

Combination of 16S rRNA variable regions provides a detailed analysis of bacterial community dynamics in the lungs of cystic fibrosis patients

Melissa S. Doud,¹ Michael Light,² Gisela Gonzalez,³ Giri Narasimhan³ and Kalai Mathee^{4*}

¹Department of Biological Sciences, College of Arts and Sciences, Florida International University, Miami, FL 33199, USA

²Department of Medicine, University of Miami, FL 33101, USA

³Bioinformatics Research Group (BioRG), School of Computing and Information Sciences, College of Engineering, Florida International University, Miami, FL 33199, USA

⁴Department of Molecular Microbiology and Infectious Diseases, College of Medicine, Florida International University, Miami, FL 33199, USA

*Correspondence to: Tel: +1 305 348 0629; Fax: +1 305 348 2193; E-mail: kalai.mathee@fiu.edu

Date received (in revised form): 13th November, 2009

Abstract

Chronic bronchopulmonary bacterial infections remain the most common cause of morbidity and mortality among patients with cystic fibrosis (CF). Recent community sequencing work has now shown that the bacterial community in the CF lung is polymicrobial. Identifying bacteria in the CF lung through sequencing can be costly and is not practical for many laboratories. Molecular techniques such as terminal restriction fragment length polymorphism or amplicon length heterogeneity–polymerase chain reaction (LH-PCR) can provide many laboratories with the ability to study CF bacterial communities without costly sequencing. The aim of this study was to determine if the use of LH-PCR with multiple hypervariable regions of the 16S rRNA gene could be used to identify organisms found in sputum DNA. This work also determined if LH-PCR could be used to observe the dynamics of lung infections over a period of time. Nineteen samples were analysed with the VI and the VI_V2 region of the 16S rRNA gene. Based on the amplicon size present in the VI_V2 region, *Pseudomonas aeruginosa* was confirmed to be in all 19 samples obtained from the patients. The VI region provided a higher power of discrimination between bacterial profiles of patients. Both regions were able to identify trends in the bacterial population over a period of time. LH profiles showed that the CF lung community is dynamic and that changes in the community may in part be driven by the patient's antibiotic treatment. LH-PCR is a tool that is well suited for studying bacterial communities and their dynamics.

Keywords: LH-PCR, 16S rRNA, VI, VI_V2, cystic fibrosis, AmpliQué, cystic fibrosis microbiome, metagenome

Introduction

Cystic fibrosis (CF) is an autorecessive disease affecting one in 3,500 Caucasian live births in the USA.¹ CF results from a mutation in the gene that encodes the CF transmembrane conductance regulator (CFTR) protein.^{2,3} A defect in the CFTR

protein leads to a malfunctioning cyclic AMP-activated chloride channel in secretory epithelia.⁴ This defect in the lung leads to the inability to secrete chloride and to the excess re-absorption of sodium.⁴ Thus, there is decreased fluid secretion and the mucus becomes immobilised

and adheres to the epithelial cells. Overproduction of mucus in the airway results in congestion of the respiratory tract and increases susceptibility to bronchopulmonary infection. CF patients often suffer from infections with *Staphylococcus aureus*, *Haemophilus influenzae*, *Pseudomonas aeruginosa* and *Burkholderia cenocepacia*.⁵ These chronic infections by highly adapted lung microbes cause inflammation and, eventually, lung damage; therefore, lung disease is the major cause of morbidity and mortality among these patients.^{6,7}

It has also been estimated that less than 1 per cent of eubacteria in the environment can be cultured.^{8,9} Thus, these identification methods will fail to detect all pathogens that might be causing the lung infection. Fortunately, with the advent of molecular techniques, culturing for identification purposes can be circumvented.¹⁰ Recent molecular studies using terminal restriction fragment length analysis and sequencing have shown that the lung community is complex. *Achromobacter (Alcaligenes) xylosoxidans*, *Rhodotorula mucilaginosa*, *Abiotrophia* spp, *Bacteroides gracilis*, *Eubacterium brachy*, *Mycobacterium mucilaginosus*, *Mycoplasma salivarium*, *Porphyromonas salivae*, *Ralstonia* spp, *Staphylococcus hominis*, *Streptococcus anginosus*, *Treponema vincentii*, *Veillonella* spp, *Burkholderia gladioli*, *Stenotrophomonas maltophilia* and *Pandoraea* atypical have all been demonstrated to be members of this polymicrobial infection.^{11–18}

Amplicon length heterogeneity–polymerase chain reaction (LH-PCR) is another molecular method that has been used to study various diverse microbial communities,^{19,20} including those of CF sputum.¹⁸ LH-PCR utilises the 16S rRNA molecular marker to analyse microbial populations.^{20–22} The sequence and length of the variable regions within the 16S rRNA gene are often used to determine phylogeny.²³ It should be noted that some bacteria have multiple copies of the 16S rRNA gene, which can show intragenomic heterogeneity in sequence and gene length.²³ This variation is unlikely to have an effect on classification at the genus and species level, and the marker is routinely used for identification.²⁴ Identification based on length variation in the hypervariable regions of the

gene is exploited by LH-PCR. In this technique, forward and reverse primers bind to the conserved regions of the 16S rRNA gene and then amplify the varying lengths seen in the hypervariable regions. LH-PCR has proven to be a robust, time-efficient and reproducible method.^{20,21,25} The main advantages of LH-PCR are that it surveys relative gene frequencies within complex mixtures of DNA, is reproducible, requires small sample sizes and can be performed with many samples simultaneously.^{20,25} Furthermore, some of the size classes emerging from LH-PCR analyses can be related at the genus level to archived database sequences.²⁰ Overall, the attributes of LH-PCR make it useful for quick assessment of the diversity of microbial communities for comparative purposes.²⁰

Prior studies have shown that LH-PCR and the hypervariable regions 1 and 2 (V1_V2) of the 16S rRNA gene can be used to detect eubacteria in sputum from CF patients.^{18,26} The technique had limited efficacy when determining overall diversity and attaining organism identification, and it was concluded that the small differences in amplicon lengths generated from the organisms were at fault.¹⁸ In this study, community profiles of CF sputa were obtained using two regions of the 16S rRNA gene (hypervariable region 1 [V1] and V1_V2). Bacterial identification based on amplicon length was attempted using a computer program, AmpliQué, which determined the theoretical V1 and V1_V2 fragment lengths for known organisms. Community profiles and identification were analysed using each region separately and combined to determine if a combinatorial approach to LH-PCR: 1) could provide a more detailed analysis of the bacterial community in the CF lung and 2) could be used to observe the dynamics of this community over a period of time. Our work has shown that the V1 region provides a more detailed community profile, which leads to a higher power of discrimination between profiles than the V1_V2 region. The LH analyses provided snapshots of the CF bacterial communities; however, the analyses are limited by their ability to identify community members with the V1 or V1_V2 regions separately or combined. In addition, through the comparisons

of V1_V2 or V1 LH profiles over time, it was established that the CF lung community is not only diverse, but also dynamic.

Materials and methods

Sample collection

Nineteen sputum samples were obtained from CF patients attending the University of Miami Adult CF Clinic (Miami, FL). Samples are referred to by UM followed by a letter. The study was carried out in accordance with the Declaration of Helsinki (2000) of the World Medical Association and was approved by the appropriate institutional review board (FIU IRB approval # 033004-02). Appropriate consent was obtained from human subjects. All 19 expectorated sputum samples were frozen at -20°C prior to DNA extraction. Sputum was liquefied using the method of Reischl *et al.*²⁷ and the metagenomic DNA was extracted and cleaned using the GeneClean Spin Kit (Qbiogene, Irvine, CA, USA) according to the manufacturer's protocol. The eluted DNA was quantified and diluted to 10 ng/ μl and refrigerated at 4°C until further use.

Control bacteria isolates

Standard DNA manipulations were used to extract chromosomal DNA from CF-related pathogens: *P. aeruginosa* strains PAO1 and PA14 (burn wound isolates) and four additional *P. aeruginosa* strains isolated from CF,²⁸ *B. cenocepacia* (ATCC BAA-246; CF isolate), *S. aureus* (ATCC 12600; pleural fluid isolate) and *S. maltophilia* (ATCC 17672; sputum isolate), were used as controls in the LH-PCR experiments.

LH-PCR

The metagenomic DNA derived from the extracted sputum samples were amplified in triplicate using eubacterial primers for the 16S rRNA gene. The V1_V2 region of the 16S rRNA gene were amplified using 27F-6FAM forward primer (5' – 6FAM – AGA GTT TGA TCM TGG – 3') and the 355R reverse primer (5' – GCT GCC TCC CGT AGG AGT – 3').²⁹ The V1 region was amplified using

fluorescent primer P1F-6FAM (5' – 6FAM – GCG GCG TGC CTA ATA CAT GC – 3') and reverse primer P2R (5' – TTC CCC ACG CGT TAC TCA CC – 3').^{29,30} All forward primers were fluorescently labelled with 6-FAMTM (Integrated DNA Technologies, Shokie, IL, USA). The final V1_V2 PCR reaction mixture was composed of: 1X PCR buffer (supplied with the enzyme by Applied Biosystems, Foster City, CA, USA), 2.5 mM MgCl_2 , 0.25 mM dNTPs, 0.5 μM forward and reverse primers, 0.1 per cent bovine serum albumin (BSA), 0.025 U of AmpliTaq Gold LD DNA polymeraseTM (Applied Biosystems) and diethylpyrocarbonate (DEPC) water for a final volume of 20 μl . A half a nanogram per microlitre of DNA template was added per reaction. The following parameters were used to amplify the selected fragments on an MJ Research Peltier Thermal Cycler 200 (Waltham, MA, USA): initial denaturation at 95°C for 11 minutes, 25 cycles of denaturation at 95°C , annealing at 55°C and extension at 72°C , each for 1 minute, and a final elongation at 72°C for 10 minutes. A negative control containing water was amplified for every PCR master mix to check for contamination. Agarose gel electrophoresis was run to confirm the success of amplification.

LH-PCR analysis

LH-PCR analysis was performed by the Forensic DNA Profiling Facility (Florida International University, Miami, FL, USA). The PCR products for the three replicate reactions were detected using the ABI Prism 310 Genetic Analyzer (Applied Biosystems). A formamide-size standard mix was prepared using a 96:1 ratio of Hi-DiTM (highly deionised) formamide and GeneScanTM 500 ROXTM internal standard (Applied Biosystems). Each PCR product was denatured by adding 0.50 μl to 9.5 μl of the formamide-size standard mix for 2 minutes at 95°C and snap cooling for 5 minutes on ice. Each sample was run for 28 minutes on the ABI Prism 310. Fragments were separated by capillary electrophoresis using polymer POP-4, matrix DS-3O_6FAM_HEX_NED_ROX and filter D (Applied Biosystems).

Electropherogram analysis

The output was collected and analysed using GeneMapper® Software Version 3.7 (Applied Biosystems) and Microsoft Excel (Microsoft, Redmond, WA, USA). The ABI Prism™ Genotyper software analysis parameters were set to the local Southern Size calling, no correction, and the minimum noise threshold was set at 70 fluorescent units.²¹ For V1_V2, amplicons were called if between 300 and 400 base pairs (bp). Amplicons for V1 region were between 60 and 120 bp.

