

## Application of Metagenomic Techniques: Understanding the Unrevealed Human Microbiota and Explaining the in Clinical Infectious Diseases

Wonyong Kim\*

*Department of Microbiology, College of Medicine, Chung-Ang University, Seoul, Korea*

Uncultured microorganisms comprise the majority of the planet's biological diversity. In many environments, as many as 99% of the microorganisms cannot be cultured by standard techniques, and the uncultured fraction includes diverse organisms that are only distantly related to the cultured ones. Therefore, culture-independent methods are essential to understand the genetic diversity, population structure, and ecological roles of the majority of microorganisms. Recently, new techniques for studying microbial communities, collectively called metagenomics, have been developed to overcome the limitations of culturing. This review assesses the potential of metagenomic techniques to analyze the relative abundance of microbial species under varying human environmental conditions and to discover infectious causes of unexplained human diseases.

**Key Words:** Metagenomics, Sanger sequencing, Next-generation DNA sequencing

### INTRODUCTION

Metagenomics is the genomic analysis of microorganisms by direct extraction and cloning of DNA from their natural environment. For the purpose of this mini-review, the terms 'unculturable' and 'as yet uncultivated' are used to describe organisms that have yet not grown on artificial media *in vitro*. It is well established that only approximately 1% of bacteria on Earth can be readily cultivated *in vitro* - the so called 'great plate count anomaly' - based on the observation that microscopic counts are considerably larger than the equivalent total viable counts (1~3). There are currently estimated to be 61 distinct bacterial phyla, of which 31 have no cultivable representatives (4). The topology of the

archaeal phylogenetic tree remains uncertain, but it is clear that the 54 species of Archaea cultured to date represent only a fraction of the total diversity, with 49 lineages mostly uncultured (5).

Because the majority of bacteria and Archaea remain unculturable, the diversity of complex bacterial communities is inevitably underestimated using standard cultivation methods. Furthermore, organisms of key importance to the community and the entire ecosystem or pathogens of plants and animals may be overlooked if they are unculturable. Consequently, with the development of molecular culture-independent techniques, there has been a move towards the characterization of mixed bacterial populations within biomass from the environment and in samples from animals (including humans) using PCR amplification of house-

Received: August 31, 2012/ Revised: October 18, 2012/ Accepted: October 22, 2012

\*Corresponding author: Wonyong Kim. Department of Microbiology, College of Medicine, Chung-Ang University, 84, Heukseok-ro, Dongjak-gu, Seoul 156-656, Korea.

Phone: +82-2-820-5685, Fax: +82-2-822-5685, e-mail: kimwy@cau.ac.kr

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>).

keeping genes, particularly the 16S rRNA gene, followed by cloning for purification and sequencing for identification (6, 7).

