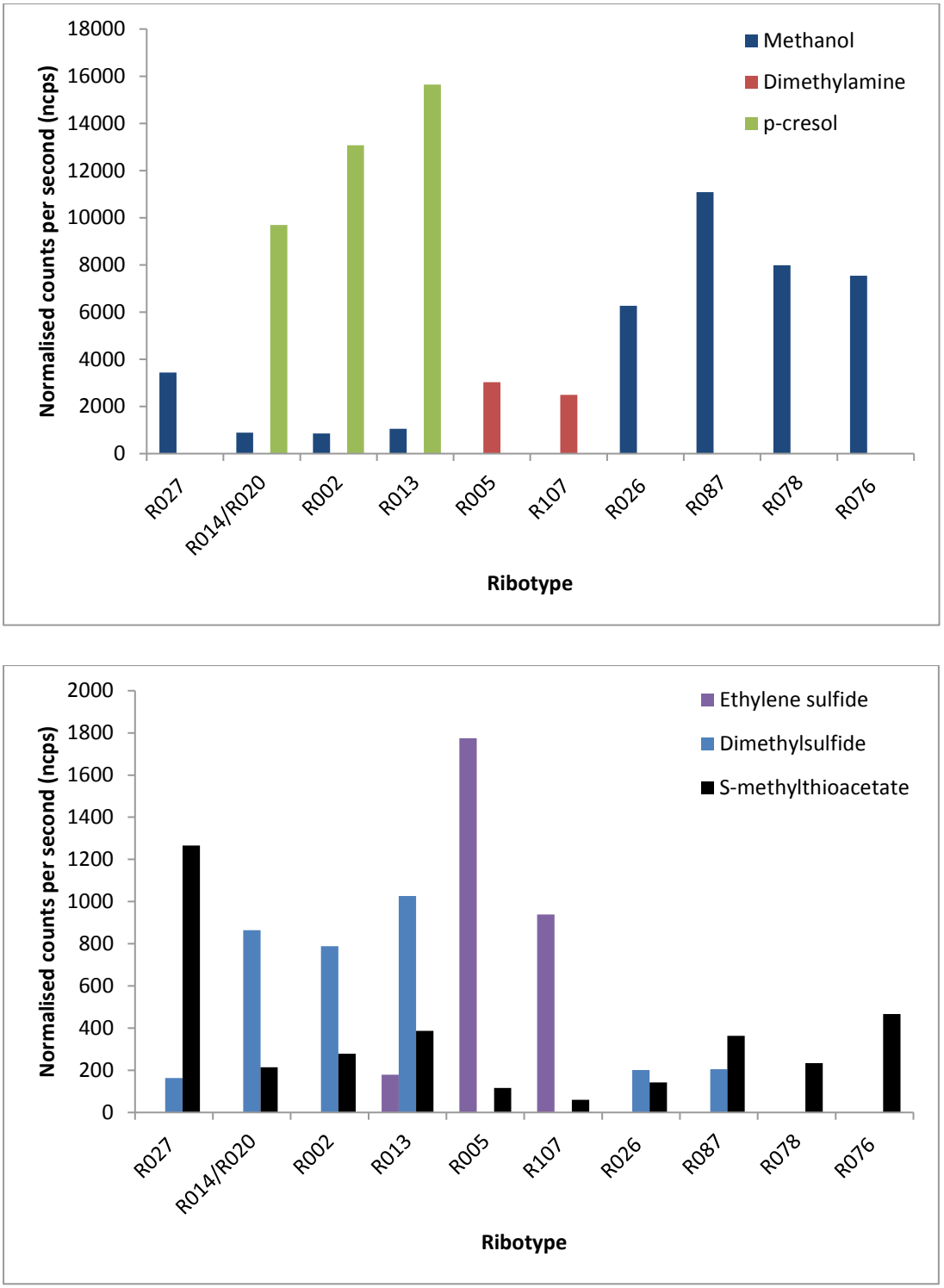


**Figure 5**. Signal intensity chart of *C. difficile* ribotypes for selected mass peaks. The measured signal levels have been subjected to subtraction from the blank spectrum. Tentative assignments of the selected peaks are for protonated versions of the following compounds: *m/z* 33 = methanol; *m/z* 46 = dimethylamine; *m/z* 61 = ethylene sulfide; *m/z* 63 = dimethyl sulfide; *m/z* 91 = methyl thioacetate; *m/z* 109 = *p*-cresol.