Statistical analysis

Data were imported into MS Excel (Microsoft), to determine mean relative ratios for each triplicate PCR reaction. The normalised data for each sample were exported into PRIMER 5 statistical software for further analyses (PRIMER E Ltd., Plymouth Marine Laboratory, Plymouth, UK). The Bray–Curtis similarity index and non-metric multi-dimensional scaling (MDS) were used to compare sample data between patients and patient profiles over time.^{31,32}

Presumptive identity analysis

GeneScan™ 500 ROX™ internal standard was used to determine amplicons sites in bp. Several approaches were used presumptively to identify an organism in the CF community based upon the length of the amplicon generated. Chromosomal DNA from isolates of *B. cenocepacia*, *Staphylococcus aureus*, *Stenotrophomonas maltophilia* and *P. aeruginosa* were used to determine the experimental amplicon length for the V1_V2 and V1 regions. Next, *in silico* analysis was performed, in which the expected amplicon length was determined based on where the primers of interest would theoretically bind to the published sequence. Lastly, the lengths of the expected amplicons for all bacteria were determined using AmpliQué, a newly-designed bioinformatics program.^{29,33} The Perl scripts are available at <http://biorg.cs.fiu.edu/AmpliQue/>. AmpliQué uses the 16S rRNA sequences downloaded from the Ribosomal

Database Project II (Release 9.60). It takes a forward and a reverse primer (both of which may be potentially degenerate) as input. The stringency of the primer binding can be set in the program to control the exactness of the primer binding site within the 16S rRNA sequences. The stringency was first set to 100 per cent ($E = 10$) to identify the organisms that have exactly the same conserved sequences as the primer and subsequently lowered to $E = 5,000$, which allowed a larger database to be created as the conserved regions of the 16S rRNA gene are not 100 per cent identical in sequence. The resulting output shows the identity of all the 16S rRNA sequences that will produce an amplicon with a particular pair of primers, along with the length(s) of the amplicons. Thus, one can query the database using any primer to determine the theoretical fragment lengths that would be generated from an organism when analysed. In addition, a list of all bacteria that produce a specific fragment length can be generated. One amplicon length may represent more than one bacterium, and thus only a presumptive identity of that amplicon can be achieved.

Results

PCR amplification of 16S rRNA genes for metagenome profiling

Variable regions V1 and V1_V2 were successfully amplified using universal eubacterial primers from the metagenomic DNA samples (see Methods). As expected, the amplicon sizes from the 19 samples ranged from 335 to 365 bp for the V1_V2 region (Table 1 and Supplementary Figure 1). Sample UMC had the most complex V1_V2 profile, with eight amplicons (Figure 1a and Table 1). Samples UMF, UMG, UMH, UMJ, UMK, UMP and UMT had the least complex LH profiles, with only one amplicon present at 342 bp (Figure 1b and Table 1). The remaining 11 samples produced LH profiles containing multiple amplicons (Table 1 and Supplementary Figure 1). Amplicon 342 was observed in all 19 samples (Table 1). Out of these 19 samples, 16 samples (84.2 per cent) had amplicon 342 as the dominant member of the profile.

Table 1: Amplicons seen in the VI and VI_V2 regions for samples.

Sample	VI_V2 (%)*†	Total # VI_V2	VI†	Total # VI
UMA	342, 349 (39.0) , 350	3	60, 62, 63, 65, 68, 69, 70, 72, 73, 75, 76, 77, 78, 81, 82, 83, 86 (17.3) , 90, 91, 92, 93, 94, 96, 98, 100, 101, 108	27
UMB	342 (61.7) , 347, 365	3	62, 63, 72, 73, 75, 76, 77, 78, 81, 82, 88 (15.7) , 91, 93, 94, 98, 100, 102, 103, 112	19
UMC	335, 342 (79.0) , 347, 349, 350, 351, 352, 365	8	61, 62, 68, 69, 70, 72, 73, 75, 76, 77, 78, 81, 82, 84, 86, 88 (12.0) , 90, 91, 92, 93, 94, 96, 98, 100, 102, 104	26
UMD	342 (93.7) , 351, 352, 365	4	64 (14.4) , 70, 73, 77, 82, 86, 88, 98, 103, 112	10
UME	342 (99.0) , 348	2	60, 62, 64, 66, 67, 68, 69, 71, 72 (12.3) , 73, 74, 75, 76, 77, 78, 80, 81, 86, 88, 91, 93, 94, 98, 100, 102, 103, 112, 113, 115, 120	30
UMF	342 (100)	1	60, 61, 62, 65, 68, 70, 72 (12.6%) , 73, 75, 77, 78, 85, 91, 93, 94, 98, 103, 112, 113, 115	20
UMG	342 (100)	1	60, 61, 62, 68, 72 (3.6) , 73, 75, 77, 78, 86, 91, 93, 98, 103, 112, 113, 115, 119, 120	19
UMH	342 (100)	1	60, 61, 62, 65, 68, 70, 73, 75, 76, 77 (3.9) , 78, 80, 81, 84, 85, 86, 88, 90, 91, 93, 94, 96, 98, 101, 103, 111, 112, 113, 115	29
UMJ	342 (100)	1	60, 61, 62, 63, 65, 68, 70, 72, 73, 75, 77, 78, 80, 81, 84, 86, 88, 91, 92, 93, 94, 98, 102 (6.6) , 112, 113, 115, 120	27
UMK	342 (100)	1	60, 61, 62, 65, 67, 68, 70, 72, 73, 75, 77, 78, 81, 85, 86, 90, 91, 93, 94, 98, 100, 101 (7.3) , 103, 111, 112, 113, 120	27
UML	342, 350, 360 (55.7) , 365	4	60, 61, 62, 63, 64, 65, 67, 68, 70, 71, 72, 73, 76, 78, 81, 85, 88, 91, 94 (31.3) , 98, 100, 102, 103, 112, 113, 115, 119, 120	28
UMM	342 (98.7) , 351, 365	3	61, 62, 65, 68, 70, 73, 75, 76, 77, 78, 81, 82, 84, 86, 88, 91, 94, 98, 101 (4.8) , 103, 111, 113, 115, 119, 120	25
UMN	339, 342 (92.7) , 347, 352	4	60, 61, 62, 65, 68, 70 (10.2) , 73, 77, 78, 81, 85, 86, 88, 90, 91, 92, 94, 98, 100, 103, 108, 112, 113, 115	24
UMO	342 (74.0) , 350, 352	3	60, 62, 65, 73, 77, 78, 80, 81, 85, 91, 98, 102 (3.1) , 103, 108, 111, 112, 113, 115, 119, 120	20
UMP	342 (100)	1	60, 61, 62, 65, 77, 78, 81, 98, 102 (5.7) , 112, 115, 120	12

Continued

Table 1: Continued

Sample	VI_V2 (%)*†	Total # VI_V2	VI†	Total # VI
UMQ	342 (41.3) , 348, 350, 365	4	60, 61, 62, 63, 64, 65, 71, 72, 76, 78, 81, 84, 85, 87, 91, 94, 98, 102 (5.3) , 103, 112, 115	21
UMR	342, 360 (63.7)	2	60, 62, 63, 65, 73, 77, 78, 80, 81, 82, 85, 94 (13.9) , 98, 102, 108, 112, 115, 120	18
UMS	339, 342 (87.7) , 347	3	60, 62, 63, 66, 72, 73, 74, 77, 78, 80, 81, 85, 88, 91, 94, 98, 102 (9.9) , 103, 111, 112, 113, 115, 119, 120	24
UMT	342 (100)	1	60, 62, 63, 65, 71, 72, 73, 76, 78, 81, 85, 92, 94, 98, 102 (13.2) , 111, 112, 115, 120	19

*The VI_V2 region ranges from 300–400 bp. The VI region ranges from 60–120 bp.

†Amplicons in **bold** are the most abundant amplicon in the profile for that particular sample. The number in parenthesis indicates the percentage abundance of the dominant amplicon for that region.

The most abundant amplicons in UMA, UML and UMR were 349, 360 and 360, respectively (Table 1).

As predicted, V1 profiles ranged from 60 to 120 bp in length (Table 1 and Supplementary Figure 2). Unlike VI_V2 profiles, all 19 samples produced more diverse profiles containing multiple

amplicons in each profile (Table 1 and Supplementary Figure 2). The most complex profile contained 30 amplicons for UME (Figure 2 and Table 1). The least diverse profile with ten amplicons belonged to UMD (Table 1 and Supplementary Figure 2). All the samples contained a 98 bp amplicon, although it was not the most abundant amplicon (Table 1). The 86 bp amplicon was present in ten out of the 19 profiles and was the most dominant peak in UMA (17 per cent) (Table 1). The 84 bp amplicon was found in five profiles. Seventeen out of 19 samples contained a 73 bp amplicon.

Comparison of LH eubacterial profiles within CF centre

The profiles for each sample were compared with each other to determine if the bacterial communities were similar between patients that were treated at the same CF centre. The normalised abundance data from the VI_V2 region were analysed using the Bray–Curtis similarity index. The similarity between samples was shown in a dendrogram using group linkage clustering. The UMF profile had 100 per cent similarity to UMG, UMH, UMJ, UMK, UMP and UMT as they all produced 100 per cent abundance for amplicon 342 (Table 1 and Figure 3a). The UMR profile had 36.0 per cent or less similarity with 13 of the

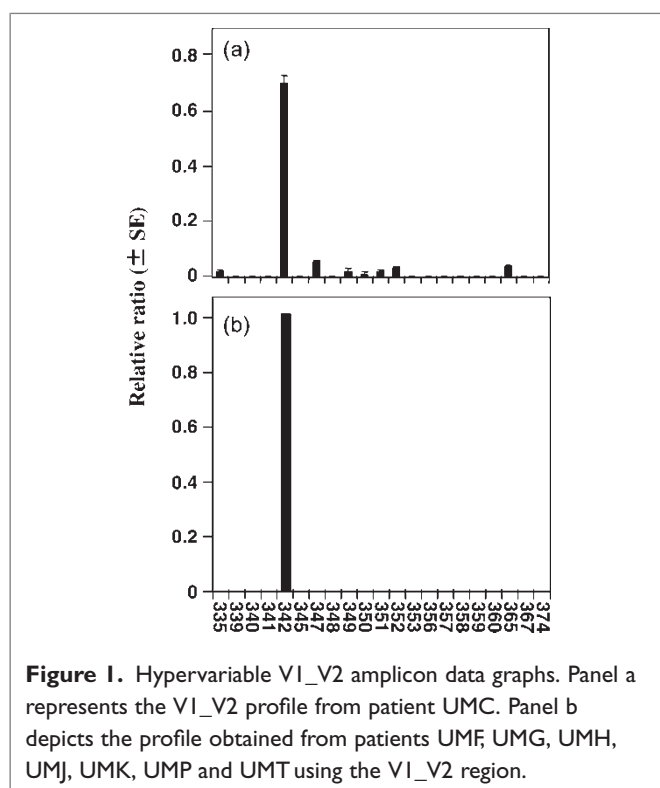


Figure 1. Hypervariable VI_V2 amplicon data graphs. Panel a represents the VI_V2 profile from patient UMC. Panel b depicts the profile obtained from patients UMF, UMG, UMH, UMJ, UMK, UMP and UMT using the VI_V2 region.