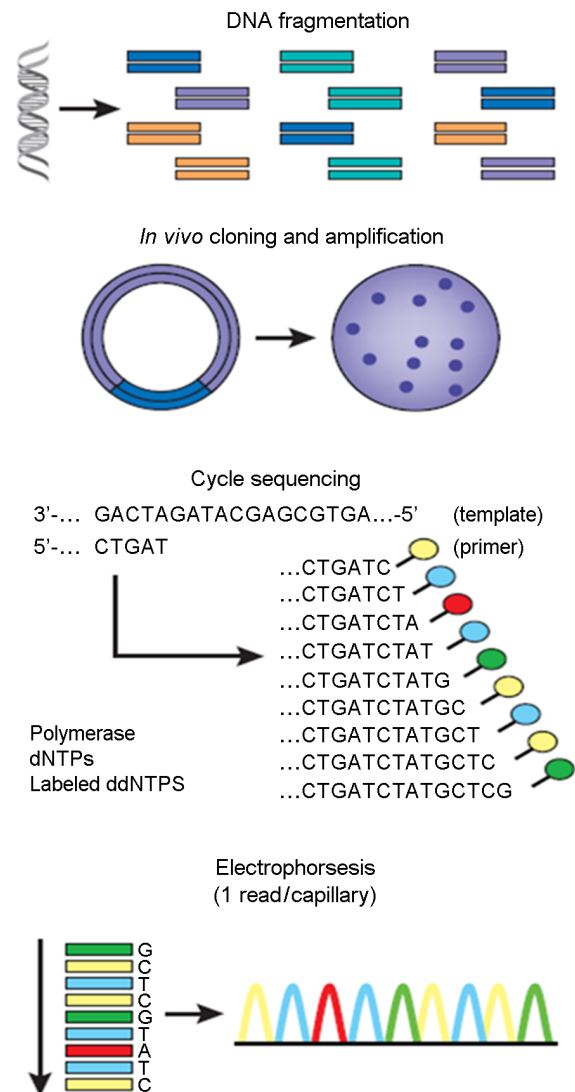
As a result, numerous novel phylotypes have been identified among bacterial communities isolated from a wide range of habitats: from seawater and soil to the health- and disease-associated microbiota of humans (8~10). Recently, 454 pyrosequencing has been used to investigate bacterial communities by sequencing 16S rRNA gene amplicons; examples of such studies include a study of nasopharyngeal microbiota in children (11), metagenomic analysis of the upper respiratory microbiota in pandemic influenza A (12), metagenomic analysis of human diarrhea (13), and analysis of the human distal gut microbiome (14, 15).

In the next section, I discuss Sanger-sequencing and next-generation DNA sequencing platforms. In addition, I highlight some applications of metagenomic techniques in the field of medical microbiology.

## 1. Sequencing Methods in Metagenomic Research

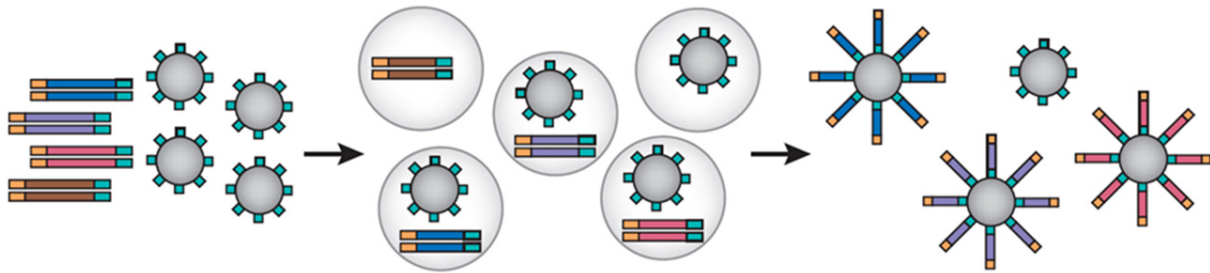
### 1-1. Sanger sequencing

Since the early 1990s, DNA sequencing has been performed almost exclusively using capillary-based, semi-automated implementations of Sanger biochemistry (16~18) (Fig. 1). In high-throughput production pipelines, the DNA to be sequenced is prepared by one of two approaches. For shotgun *de novo* sequencing, randomly fragmented DNA is cloned into a high-copy-number plasmid, which is then transformed into *Escherichia coli*. Alternatively, for targeted resequencing, PCR amplification is carried out with primers that flank the target. The output of both approaches is an amplified template, either in the form of many 'clonal' copies of a single plasmid insert present within a spatially isolated bacterial colony that can be picked, or as many PCR amplicons present within a single reaction volume. The sequencing biochemistry takes place in a 'cycle sequencing' reaction, in which cycles of template denaturation, primer annealing, and primer extension are performed. The primer is complementary to the known sequence immediately flanking the region of interest. Each round of primer



**Figure 1. In high-throughput shotgun Sanger sequencing, genomic DNA is fragmented, then cloned into a plasmid vector and transformed into *E. coli*.** For each sequencing reaction, a single bacterial colony is picked and plasmid DNA isolated. Each cycle sequencing reaction takes place within a microliter-scale volume, generating a ladder of ddNTP-terminated, dye-labeled products that are subjected to high-resolution electrophoretic separation within one of 96 or 384 capillaries in one run of a sequencing instrument. As fluorescently labeled fragments of discrete sizes pass a detector, the four-channel emission spectrum is used to generate a sequencing trace (26).

extension is stochastically terminated by the incorporation of fluorescently labeled di-deoxynucleotides (ddNTPs). In the resulting mixture of end-labeled extension products, the label on the terminating ddNTP of any given fragment corresponds to the nucleotide identity of its terminal position.