**References**


Barber, S., et al. (2012). Increased Sensitivity in Proton Transfer Reaction Mass Spectrometry by Incorporation of a Radio Frequency Ion Funnel. Analytical Chemistry 84, 5387-5391 doi:10.1021/ac300894t

Bartlett, J. G. (1990). Clostridium-difficile - Clinical considerations. Reviews of Infectious Diseases 12, S243-S251

Bauer, M. P., et al. (2011). Clostridium difficile infection in Europe: a hospital-based survey. Lancet 377, 63-73 doi:10.1016/s0140-6736(10)61266-4

Berg, J. D., R. G. Mills, D. J. Coleman (1985). Improved Gas-Liquid-Chromatography Method for the indentification of clostridium-difficile. Journal of Clinical Pathology 38, 108-110 doi:10.1136/jcp.38.1.108

Blake, R. S., P. S. Monks, A. M. Ellis (2009). Proton-Transfer Reaction Mass Spectrometry. Chemical Reviews 109, 861-896 doi:Doi 10.1021/Cr800364q

Blake, R. S., C. Whyte, C. O. Hughes, A. M. Ellis, P. S. Monks (2004). Demonstration of proton-transfer reaction time-of-flight mass spectrometry for real-time analysis of trace volatile organic compounds. Analytical Chemistry 76, 3841-3845 doi:Doi 10.1021/Ac0498260

Bruins, M., et al. (2009). Device-independent, real-time identification of bacterial pathogens with a metal oxide-based olfactory sensor. European Journal of Clinical Microbiology & Infectious Diseases 28, 775-780 doi:10.1007/s10096-009-0700-1

Bunge, M., et al. (2008). On-line monitoring of microbial volatile metabolites by proton transfer reaction-mass spectrometry. Applied and Environmental Microbiology 74, 2179-2186 doi:10.1128/aem.02069-07

Burns, D. A., N. P. Minton (2011). Sporulation studies in Clostridium difficile. Journal of Microbiological Methods 87, 133-138 doi:10.1016/j.mimet.2011.07.017

Cheknis, A. K., et al. (2009). Distribution of Clostridium difficile strains from a North American, European and Australian trial of treatment for C. difficile infections: 2005-2007. Anaerobe 15, 230-233 doi:10.1016/j.anaerobe.2009.09.001

Clements, A. C. A., R. J. S. Magalhaes, A. J. Tatem, D. L. Paterson, T. V. Riley (2010). Clostridium difficile PCR ribotype 027: assessing the risks of further worldwide spread. Lancet Infectious Diseases 10, 395-404

Dawson, L. F., et al. (2011). The analysis of para-cresol production and tolerance in Clostridium difficile 027 and 012 strains. BMC Microbiol 11, 86 doi:10.1186/1471-2180-11-86

Elsden, S. R., M. G. Hilton, J. M. Waller (1976). End products of metabolism of aromatic amino-acids by clostridia Archives of Microbiology 107, 283-288 doi:10.1007/bf00425340

Freeman, J., et al. (2010). The Changing Epidemiology of Clostridium difficile Infections. Clinical Microbiology Reviews 23, 529-+ doi:10.1128/cmr.00082-09

Garner, C. E., et al. (2007). Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. Faseb Journal 21, 1675-1688 doi:10.1096/fj.06-6927com

Grein, J. D., M. Ochner, H. Hoang, A. Jin, M. A. Morgan, A. R. Murthy (2014). Comparison of testing approaches for Clostridium difficile infection at a large community hospital. Clinical Microbiology and Infection 20, 65-69 doi:10.1111/1469-0691.12198

Hargreaves, K. R., H. V. Colvin, K. V. Patel, J. J. P. Clokie, M. R. J. Clokie (2013). Genetically Diverse Clostridium difficile Strains Harboring Abundant Prophages in an Estuarine Environment. Applied and Environmental Microbiology 79, 6236-6243 doi:10.1128/aem.01849-13

Indra, A., et al. (2008). Characterization of Clostridium difficile isolates using capillary gel electrophoresis-based PCR ribotyping. Journal of Medical Microbiology 57, 1377-1382 doi:10.1099/jmm.0.47714-0

Kai, M., M. Haustein, F. Molina, A. Petri, B. Scholz, B. Piechulla (2009). Bacterial volatiles and their action potential. Applied Microbiology and Biotechnology 81, 1001-1012 doi:10.1007/s00253-008-1760-3

Knoop, F. C., M. Owens, I. C. Crocker (1993). Clostridium-Difficile - Clinical Disease and Diagnosis. Clinical Microbiology Reviews 6, 251-265

Labbe, A. C., et al. (2008). Clostridium difficile infections in a Canadian tertiary care hospital before and during a regional epidemic associated with the BI/NAP1/027 strain. Antimicrobial Agents and Chemotherapy 52, 3180-3187 doi:10.1128/aac.00146-08

Nale, J. Y., J. Y. Shan, P. T. Hickenbotham, W. N. Fawley, M. H. Wilcox, M. R. J. Clokie (2012). Diverse Temperate Bacteriophage Carriage in Clostridium difficile 027 Strains. Plos One 7, doi:10.1371/journal.pone.0037263

Nunez-Montiel, O. L., F. S. Thompson, V. R. Dowell (1983). Norleucine-Tyrosine broth for rapid indetification of clostridium difficile by gas-liquid-chromatography. Journal of Clinical Microbiology 17, 382-385