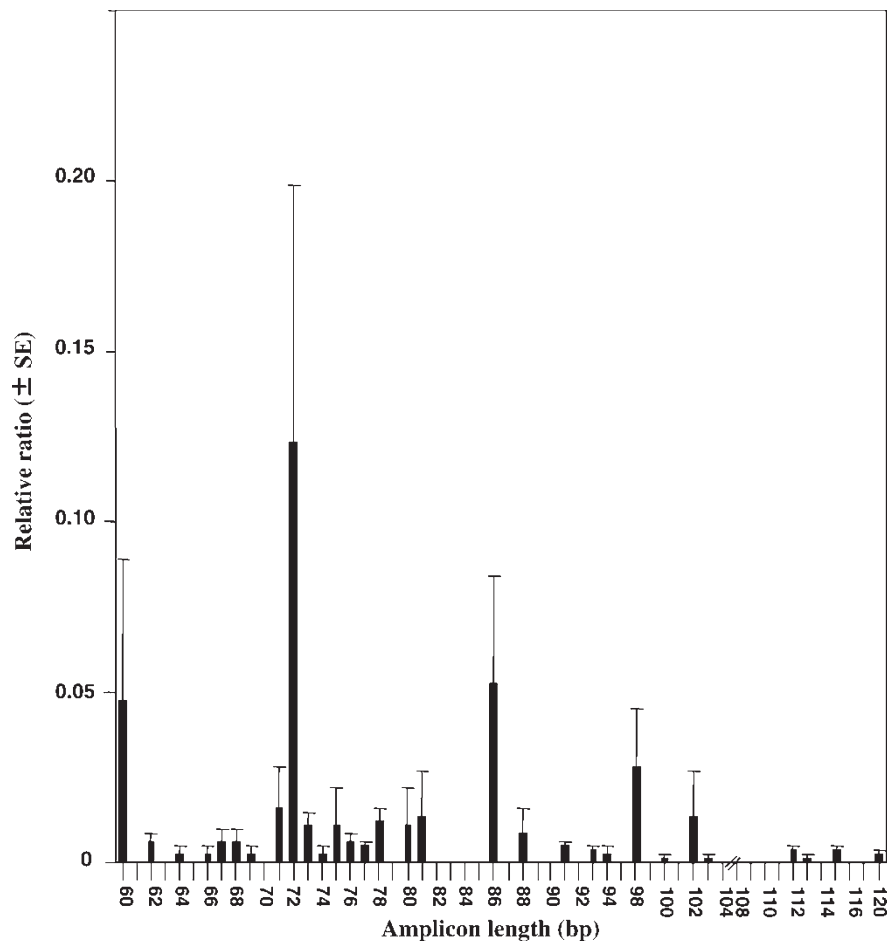


Figure 2. Hypervariable V1 amplicon data. Graph depicts the V1 profile from patient UME, which was the most diverse profile found in this study.

samples, but had a 91.0 per cent similarity with UML (Figure 3a). The UMR and UML profiles had two amplicons in common. The V1 region was able to discriminate between all samples; no two profiles were exactly the same (Table 1 and Figure 3b). The highest similarity between two profiles was 79.2 per cent between UMF and UMG due to the presence of many shared amplicons (Figure 3b). The most dissimilar profiles were those of UMD and UMQ, with just 6.2 per cent similarity between them. Four out of the ten amplicons were common between UMD and UMQ (Table 1).

The similarity of the patients' bacterial population with both the V1_V2 and the V1 data was also compared. The combination of regions decreased the degree of similarity between

V1_V2 profiles, yet the LH profiles were most dissimilar to the V1 region data (data not shown).

Patient CF flora changes over time

The lung flora changes over time in the CF lung. It is also known that the composition of the bacterial community in the lung at any one time varies due to the compartmentalised structure of the organ.³⁴ When working with expectorated sputum, it was assumed that the bacterial community within a given patient may vary from sample to sample. In order to study the dynamics of lung flora over a period of time, sampling bias was addressed — that is, any changes seen in community profiles over time may potentially be an artefact from the sampling method. A control experiment was performed to understand the effect of sampling.

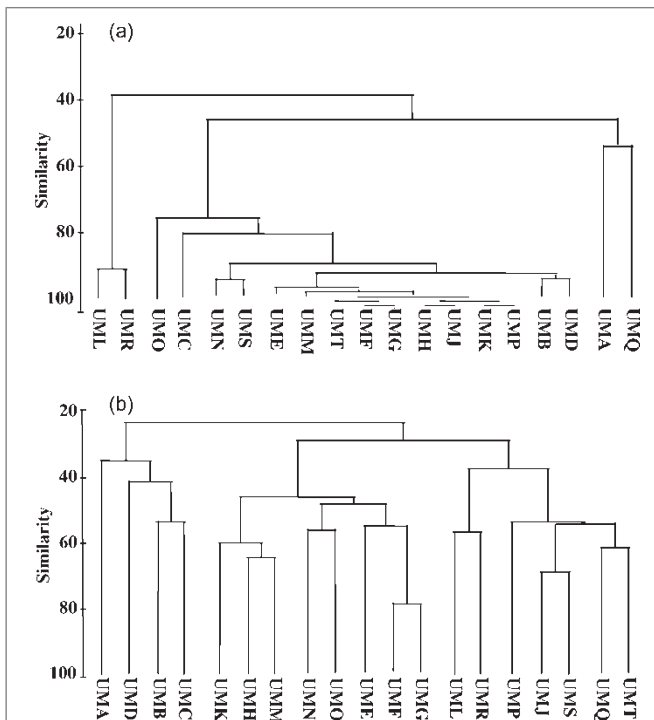


Figure 3. Similarity-based clustering between LH profiles of patients. Panels a and b are dendrograms indicating the relatedness of LH profiles of patients generated from the VI_V2 and VI regions, respectively. Bray–Curtis similarity index and group-linked clustering were used to determine similarity of bacterial communities.

Sputa samples from patient UML were analysed over the course of two years (Month 1 to Month 22). The samples were referred to as M0, M3, M16, M17, M22.1, M22.2, M22.3, M22.4 and M22.5. Five samples were taken on the same day, every three hours for 12 hours (M22.1–M22.5). The sampling scheme was designed to account for the potential variability contributed by the lung structure. All samples were analysed using the V1 and V1_V2 regions. The Bray–Curtis similarity index was used to determine the variation between the community profiles of the patient samples. The relatedness of bacterial communities was shown with dendrograms created from the calculated similarity (Figure 4). In addition, the relationship between profiles was graphically depicted using non-metric MDS, which represents the samples as points in a low-dimensional space, so that the relative distances of the symbols is correlated with the same rank

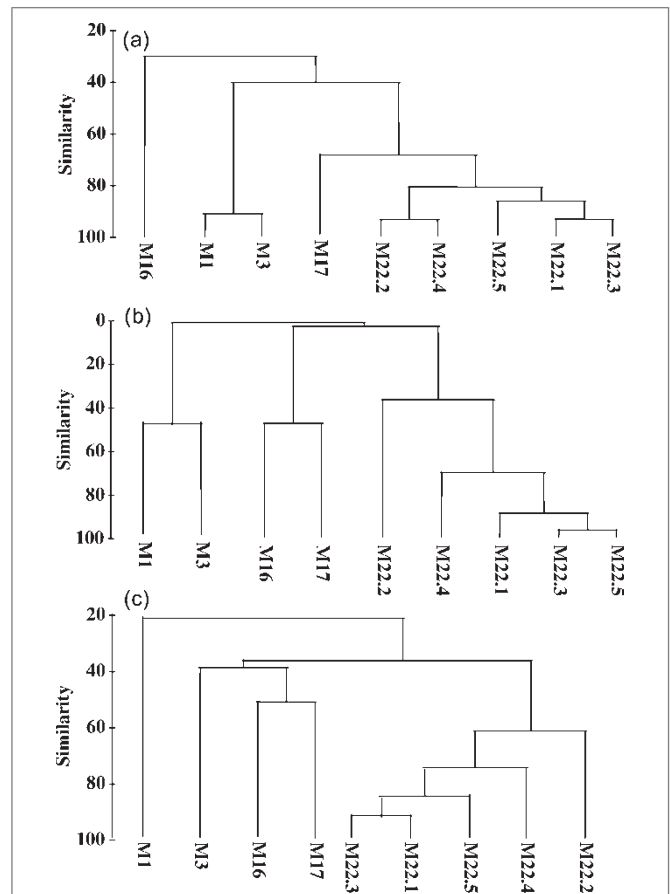
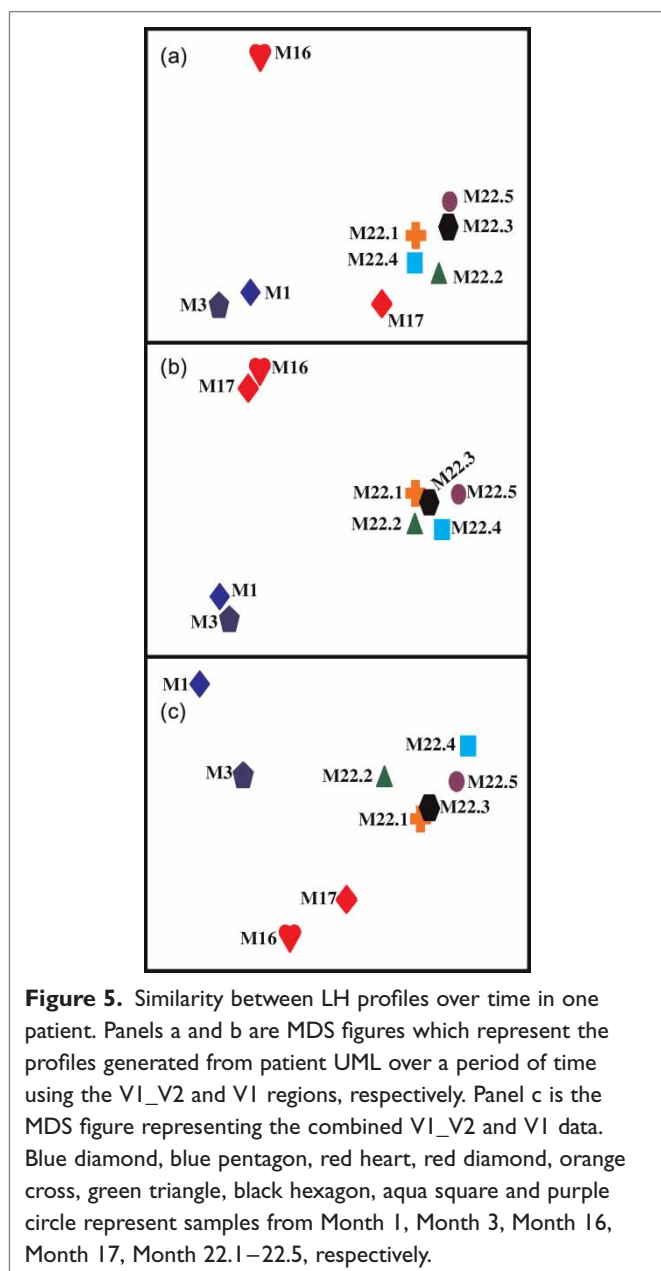


Figure 4. Similarity-based clustering between LH profiles over time in one patient (UML). Panels a and b are dendrograms indicating the relatedness of LH profiles from UML generated from the VI_V2 and VI regions, respectively. Panel c describes the community similarity when both VI_V2 and VI data are combined. The Bray–Curtis similarity index and group-linked clustering were used to determine similarity of bacterial communities.

order determined by the Bray–Curtis similarity index (Figure 5). These analyses showed that samples taken on the same day were more similar than LH profiles from earlier time points (Figures 4 and 5). For the V1_V2 region, the samples in M22 ranged from 70.6 per cent to 93.4 per cent in similarity. The first three samples M1, M3 and M16 had less than 42.9 per cent similarity to any profile produced from the hourly samples. The LH profile M17 and the five profiles in M22 were related, with a range of 52.8 per cent to 78.3 per cent similarity. M22.4 was more similar to M17 than the other same-day samples (Figure 4a).