**Figure 2. Clonal amplification of sequencing features.** The 454, Polonator, and SOLiD platforms rely on emulsion PCR20 to amplify clonal sequencing features. In brief, an *in vitro*-constructed adaptor flanked shotgun library (shown as gold and turquoise adaptors flanking unique inserts) is PCR amplified (that is, multi-template PCR, not multiplex PCR, as only a single primer pair is used, corresponding to the gold and turquoise adaptors) in the context of a water-in-oil emulsion. One of the PCR primers is tethered to the surface (5'-attached) of micron-scale beads that are also included in the reaction. A low template concentration results in most bead-containing compartments having either zero or one template molecule present. In productive emulsion compartments (where both a bead and template molecule are present), PCR amplicons are captured on the surface of the bead. After breaking the emulsion, beads bearing amplification products are selectively enriched. Each clonally amplified bead has PCR products on its surface corresponding to amplification of a single molecule from the template library (26).

Sequence is determined by high-resolution electrophoretic separation of the single-stranded, end-labeled extension products in a capillary-based polymer gel. The reported error rates of Sanger sequencing range from 0.001% (19, 20) to more than 1% (21, 22) and seem to depend on the software that is used for post-processing of the reads. Pyrosequencing, also known as "454 sequencing", produces shorter reads (23, 24). When the technology was first introduced, the read length was about 110 nt, but it has now increased to more than ~450 nt. Huse *et al.* (2007) reported an error rate of 0.49% for 100~200 nt-long reads (25), while the read simulation software MetaSim (22) produced reads with an error rate of 2.8% with parameters based on the original 454 publication (24). Pyrosequencing is still actively being researched and optimized (26).

### 1-2. Next-generation DNA sequencing

The current sequencing revolution is being driven by three commercially available platforms: the 454 (Roche), Genome Analyzer (Illumina/Solexa), and ABI-SOLiD (Applied Biosystems) (27, 28). These new technologies are based on different principles than the classical Sanger-based method (29), and these methods are collectively referred to as either 'next-generation' sequencing (NGS), 'high throughput' sequencing, 'ultra-deep' sequencing, or 'massively parallel' sequencing. The 454 system was the first NGS platform

available as a commercial product (24) (Fig. 2). Using this approach, libraries can be constructed using any method that gives rise to a mixture of short and adaptor-flanked fragments. Clonal sequencing features are generated by emulsion PCR (30) with amplicons captured on the surface of 28  $\mu\text{m}$  beads. During sequencing, one side of the semi-ordered array functions as a flow cell for introducing and removing sequencing reagents, whereas the other side is bonded to a fiber-optic bundle for charge coupled device (CCD)-based signal detection. During each of several hundred cycles, a single species of unlabeled nucleotide is introduced. On templates where this results in an incorporation event, a pyrophosphate is released. Due to the activity of ATP sulfurylase and luciferase, incorporation events immediately drive the generation of a burst of light, which is detected by the CCD as corresponding to the array coordinates of specific wells. In contrast with other platforms, therefore, this sequencing by synthesis must be monitored by the camera does not move relative to the array. Across multiple cycles, the pattern of detected incorporation events reveals the sequence of templates represented by individual beads. The sequencing is 'asynchronous' in that some features may get ahead or behind other features depending on their sequence relative to the order of base addition. A major limitation of the 454 technology is related to homopolymers. Because there is no terminating moiety