Perl, T., et al. (2011). Detection of characteristic metabolites of Aspergillus fumigatus and Candida species using ion mobility spectrometry - metabolic profiling by volatile organic compounds. Mycoses 54, E828-E837 doi:10.1111/j.1439-0507.2011.02037.x

Phillips, K. D., P. A. Rogers (1981). Rapid detection and presumptive indetification of Clostridium-difficile by p-cresol production of a selective medium. Journal of Clinical Pathology 34, 642-644 doi:10.1136/jcp.34.6.642

Pons, J. L., A. Rimbault, J. C. Darbord, G. Leluan (1985). Gas-Chromatographic mass-spectrometric analysis of volatile amines produced by several strains of Clostridium. Journal of Chromatography 337, 213-221 doi:10.1016/0378-4347(85)80034-7

Probert, C. S. J. (2011). Role of faecal gas analysis for the diagnosis of IBD. Biochemical Society Transactions 39, 1079-1080 doi:10.1042/bst0391079

Probert, C. S. J., P. R. H. Jones, N. M. Ratcliffe (2004). A novel method for rapidly diagnosing the causes of diarrhoea. Gut 53, 58-61 doi:10.1136/gut.53.1.58

Rimbault, A., P. Niel, J. C. Darbord, G. Leluan (1986). Headspace gas-chromatographic mass-spectrometric analysis of light-hydrocarbons and voaltile organosulfur compounds in reduced-pressure cultures of Clostridium. Journal of Chromatography 375, 11-25 doi:10.1016/s0378-4347(00)83687-7

Scotter, J. M., R. A. Allardyce, V. Langford, A. Hill, D. R. Murdoch (2006). The rapid evaluation of bacterial growth in blood cultures by selected ion flow tube-mass spectrometry (SIFT-MS) and comparison with the BacT/ALERT automated blood culture system. Journal of Microbiological Methods 65, 628-631 doi:10.1016/j.mimet.2005.09.016

Settle, C., K. G. Kerr (2011). Diarrhoea after broad spectrum antimicrobials. BMJ (Clinical research ed) 342, d3798-d3798 doi:10.1136/bmj.d3798

Stabler, R. A., et al. (2012). Macro and Micro Diversity of Clostridium difficile Isolates from Diverse Sources and Geographical Locations. Plos One 7, doi:10.1371/journal.pone.0031559

Stotzky, G., S. Schenck (1976). Volatile organic compounds and microorganisms. CRC critical reviews in microbiology 4, 333-82 doi:10.3109/10408417609102303

Tait, E., J. D. Perry, S. P. Stanforth, J. R. Dean (2014). Use of volatile compounds as a diagnostic tool for the detection of pathogenic bacteria. Trac-Trends in Analytical Chemistry 53, 117-125 doi:10.1016/j.trac.2013.08.011

Thorn, R. M. S., D. M. Reynolds, J. Greenman (2011). Multivariate analysis of bacterial volatile compound profiles for discrimination between selected species and strains in vitro. Journal of Microbiological Methods 84, 258-264 doi:10.1016/j.mimet.2010.12.001

Varmuza, K., P. Filzmoser (2009). Introduction to multivariate statistical analysis in chemometrics, CRC Press.

Warny, M., et al. (2005). Toxin production by an emerging strain of Clostridium difficile associated with outbreaks of severe disease in North America and Europe. The Lancet 366, 1079-1084

White, I. R., et al. (2013). Real-time multi-marker measurement of organic compounds in human breath: towards fingerprinting breath. Journal of Breath Research 7, 017112 doi:doi:10.1088/1752-7155/7/1/017112

Wiegand, P. N., D. Nathwani, M. H. Wilcox, J. Stephens, A. Shelbaya, S. Haider (2012). Clinical and economic burden of Clostridium difficile infection in Europe: a systematic review of healthcare-facility-acquired infection. Journal of Hospital Infection 81, 1-14 doi:10.1016/j.jhin.2012.02.004

Wilcox, M. H., et al. (2012). Changing Epidemiology of Clostridium difficile Infection Following the Introduction of a National Ribotyping-Based Surveillance Scheme in England. Clinical Infectious Diseases 55, 1056-1063 doi:10.1093/cid/cis614

Wyche, K. P., et al. (2007). Technical note: Performance of Chemical Ionization Reaction Time-of-Flight Mass Spectrometry (CIR-TOF-MS) for the measurement of atmospherically significant oxygenated volatile organic compounds. Atmospheric Chemistry and Physics 7, 609-620

Wyche, K. P., R. S. Blake, K. A. Willis, P. S. Monks, A. M. Ellis (2005). Differentiation of isobaric compounds using chemical ionization reaction mass spectrometry. Rapid Communications in Mass Spectrometry 19, 3356-3362 doi:Doi 10.1002/Rcm.2202