The V1 profiles for these samples were also analysed and compared. Overall, the similarities between samplings were much lower when using the V1 versus the V1_V2 region (Figures 4b and 5b). The highest profile similarities were seen within M22 samples; similarities as high as 96.0 per cent were seen between M22.3 and M22.5. M1 and M3 had a similarity of 48.1 per cent; M16 and M17 were 47.1 per cent similar. Some of the samplings taken months prior to the daily sampling in

M22 were completely different (Figure 4b). The MDS analysis clearly showed that samples taken on the same day (M22) clustered (Figure 5b).

Combining LH data from both the V1 and the V1_V2 regions further indicated the overall similarity of the bacterial community within one day versus over a period of time (Figures 4c and 5c). The additive effect of the regions created a tighter clustering of the same-day samples and increased distances between samples taken in different months (Figure 4c).

Antibiotics drive the CF community

After the initial time study with patient UML, two more patients were followed for a period of time to determine the effect of antibiotics on the lung microbial flora. Azithromycin, co-trimoxazole (Bactrim), tobramycin and polymyxin E (Colistin) were administered at different time points for each patient. Azithromycin, which is most often used to treat Gram-negative bacteria, is a macrolide antibiotic which interferes with protein synthesis, preventing growth of bacteria.³⁵ Co-trimoxazole (Bactrim) is a combination of sulfamethoxazole and trimethoprim and interferes with DNA synthesis in bacteria that cause respiratory infections, especially in *Stenotrophomonas maltophilia*.^{36,37} Tobramycin, an aminoglycoside, prevents translation in bacteria. The inhaled form of the drug targets *P. aeruginosa* in the lungs.³⁸ Polymyxin E is effective against most Gram-negative bacilli and is often used against multidrug-resistant strains of *P. aeruginosa*.³⁹ The drug regime and sputa samples from two patients, referred to as patient UMX and patient UMY, were analysed every other month for five months. The three time points are referred to as M0, M3 and M5. The amplicons of size 60–120 bp and 300–400 bp were obtained from the V1 and the V1_V2 regions, respectively (Table 2).

Patient UMX was under triple combination therapy with azithromycin, co-trimoxazole (Bactrim) and tobramycin at least two weeks prior and up to the sampling at month 0 (M0) and month 3 (M3). Two weeks prior to month 5 (M5), the patient stopped taking tobramycin but

Table 2: Amplicon contributions for Patients UMX and UMY over a five-month time period.

Patient UMX				Patient UMY			
Amplicon (bp)*	M0 ^a (%)	M3 ^a (%)	M5 ^b (%)	Amplicon (bp)*	M0 ^c (%)	M3 ^d (%)	M5 ^d (%)
Unique amplicons							
352	P (2.34)		P (13.00)				
351	P (1.67)						
348				348		P (1.00)	P (33.88)
347				347	P (35.33)		
120		P (1.88)	P (5.88)	120			
119		P (1.88)	P (3.53)	119			
113		P (3.77)	P (2.35)	113			
111			P (3.53)	111			
108			P (18.82)	108			
100				100	P (4.66)		
94				94	P (0.99)	P (0.81)	P (1.85)
87				87			P (2.47)
84				84			P (1.23)
76				76	P (1.84)	P (1.63)	P (4.94)
74				74		P (0.81)	
71				71		P (4.07)	P (6.79)
70	P (4.36)			70			
69				69		P (0.81)	
67				67		P (0.81)	
66				66		P (0.81)	
65				65			P (3.70)
63				63	P (3.25)		P (3.09)
Shared amplicons							
342	P (93.67)	P (100)	P (74.00)	342	P (61.67)	P (99.00)	P (34.16)
365	P (2.00)			365	P (3.00)		P (2.20)
350			P (13.00)	350			P (29.75)
115		P (3.77)	P (2.35)	115		P (0.81)	P (2.47)
112	P (16.50)	P (3.77)	P (4.71)	112	P (10.31)	P (0.81)	P (3.09)

Continued

Table 2: Continued

Patient UMX				Patient UMY			
Amplicon (bp)*	M0 ^a (%)	M3 ^a (%)	M5 ^b (%)	Amplicon (bp)*	M0 ^c (%)	M3 ^d (%)	M5 ^d (%)
103	P (4.85)	P (9.43)	P (5.88)	103	P (2.68)		P (1.23)
102			P (10.59)	102	P (8.62)	P (3.25)	P (29.63)
98		P (5.66)	P (8.24)	98	P (6.92)	P (6.50)	P (3.70)
93		P (3.77)		93	P (0.99)	P (0.81)	
91		P (3.77)	P (4.71)	91	P (3.53)	P (1.63)	P (2.47)
88	P (12.13)			88	P (19.92)	P (1.63)	
86		P (1.88)		86		P (13.01)	
85			P (7.06)	85			P (6.79)
82	P (14.07)			82	P (2.97)		
81			P (1.18)	81	P (2.12)	P (3.25)	P (3.70)
80			P (1.18)	80		P (2.44)	
78		P (3.77)	P (1.18)	78	P (1.69)	P (3.25)	P (1.85)
77	P (8.73)	P (3.77)	P (5.88)	77	P (4.24)	P (1.63)	
75		P (1.88)		75	P (2.26)	P (2.44)	
73	P (8.73)	P (3.77)	P (3.53)	73	P (2.26)	P (3.25)	
72		P (20.75)		72	P (1.27)	P (30.08)	P (11.73)
68		P (3.77)		68		P (0.81)	
64	P (21.35)			64		P (0.81)	P (1.23)
62		P (3.77)	P (4.71)	62	P (18.79)	P (1.63)	P (2.47)
61		P (3.77)		61			P (1.23)
60		P (15.09)	P (4.71)	60		P (12.20)	P (3.70)

*V1_V2 amplicons range from 300–400 bp whereas V1 amplicons range from 60–120 bp.

M0, M3 and M5 refers to collection of samples at time 0, three and five months

P indicates presence of amplicon with its percentage abundance in the individual profile for that specific region.

^aPatient UMX was treated with azithromycin, co-trimoxazole and tobramycin

^bPatient UMX was treated with co-trimoxazole

^cPatient UMY was treated with levofloxacin (Levaquin)

^dPatient UMY was treated with polymyxin E

continued treatment with azythromycin and co-trimoxazole (Bactrim). The number of amplicons for the V1_V2 and V1 LH profiles varied for the three time points (Table 2). The dominant 342 bp amplicon was present at each time point (93.7 per cent, 100 per cent and 74 per cent at M0, M3, and M5, respectively) for the V1_V2 region (Table 2).

The dominant amplicons in V1 for M0, M3 and M5 were those with 64 (21.4 per cent), 72 (20.8 per cent) and 108 (18.8 per cent) bp, respectively. For the V1 profile, amplicons with 73, 77, 103 and 112 bp were present at all time points. Amplicons with 64, 70, 82, 88, 351 and 365 bp were present only during M0. The 352 bp amplicon was found

during M0, was absent in M3 and was present again at M5. Amplicons with 61, 68, 72, 75, 86 and 93 bp were only present during M3. Amplicons with 60, 62, 78, 91, 98, 113, 115, 119 and 120 bp appeared at M3 and were also detected at M5. M5 unique amplicons had 80, 81, 85, 102, 108, 111 and 350 bp.

Patient UMY was taking levofloxacin for two weeks prior to the M0 sampling. At least two weeks prior to the second sampling (M3), the therapy was changed to colistin. The LH profiles (number and abundance) varied over time (Table 2). The 342 bp amplicon in the V1_V2 region was the most abundant, with 61.7 per cent at M0 and 99.0 per cent at M3; however, this amplicon was drastically decreased at M5 (34.2 per cent) (Table 2). The V1 amplicons with 62, 72, 76, 78, 81, 91, 94, 98, 102 and 112 bp were found at all three time points. The amplicons with 63, 103 and 365 bp were present in M0 and M5 but absent in M3. Amplicons unique to M0 had 82, 100 and 347 bp. Amplicons present at M0 and M3 but absent in M5 had 73, 75, 77, 88 and 93 bp. M3 and M5 shared amplicons with 60, 64, 71 and 115 bp. The six unique amplicons present in M5 had 61, 65, 84, 85, 87 and 350 bp (Table 2).

Presumptive identity analysis

Based on amplicon lengths for both the V1_V2 and V1 regions, separately or combined, a presumptive identity for an organism can be determined. In an attempt to achieve identification of an amplicon, chromosomal DNA from several CF-related pathogens was amplified using V1_V2 primers and V1 primers. All strains of *P. aeruginosa* used in this study amplified a 342 and an 83 bp amplicon for the V1_V2 and V1 regions, respectively. *Stenotrophomonas maltophilia* amplified a 350 and an 87 bp amplicon for the V1_V2 and V1 regions, respectively. A 353 bp amplicon was generated for *Staphylococcus aureus* for the V1_V2 region; however, V1 primers amplified two peaks at 77 and 88 bp for *S. aureus*. *B. cenocepacia* produced a 339 bp and an 85 bp amplicon for V1_V2 and V1, respectively.

Many LH profiles contained amplicons that correspond to these bacterial isolates' fragment sizes. Nineteen LH V1_V2 profiles may have contained the pathogen *P. aeruginosa* but only one sample contained the corresponding V1 amplicon (83 bp) (Table 1). Five samples had the 350 bp amplicon, which may be from *Stenotrophomonas maltophilia*, but only one sample had the corresponding V1 amplicon. None of the samples had the *Staphylococcus aureus*-specific V1_V2 fragment of 353 bp; however, the V1-specific fragments (77/88) were present in 16 samples; nine of these samples contained both fragments. The 339 and 85 bp amplicons, which may correspond to *B. cenocepacia*, were present in two and ten samples, respectively. Two samples contained both amplicons.