that prevents multiple consecutive incorporations in a given cycle, the length of all homopolymers must be inferred from the signal intensity. In contrast, the Sanger method is based on DNA synthesis on a single-stranded template and di-deoxy chain termination (31, 32). Improvements in the cost-effectiveness of references and the development of high-throughput techniques have enabled direct sequencing of clones without laborious prior screening by restriction analysis (18). Relative to other next-generation platforms, the key advantage of the 454 platform is read length. For example, the 454 FLX instrument generates ~400,000 reads of 200 to 300 bp per instrument-run. Currently, the per-base cost of sequencing with the 454 platform is much greater than that of other platforms (e.g., SOLiD: Applied Biosystems and Solexa: Illumina) but it may be the method of choice for certain applications where long read lengths are critical (26).

## **2. Metagenomic Approaches For Discovering Novel Microorganisms**

In this section, I highlight the utility of using metagenomic approaches to explore human microbial diversity and the potential diagnostic value of a metagenomic approach.

### **2-1. Determination of the diversity of bacterial communities (Bacteriome) using sanger sequencing technique**

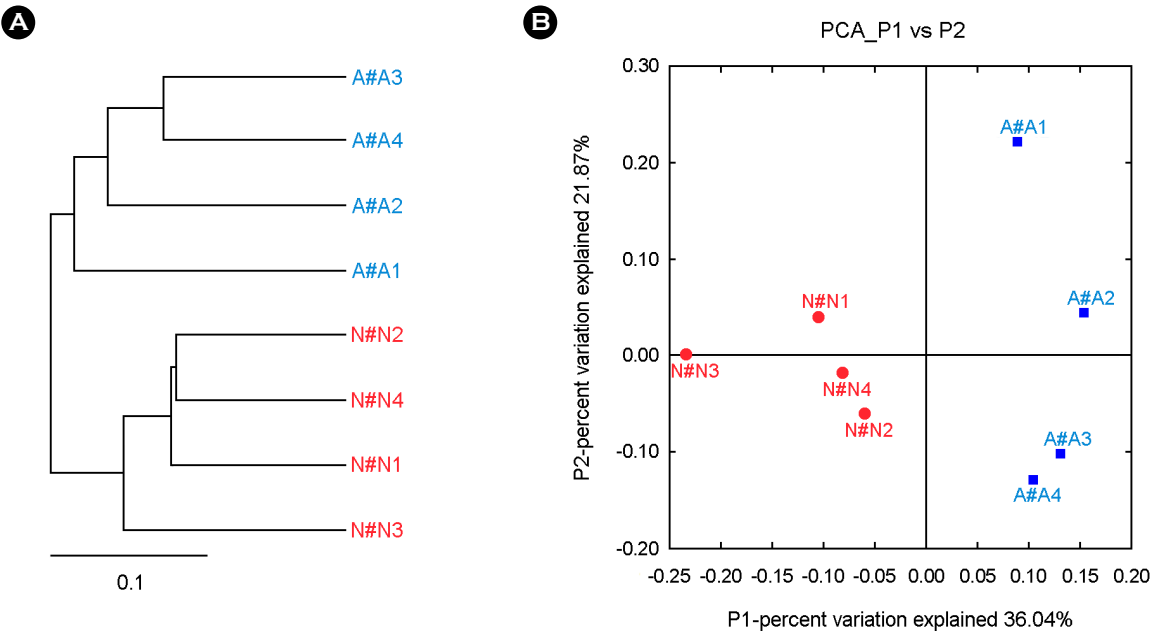
#### **2-1-1. Background**

Asthma is a heterogeneous airway inflammation syndrome with an increasing incidence in populations around the world. Its heterogeneity has made it difficult to determine pathogenesis and reasons for frequent exacerbation. Chronic infections cause a decline in lung function and may be a risk factor for the development and triggering of bronchial asthma. The association between respiratory infections and atopy or asthma remains to be determined and may be highly dependent on different infectious agents (33). The clinical relevance of the microbiota present on the oropharyngeal wall is unknown. However, it is known the diversity of the oral microflora can be altered by the use of antibiotics, probiotics, diet, and the microbiota in the gut. Changes in