The theoretical amplicon lengths for bacteria were determined manually using *in silico* analysis and AmpliQué and then compared with the LH profiles of the strains. The manual *in silico* analysis determined the V1_V2 fragment length for one specific strain of a CF pathogen. These results agreed with the experimental data. The AmpliQué program was developed to determine the V1_V2 fragment lengths for all bacterial strains which had a 16S rRNA sequence in Ribosomal Database Project II (RDPII). The data generated from the program agreed with the experimental data for some strains, yet showed different fragments for other strains. For example, in the database, some strains of *Stenotrophomonas maltophilia* produced the same 350 bp fragment as our LH result. In addition, many more fragment lengths (304–344 and 346–351 bp) were identified for various strains of *Stenotrophomonas maltophilia*. Various isolates of *P. aeruginosa* produced theoretical fragments ranging from 337–345 bp. Only two fragments, 347 bp and 348 bp, were produced for this region based on *H. influenzae* sequences. AmpliQué determined that different isolates of *Staphylococcus aureus* have a V1_V2 of 340–344 and 346–351 bp. Based on *B. cenocepacia* sequences, any of the lengths 327, 335, 337, 338, 340, 344 or 346 bp fragments could be generated using the bacterial primers. Most bacterial species produced multiple hits for the V1_V2 region (data not shown).

AmpliQué was unable to determine the fragment sizes for many CF-related pathogens, including

P. aeruginosa. *In silico* analysis was able to determine the fragment lengths of CF pathogens, as the site of primer binding could be detected manually. Again, fragment lengths generated from sequences of individual strains varied from the actual LH results for the specific control strain. Four different isolates of *P. aeruginosa* produced an 86 bp fragment for the V1 region, while experimentally the region was 83 bp long. Some strains of *Stenotrophomonas maltophilia* produced the same 87 bp fragment as the control strain, while others were determined to have a fragment of 85 or 88 bp in length. *Staphylococcus aureus* strains were 91 bp long in that region, but the test strain did not produce this fragment. The fragment from *B. cenocepacia* was three bp longer (88 bp) than the LH isolate's profile (85 bp).

Discussion

LH analysis of 16S rRNA genes has been used to detect both known and novel organisms that may be present in many complex microbial communities, including sputa from CF patients.^{18,20,21,26} In these previous studies, only one region of the 16S rRNA was used to identify bacteria present in the community at one time point.^{18,26} Identification based on one region is difficult, and it is hypothesised that the use of multiple regions might provide more information about the community. In this study, the V1 and V1_V2 regions (separately and combined) were analysed with LH-PCR to determine the diversity and dynamics of the CF lung. Identification of the most abundant amplicon lengths in the LH profiles was attempted through *in silico* analysis and AmpliQué, a newly designed bioinformatics program. Lastly, the dynamics of the eubacterial community in the sputum over a period of time was further studied using data generated from the LH profiles.

The V1 region of the 16S rRNA gene provides a more detailed look at the complex bacterial community in the CF lung

The 16S rRNA gene contains nine variable regions. In this study, the V1 region and the V1_V2 regions were used to produce bacterial LH profiles of the

CF sputa samples (Table 1). The V1_V2 profiles were identical for seven patients and hence were unable to discriminate between patients. The 342 bp amplicon produced a high relative fluorescence in the electropherograms, which may have caused the less dominant amplicons to be below the threshold limit and inadvertently ignored (Table 1). PCR amplification bias may also have decreased the detection of other bacteria that were less abundant than the organism(s) represented by the 342 bp amplicon. Therefore, it appeared that seven CF samples had identical bacterial communities. Statistically, the V1_V2 profiles cannot discriminate between patients at the UM centre (Figure 3a). The V1 LH profiles were more complex, and statically discriminated between samples to a higher degree than the V1_V2 region (Figure 3). The V1 profiles which were unique to each patient contained multiple amplicons per sample (Table 1).

The use of multiple regions to increase discrimination between samples was not beneficial. The identical communities shown in the V1_V2 profiles negated the discriminatory power of the V1 region when analysed in conjunction with each other. The discriminatory strength of the V1 region, as compared with the V1_V2 region, was also evident in a soil community LH study.²⁵ Prior LH analysis on CF sputa used only the V1_V2 region, and more information may have been gained by using the V1 region.¹⁸ Using the profiles from the two regions together did not differentiate profiles to the same extent as the V1 region. The combinatorial approach may be of more use in amplicon identification or observing community dynamics.

CF patients harbour diverse bacterial communities

The 19 sputum samples obtained from South Florida patients produced unique profiles. No two LH profiles for the V1 region were the same (Table 1 and Figure 3b). The V1_V2 region profiles all contained an amplicon of length 342 bp, which presumptively could be identified as *Pseudomonas* species, *Burkholderia* species (but not *B. cenocepacia*) or *Ralstonia* species. The 342 bp amplicon is indicative of *P. aeruginosa*, since control CF isolates

produced a fragment of this size. In addition, the patients were clinically diagnosed with *P. aeruginosa*. The high relative fluorescence of the 342 bp amplicon also indicated a severe infection, which would be in agreement with culturing results. The intense fluorescence of the fragment in seven samples may have skewed the community profile, causing it to appear as if there was only one type of bacteria present. The V1 region for these seven samples show some shared amplicons, but largely they had very diverse microbial communities. Although some amplicons (presence of, not abundance) are common to patients, no two overall profiles were identical to each other (Table 1 and Figure 3b).

It was thought that there are only a few pathogens commonly found in the lungs of CF patients, due to clinical diagnostic procedures.⁴⁰ Molecular techniques have revealed that many more organisms plague the CF lung, which is in agreement with our findings.^{17,18} Many of the samples in this study produced LH profiles that contained numerous amplicons, some in low abundance. These less dominant amplicons may represent bacteria that are not routinely cultured and are potentially responsible for some of the differences seen in disease manifestation between CF patients with the same genetic mutations. Recent Sanger sequencing projects have identified uncultured bacteria in the CF lung.^{11,17} Deeper sequencing techniques, such as 454, could be used to identify other, less abundant and potentially uncultured organisms.⁴¹ In order to understand the significance of these organisms within the CF lung, proper statistical tools for comparative metagenomics need to be developed.

Challenges of identifying organisms in complex communities

One of the downsides to the LH technique is that the organism found in a sample cannot be conclusively identified. We hypothesised that organisms could be presumptively identified based on the V1_V2 and V1 amplicon lengths that would be generated from LH-PCR. In this study, the fragment lengths for both regions were experimentally determined for one strain of four known CF pathogens. In addition, *in silico* analysis and AmpliQué

were used to determine the expected amplicon length of a variable region for any given organism. This information was correlated with the LH profile in an attempt to identify the bacteria present in the CF sputum. The bioinformatics approach highlighted the presence of intra-species length variation present in both regions. The high degree of variation seen in different isolates drastically decreased the ability to identify a bacteria based on the length of one hypervariable region. It was previously known that one amplicon length could represent multiple species, thereby allowing only a presumptive identification. Based on the absence of a peak, a specific genus or species could be ruled out from being a member of the community. The AmpliQué output clearly demonstrates that one V1_V2 fragment could represent a multitude of species or genera. The wide range of fragment lengths generated for different isolates of the species may be accurate or it may be due to problems related to the database used in the program. The database contains sequences which are determined by various users. It is possible that some of the length heterogeneity at the isolate level may arise from poor sequencing reads, incorrect trimming of sequences, and old taxonomy references (some organism names have changed). Database issues will change as more genome-wide sequencing is performed. As it stands currently, identification is nearly impossible based solely on amplicon length from the V1_V2 region.

The V1 region, which was more informative when looking at the community profile as a whole, proved to be more problematic when analysing individual peaks. Chromosomal DNA of *P. aeruginosa*, *Staphylococcus aureus*, *Stenotrophomonas maltophilia* and *B. cenocepacia* were amplified using the V1 primers and the fragment lengths were detected. *In silico* analysis was used to determine fragment lengths for the CF-associated lung pathogens, including those listed above (although not always the same isolate). The initial analysis indicated the presence of length heterogeneity within species. AmpliQué was used to determine the degree of strain variation present in the V1 region. The primer binding parameter was initially set to 100

per cent stringency, which resulted in a very small database of organisms and their corresponding fragment lengths. Many CF pathogens, including *P. aeruginosa*, were not present at this high stringency. By lowering the stringency, AmpliQué generated more results, yet no information was given for *P. aeruginosa*. The V1 region primers (P1F and P2R) experimentally amplified *P. aeruginosa* PAO1. To determine the putative primer binding site, the 16S rRNA sequence for PAO1 was manually examined. Both the forward and reverse primer sequences show some degeneracy. Lowering the stringency parameter was unable to correct for this degeneracy. This issue can be resolved by determining the binding motif of the primers using programs such as iterative enhancement of motifs (IEM)⁴² and then using that motif in AmpliQué. Although P1F is a common V1 primer used extensively in the laboratory, other, less degenerate V1 forward primers, such as 27F (which was used to amplify V1_V2), could be used when attempting to identify organisms based on amplicon length. Selection of primers for LH-PCR is critical to ensure that all bacteria are being amplified and to increase the success of identification based on fragment length. The widely used primers, 27F and 355R, are based on *Escherichia coli* sequence and should be thought of as generalised primers and not universal primers.⁴³ Primers that have been developed with a bioinformatics approach may prove to be more useful in future LH studies.⁴³

Due to the limitations of the V1 AmpliQué database, identification using two regions simultaneously could not be performed. Therefore, identification based on two regions has yet to be proven. Further modification of AmpliQué and/or the use of different primers or variable regions may eventually lead to LH-PCR being used as an identification technique. At this time, LH-PCR can still be useful when looking at a bacterial community as a whole.

The eubacterial communities in CF lungs are dynamic

The LH-PCR profiles provided a view of the CF bacterial community as a whole. It is known that in soil and water samples, the members of a community

and their abundances change, and these changes can be driven by external factors.^{44,45} Thus, it is hypothesised that the bacterial community within the CF lung also changes based on external factors. To understand the changes in the community, LH-PCR was used to profile three patients over a period of time. Samples from patient UML were used to determine the bacterial community variation over a short period of time (three hours) and over a long period of time (two years). Samples from patients UMY and UMX were used to study how antibiotics may affect the lung community.

Sputum samples from patient UML were analysed over the course of two years (M0, M3, M16, M17, M22.1, M22.2, M22.3, M22.4 and M22.5). Five samples were taken on the same day, every three hours for 12 hours (M22.1–M22.5). For both the V1 and the V1_V2 LH profiles, the overall bacterial community was more similar within the one day than across multiple time periods over two years (Figures 4 and 5). The LH profiles from the samples taken in one day were not identical. These variations may arise from transient bacteria that were present in the oral cavity at the time of sampling (such as *Lactobacillus* from yogurt). Analysing the samples with both regions further demonstrated that bacterial communities change over time. By combining regions, the data indicated increased similarity of the samples taken in one day, which was due to the more abundant organisms, not the less dominant amplicons in the profile that may represent the transient bacteria in the oral community. The dissimilarity between the samples taken months apart was increased due to the presence of the more dominant, possibly disease-causing, organisms (Figures 4c and 5c). This LH analysis is the first to show that colonisation of the CF lung is dynamic and that these changes can be tracked.

To date, the dynamics of the bacterial community in the CF lung in response to external drivers such as the presence of antibiotics has yet to be studied. Thus, a short-term study was performed in an attempt to understand how the overall community in the lung changes when under attack by antibiotics. To begin, CF sputum samples from patient

UMX and patient UMY were analysed every other month for five months. During this time, patients were taking a combination of antibiotics.

During the five-month period, patient UMX was taking azithromycin, co-trimoxazole and tobramycin. Shortly before the third time point, the patient stopped taking tobramycin. The patient's profiles contained a large number of V1 amplicons, whose presence and abundance changed throughout the time course (Table 2). For V1_V2, patient UMX's profiles consisted predominantly of amplicon 342 at all time points (Table 2). This fragment was presumptively identified as *P. aeruginosa*, which concurred with the patient's clinical diagnosis. Interestingly, UMX had been prescribed tobramycin (which targets *P. aeruginosa*) for the majority of the sampling period, yet the 342 bp amplicon was always present in large abundance (74–100 per cent). LH analysis with both the V1 and V1_V2 regions indicated a dynamic environment within the lungs of UMX. These changes seen over the time period are likely to be an effect of the antibiotic regime.

Patient UMY's profile was more dynamic than that of patient UMX. Patient UMY's sputum showed a high level of abundance for the 342 bp amplicon between the first two samplings, followed by a drastic decrease in the abundance of the 342 bp amplicon in the profile from the third sampling (Table 2). This change in profile may correspond to the patient's antibiotic regime polymyxin E specifically, replacing levofloxacin. This may have caused the pathogen associated with that amplicon to disappear. As the abundance of the 342 bp amplicon decreased, a 348 bp amplicon emerged. It is possible that as one pathogen was cleared, it was replaced with another bacterium that fitted that niche.⁴⁶ The decreasing 342 bp amplicon may represent *P. aeruginosa*, which had been cultured from the patient. Levofloxacin and polymyxin E are both used to target against this pathogen. There are studies that have demonstrated that levofloxacin is less effective in treating *P. aeruginosa* than the first-generation fluoroquinolones.⁴⁷ Antibiotic-resistant strains of *P. aeruginosa* are often treated with polymyxin E.^{48,49} It could be hypothesised that patient UMY was infected with an antibiotic-

resistant strain of *P. aeruginosa* which may have been susceptible to the stronger antibiotic. The patient's lung infection was then dominated by a bacterium represented by the 348 bp amplicon, which may belong to *H. influenzae* or *Stenotrophomonas maltophilia*. Changes in the bacterial community were also seen in the V1 profiles (Table 2). Due to the limitations of AmpliQué, it is not clear if similar changes in abundance were seen in peaks that may represent *P. aeruginosa*.

The microbial community was more dynamic in UMY, a 60-year-old Caucasian man, than in UMX, a 23-year-old Hispanic woman. The number of unique amplicons in UMX and UMY were eight and 14, respectively. UMX had five (73, 77, 103, 112 and 342 bp) and UMY had 11 (62, 72, 76, 78, 81, 91, 94, 98, 102, 112 and 342 bp) amplicons that were present at all three time points. These constant amplicons may come from chronic colonisers that no longer respond to the treatment regimen. The uniqueness, diversity and selective abundance of certain amplicons could be attributed to the age, gender and ethnicity. The patients' respective lung flora may have adapted to their immune response and the presence of antibiotics differently. The variations in profiles due to age and antibiotic treatment have been previously demonstrated.^{5,50}

Clearly, LH can be used to analyse changes in a microbial community. Although this study used a small sample size over a relatively short period of time, changes in the presence and abundance of amplicons in the LH profiles could be seen. To identify the drivers of the community accurately, a more comprehensive long-term study needs to be performed. Armed with the patient's antibiotic information and long-term sampling, it should be possible to determine the effectiveness of a drug against pathogens in the CF lung. This could eventually lead to more effective treatments.

In conclusion, LH-PCR analysis detected the dynamic and complex flora that is present in South Florida CF patients. Although LH-PCR could not be used to identify bacteria in this study, it still proved to be a useful community profiling technique. LH-PCR of the V1_V2 and, to a greater extent, the V1 region of the 16S rRNA gene can be used to

compare microbial communities between samples. Interpreting LH profiles may provide insight into the evolution of a microbial community and identify the factors that drive these changes. Technologies such as pyrosequencing will probably be used to identify the members present in the bacterial community; however, LH-PCR is still an accessible tool that can be implemented to study microbial communities.

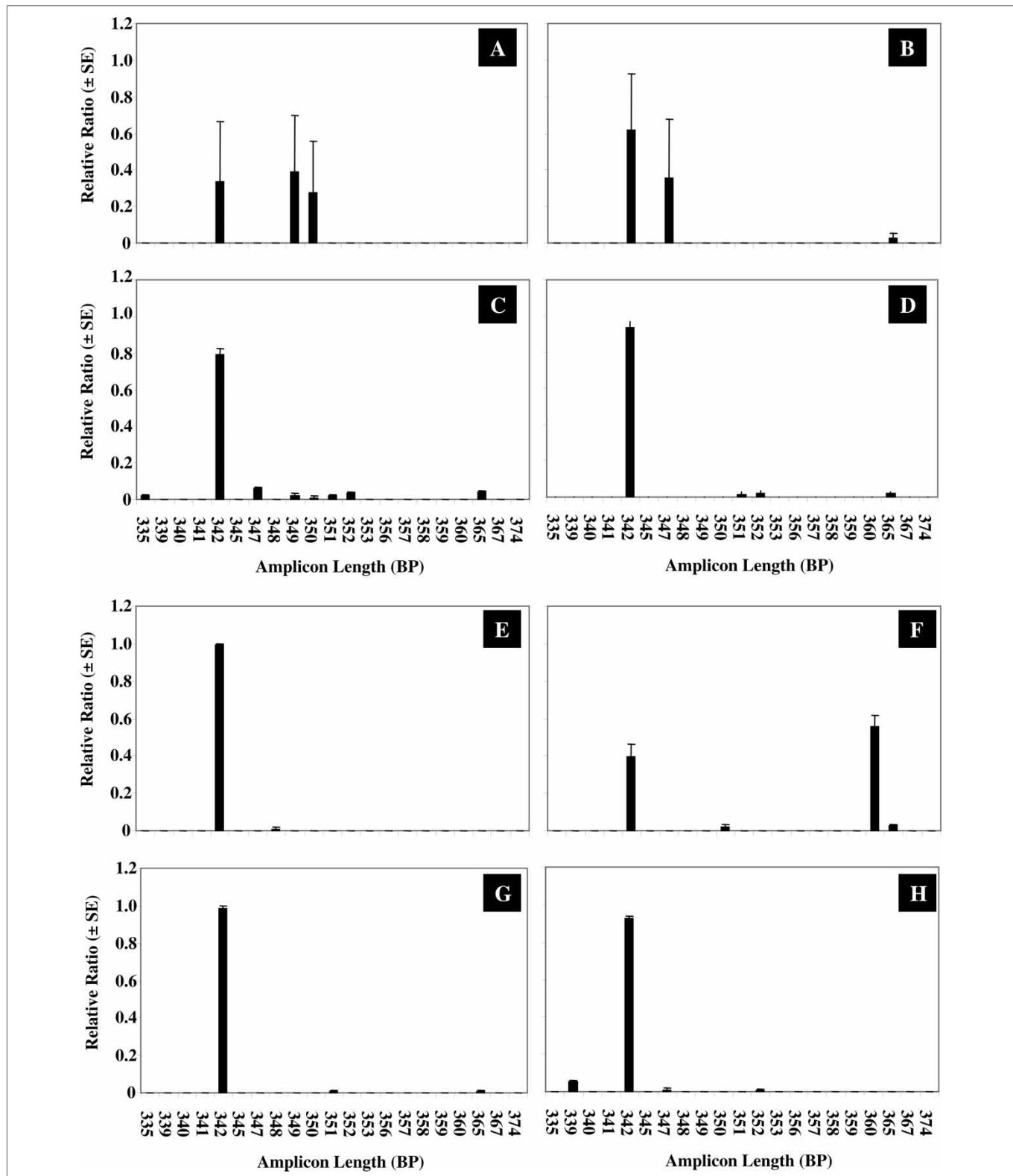
Acknowledgments

This work was supported by Florida International Teaching Assistantship (M.D.) and MBRS Research Initiative for Scientific Enhancement (RISE) Program NIH\NIGMS R25 GM061347 (MD). We would like to thank DeEtta Mills for her many conversations that helped M.D. to troubleshoot the project over the years, Sasha Miller for training M.D. in the laboratory and Lilliana Moreno for help with the statistical analysis. We are grateful to Robert Sautter for developing the macros used in the analysis. We also thank Burkhard Tümmler for use of his *P. aeruginosa* strains and Lisa Schneper for the final reading of the manuscript. All LH-PCR samples were analysed at the FIU Forensic DNA Profiling Facility. Lastly, we are indebted to all South Floridian CF patients who were willing to enrol in our study.

References

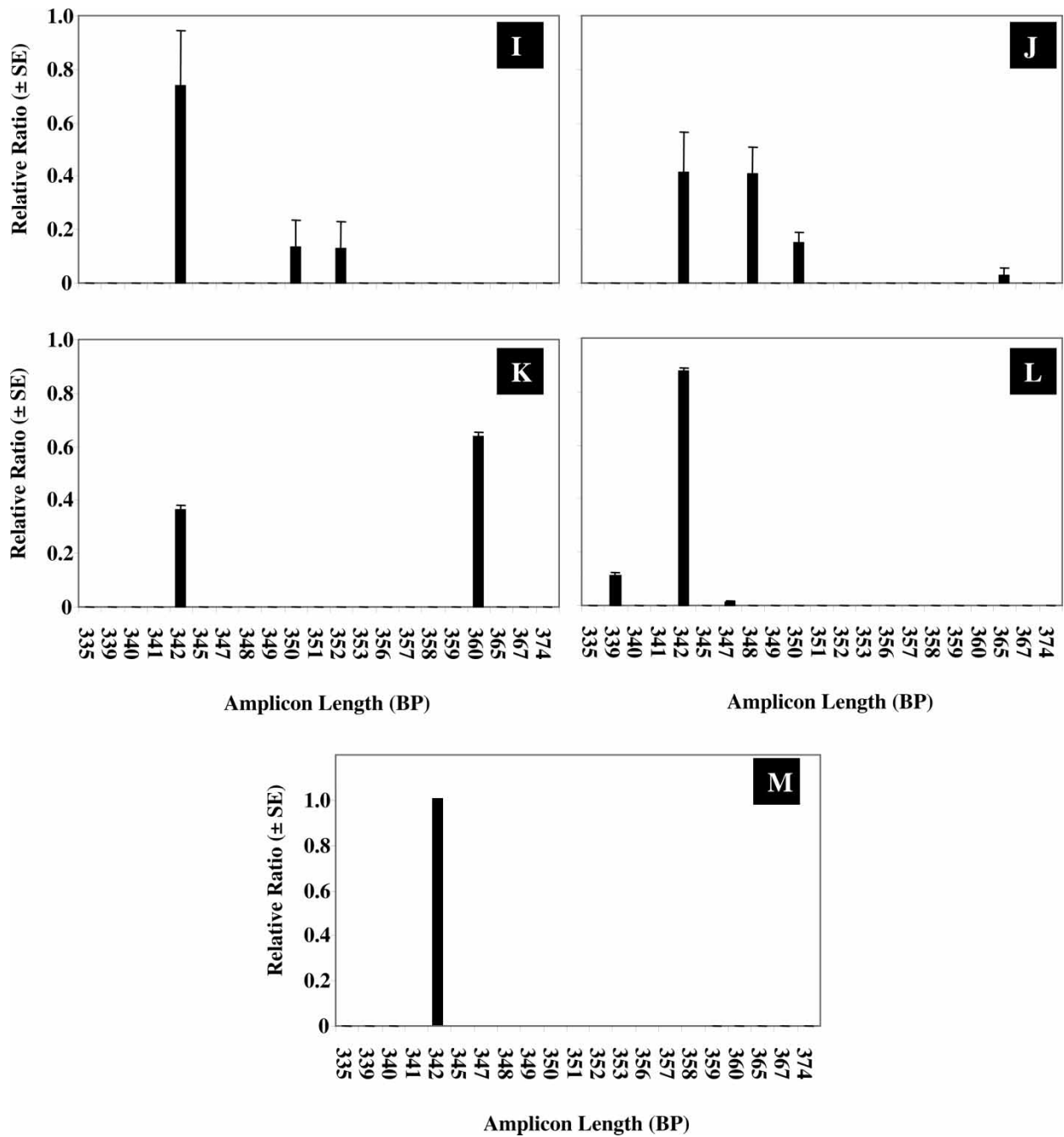
- Cystic Fibrosis Foundation. (2008), 'Patient Registry. Annual Data Report 2008', Available at: www.cff.org/LivingWithCF/QualityImprovement/PatientRegistryReport/ [last accessed 1st March, 2010].
- Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N. *et al.* (1989), 'Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA', *Science* Vol. 245, pp. 1066–1073.
- Wainwright, B.J., Scambler, P.J., Schmidtke, J., Watson, E.A. *et al.* (1985), 'Localization of cystic fibrosis locus to human chromosome 7cen-q22', *Nature* Vol. 318, pp. 384–385.
- Kunzelmann, K. (1999), 'The cystic fibrosis transmembrane conductance regulator and its function in epithelial transport', *Rev. Physiol. Biochem. Pharmacol.* Vol. 137, pp. 1–70.
- Lyczak, J.B., Cannon, C.L. and Pier, G.B. (2002), 'Lung infections associated with cystic fibrosis', *Clin. Microbiol. Rev.* Vol. 15, pp. 194–222.
- Hull, J. (2003), 'Basic science of cystic fibrosis', *Curr. Paediatr.* Vol. 13, pp. 253–258.
- Mathee, K., Ciofu, O., Sternberg, C., Lindum, P.W. *et al.* (1999), 'Mucoicid conversion of *Pseudomonas aeruginosa* by hydrogen peroxide: A mechanism for virulence activation in the cystic fibrosis lung', *Microbiology* Vol. 145, pp. 1349–1357.
- Ward, D., Weller, R. and Bateson, M. (1990), '16S rRNA sequences reveal numerous uncultured microorganisms in a natural community', *Nature* Vol. 345, pp. 63–65.
- Ward, D.M., Weller, R. and Bateson, M.M. (1990), '16S rRNA sequences reveal uncultured inhabitants of a well-studied thermal community', *FEMS Microbiol. Rev.* Vol. 6, pp. 105–115.
- Belkum, A., Renders, N.H.M., Smith, S., Overbeek, S.E. *et al.* (2000), 'Comparisons of conventional and molecular methods for the detection of bacterial pathogens in sputum samples from cystic fibrosis patients', *FEMS Immunol. Med. Microbiol.* Vol. 27, pp. 51–57.
- Bittar, F., Richet, H., Dubus, J.C., Reynaud-Gaubert, M. *et al.* (2008), 'Molecular detection of multiple emerging pathogens in sputa from cystic fibrosis patients', *PLoS ONE* Vol. 3, p. e2908.
- Burns, J.L., Emerson, J., Stapp, J.R., Yim, D.L. *et al.* (1998), 'Microbiology of sputum from patients at cystic fibrosis centers in the United States', *Clin. Infect. Dis.* Vol. 27, pp. 158–163.
- Coenye, T., Liu, L., Vandamme, P. and LiPuma, J.J. (2001), 'Identification of *Pandora* species by 16S ribosomal DNA-based PCR assays', *J. Clin. Microbiol.* Vol. 39, pp. 4452–4445.
- Coenye, T., Goris, J., Spilker, T., Vandamme, P. *et al.* (2002), 'Characterization of unusual bacteria isolated from respiratory secretions of cystic fibrosis patients and description of *Inquilinus limosus* gen. nov., sp. nov.', *J. Clin. Microbiol.* Vol. 40, pp. 2062–2069.
- Coenye, T., Goris, J., De Vos, P., Vandamme, P. *et al.* (2003), 'Classification of *Ralstonia pickettii*-like isolates from the environment and clinical samples as *Ralstonia insidiosus* sp. nov.', *Int. J. Syst. Evol. Microbiol.* Vol. 53, pp. 1075–1080.
- Coenye, T., Spilker, T., Reik, R., Vandamme, P. *et al.* (2005), 'Use of PCR analyses to define the distribution of *Ralstonia* species recovered from patients with cystic fibrosis', *J. Clin. Microbiol.* Vol. 43, pp. 3463–3466.
- Harris, J.K., De Groote, M.A., Sagel, S.D., Zemanick, E.T. *et al.* (2007), 'Molecular identification of bacteria in bronchoalveolar lavage fluid from children with cystic fibrosis', *Proc. Natl. Acad. Sci. USA* Vol. 104, pp. 20529–20533.
- Rogers, G.B., Hart, C.A., Mason, J.R., Hughes, M. *et al.* (2003), 'Bacterial diversity in cases of lung infection in cystic fibrosis patients: 16S ribosomal DNA (rDNA) length heterogeneity PCR and 16S rDNA terminal restriction fragment length polymorphism profiling', *J. Clin. Microbiol.* Vol. 41, pp. 3548–3558.
- Bernhard, A.E., Colbert, D., McManus, J. and Field, K.G. (2005), 'Microbial community dynamics based on 16S rRNA gene profiles in a Pacific Northwest estuary and its tributaries', *FEMS Microbiol. Ecol.* Vol. 52, pp. 115–128.
- Suzuki, M., Rappe, M.S. and Giovannoni, S.J. (1998), 'Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity', *Appl. Environ. Micro.* Vol. 64, pp. 4522–4529.
- Mills, D.K., Fitzgerald, K., Litchfield, C.D. and Gillevet, P.M. (2003), 'A comparison of DNA profiling techniques for monitoring nutrient impact on microbial community composition during bioremediation of petroleum-contaminated soils', *J. Microbiol. Methods* Vol. 54, pp. 57–74.
- Ritchie, N.J., Schutter, M.E., Dick, R.P. and Myrold, D.D. (2000), 'Use of length heterogeneity PCR and fatty acid methyl ester profiles to characterize microbial communities in soil', *Appl. Environ. Microbiol.* Vol. 66, pp. 1668–1675.
- Coenye, T. and Vandamme, P. (2003), 'Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes', *FEMS Microbiol. Lett.* Vol. 228, pp. 45–49.
- Case, R.J., Boucher, Y., Dahllof, I., Holmstrom, C. *et al.* (2007), 'Use of 16S rRNA and *rpoB* genes as molecular markers for microbial ecology studies', *Appl. Environ. Microbiol.* Vol. 73, pp. 278–288.
- Moreno, L.I., Mills, D.K., Entry, J., Sautter, R.T. *et al.* (2006), 'Microbial metagenome profiling using amplicon length heterogeneity-polymerase chain reaction proves more effective than elemental analysis in discriminating soil specimens', *J. Forensic Sci.* Vol. 51, pp. 1315–1322.
- Miller, S.Z. (2003), 'Molecular identification of microbes in the sputum of cystic fibrosis patients', Thesis, Florida International University, Miami, FL.
- Reischl, U., Pulz, M., Ehret, W. and Wolf, H. (1994), 'PCR based detection of *Mycobacterium* in sputum samples using a simple and reliable DNA extraction protocol', *Biotechniques* Vol. 17, pp. 844–845.
- Klockgether, J., Würdemann, D., Reva, O., Wiehlmann, L. *et al.* (2007), 'Diversity of the abundant pKLC102/PAGI-2 family of genomic islands in *Pseudomonas aeruginosa*', *J. Bacteriol.* Vol. 189, pp. 2443–2459.
- Doud, M., Zeng, E., Schneper, L., Narasimhan, G. *et al.* (2009), 'Approaches to analyse dynamic microbial communities such as those seen in cystic fibrosis lung', *Hum. Genomics* Vol. 3, pp. 246–256.

30. Coccolin, L., Manzano, M., Cantoni, C. and Comi, G. (2001), 'Denaturing gradient gel electrophoresis analysis of the 16S rRNA gene V1 region to monitor dynamic changes in the bacterial population during fermentation of Italian sausages', *Appl. Environ. Microbiol.* Vol. 67, pp. 5113–5121.
31. Rees, G., Baldwin, D., Watson, G., Perryman, S. et al. (2004), 'Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms: Application of multivariate statistics', *Antonie Leeuwenhoek.* Vol. 86, pp. 339–347.
32. Hartmann, M., Frey, B., Kolliker, R. and Widmer, F. (2005), 'Semi-automated genetic analyses of soil microbial communities: Comparison of T-RFLP and RISA based on descriptive and discriminative statistical approaches', *J. Microbiol. Methods* Vol. 61, pp. 349–360.
33. Gonzalez, G., Doud, M., Mathee, K. and Narasimhan, G. (2009), 'Computer-assisted bacterial identification using 16S rRNA sequence data', *25th Southern Biomedical Engineering Conference: IFMBE Proceedings*, Springer Berlin Heidelberg, Miami, FL, pp. 239–240.
34. Harrison, F. (2007), 'Microbial ecology of the cystic fibrosis lung', *Microbiology* Vol. 153, pp. 917–923.
35. Retsema, J., Girard, A., Schelkly, W., Manousos, M. et al. (1987), 'Spectrum and mode of action of azithromycin (CP-62,993), a new 15-membered-ring macrolide with improved potency against Gram-negative organisms', *Antimicrob. Agents Chemother.* Vol. 31, pp. 1939–1947.
36. Hughes, D.T. and Russell, N.J. (1982), 'The use of trimethoprim-sulfamethoxazole in the treatment of chest infections', *Rev. Infect. Dis.* Vol. 4, pp. 528–532.
37. Muder, R.R., Harris, A.P., Muller, S., Edmond, M. et al. (1996), 'Bacteremia due to *Stenotrophomonas (Xanthomonas) maltophilia*: A prospective, multicenter study of 91 episodes', *Clin. Infect. Dis.* Vol. 22, pp. 508–512.
38. Ramsey, B.W., Pepe, M.S., Quan, J.M., Otto, K.L. et al. (1999), 'Intermittent administration of inhaled tobramycin in patients with cystic fibrosis', *N. Engl. J. Med.* Vol. 340, pp. 23–30.
39. Beringer, P. (2001), 'The clinical use of colistin in patients with cystic fibrosis', *Curr. Opin. Pulm. Med.* Vol. 7, pp. 434–440.
40. Govan, J.R.W. and Nelson, J.W. (1993), 'Microbiology of cystic fibrosis lung infections: Themes and issues', *J. R. Soc. Med.* Vol. 86, pp. 11–18.
41. Margulies, M., Egholm, M., Altman, W.E., Attiya, S. et al. (2005), 'Genome sequencing in microfabricated high-density picolitre reactors', *Nature* Vol. 437, pp. 376–380.
42. Zeng, E., Mathee, K. and Narasimhan, G. (2007), 'IEM: An algorithm for iterative enhancement of motifs using comparative genomics data', *Comput. Syst. Bioinformatics Conf.* Vol. 6, pp. 227–235.
43. Wang, Y. and Qian, P.Y. (2009), 'Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies', *PLoS ONE* Vol. 4, p. e7401.
44. Fierer, N. and Jackson, R.B. (2006), 'The diversity and biogeography of soil bacterial communities', *Proc. Natl. Acad. Sci. USA* Vol. 103, pp. 626–631.
45. Shade, A., Jones, S.E. and McMahon, K.D. (2008), 'The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics', *Environ. Microbiol.* Vol. 10, pp. 1057–1067.
46. Boni, M.F. and Feldman, M.W. (2005), 'Evolution of antibiotic resistance by human and bacterial niche construction', *Evolution* Vol. 59, pp. 477–491.
47. Phillips, I., King, A. and Shannon, K. (2009), 'Comparative in-vitro properties of the quinolones', in: Andriole, V. (ed.), *The Quinolones*, Academic Press, San Diego, CA, pp. 99–137.
48. Scheld, M.W. (2003), 'Maintaining fluorquinolone class efficacy: Review of influencing factors', *Emerg. Infect. Dis.* Vol. 9, pp. 1–9.
49. Canton, R., Cobos, N., de Gracia, J., Baquero, F. et al. (2005), 'Antimicrobial therapy for pulmonary pathogenic colonisation and infection by *Pseudomonas aeruginosa* in cystic fibrosis patients', *Clin. Microbiol. Infect.* Vol. 11, pp. 690–703.
50. Beringer, P.M. and Appleman, M.D. (2000), 'Unusual respiratory bacterial flora in cystic fibrosis: Microbiologic and clinical features', *Curr. Opin. Pulm. Med.* Vol. 6, pp. 545–550.

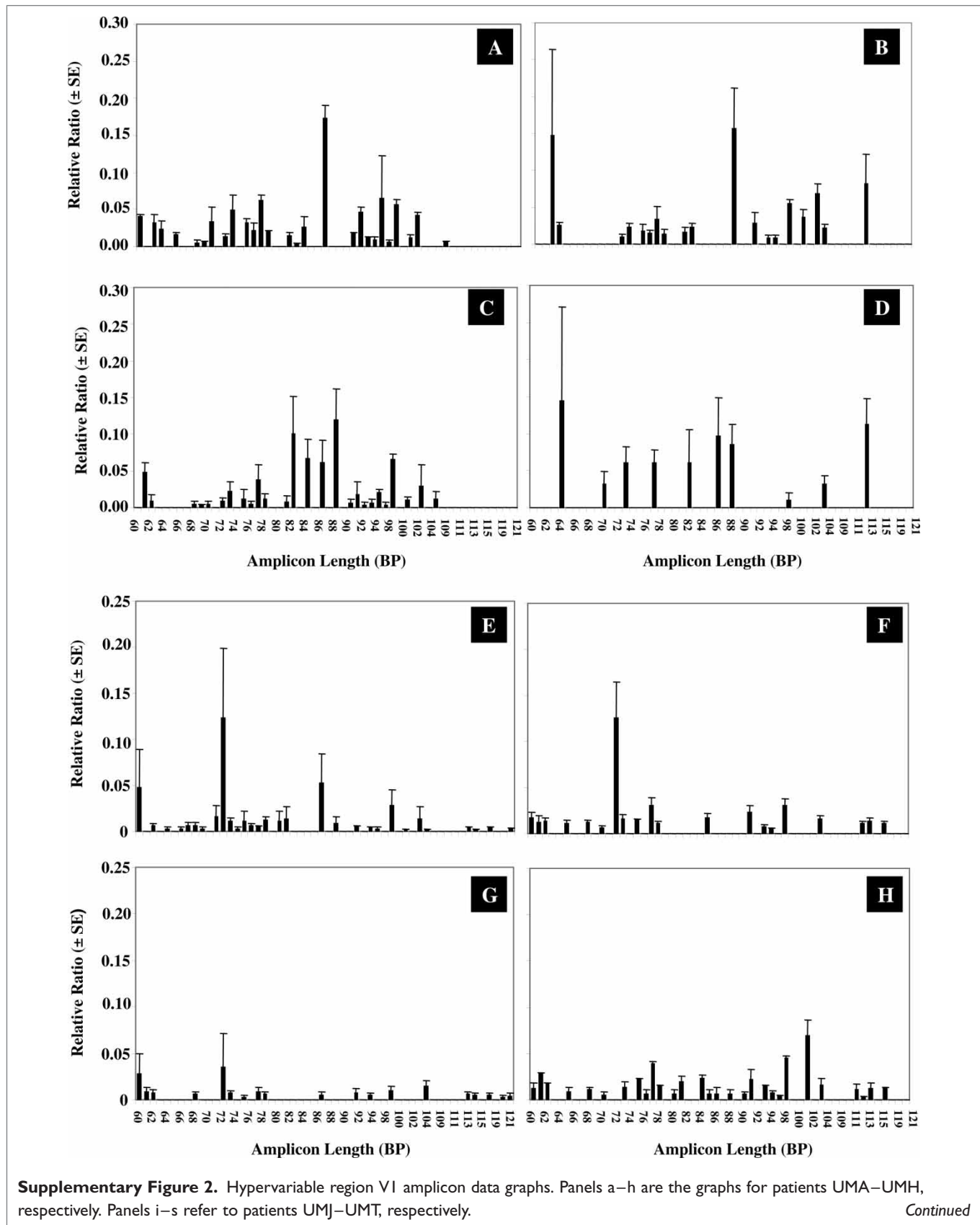


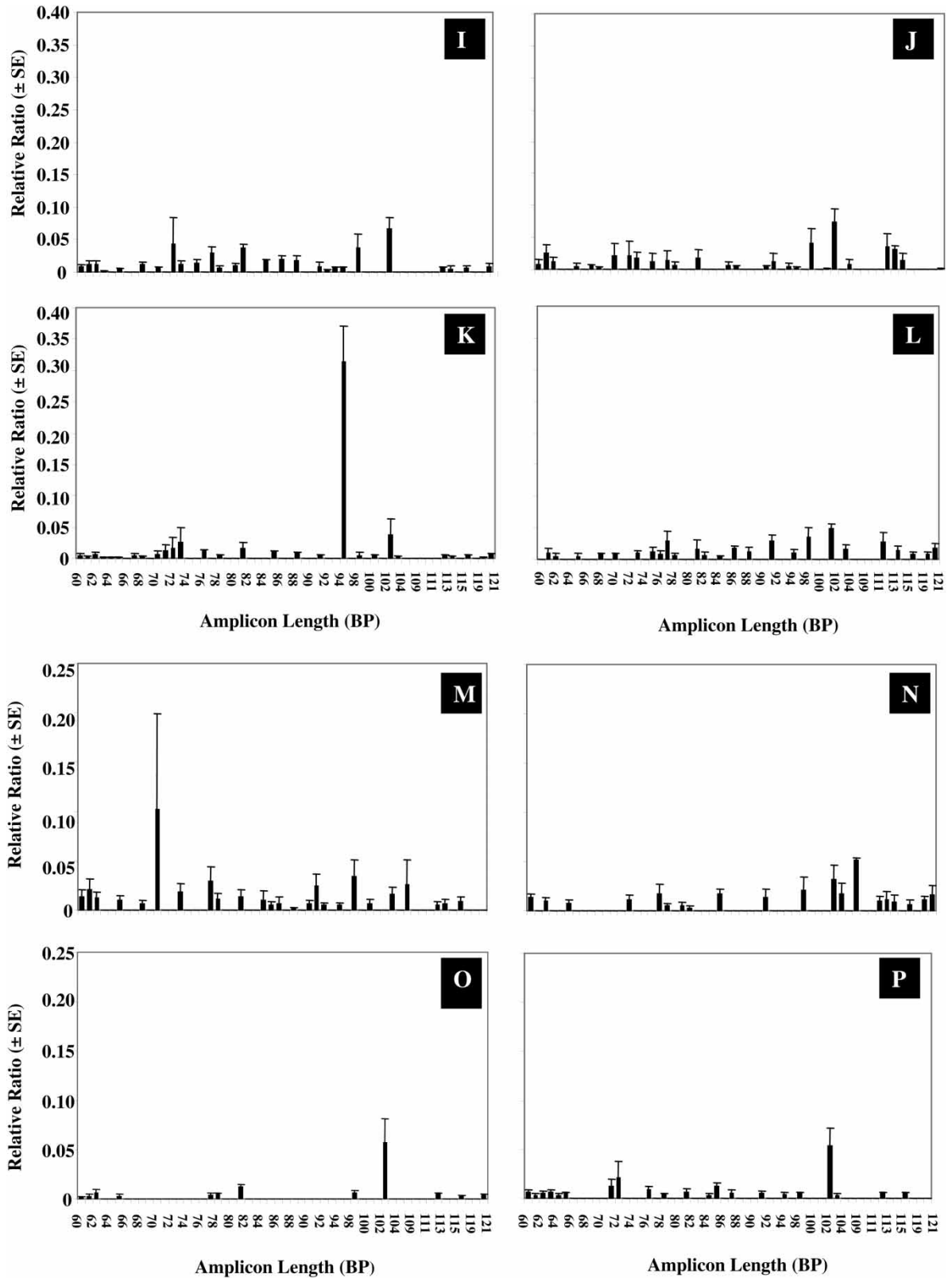
Supplementary Figure I. Hypervariable region VI_V2 amplicon data graphs. Panels a–e represent patients UMA–UME. Panels f–i refer to patients UML, UMM, UMN and UMO, respectively. Panels j, k and l represent patients UMQ, UMR and UMS, respectively. Panel m represents profiles from patients UMF, UMG, UMH, UMJ, UMK and UMT.

Continued



Supplementary Figure 1.





Supplementary Figure 2.

Continued