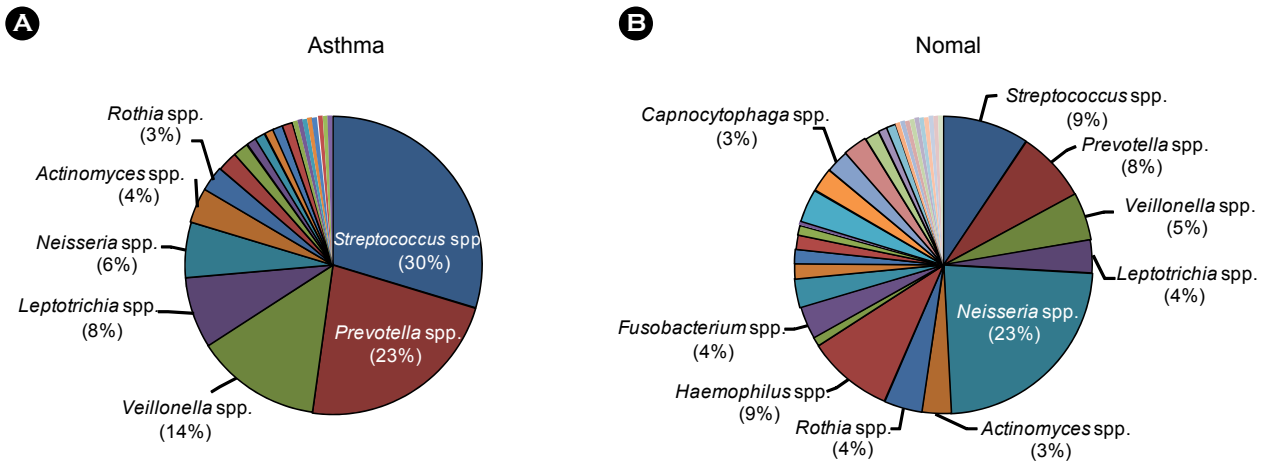
the microbiota of the GI tract can affect mucosal immunity in the lungs, as well as the gut, by bystander suppression (34). The link between oral and airway tolerance may be related to the induction or suppression of systemic immune tolerance due to challenge by antigens. Therefore, defining the characteristics of the oropharyngeal microbiota may increase our understanding of the pathogenic roles of these microorganisms and aid in the discovery of novel therapies for the prevention and treatment of oral complications (35). To define the difference in the oropharyngeal microbiota between patients with bronchial asthma and non-asthmatic controls, the 16S rRNA gene from bacterial species collected from healthy bronchial and asthma samples was analyzed by sequence analysis.

#### **2-1-2. Comparison of bacterial microbiota between asthma patients and non-asthmatic controls**

The diversity of the oropharyngeal swab sample microbiomes was characterized by sequencing the V3 region of the 16S rRNA gene that was amplified from genomic DNA samples of four asthma patients and four controls. About 50 clones were sequenced from each sample. Based on an analysis of 376 16S rRNA clones, the diversity of the microflora from the bronchial asthmatic and non-asthmatic control groups was striking; a total of 39 different bacterial genera representing six different bacterial phyla were documented in the non-asthmatic control samples. The majority of sequences in the metagenome library from the control group had *Neisseria* spp. (23.3%), *Prevotella* spp. (7.8%) and *Streptococcus* spp. (9.3%) according to GenBank, Eztaxon, and RDP database analysis tools. In contrast, a total of 21 different bacterial taxa were present in the samples from the group with bronchial asthma. The majority of sequences in the bronchial asthma library were identified as *Streptococcus* spp. (29.5%), *Prevotella* spp. (22.4%), *Neisseria* spp. (6.0%), and *Veillonella* spp. (13.7%). To examine the diversity of the microbial community present in oropharyngeal swab samples, assembled sequences were assigned to bacterial genera based on the taxonomy of their closest relatives as judged by Blastn analysis. While the two data sets showed broad taxonomic similarity, there was notable variation in each sample at the species level. It should be



**Figure 3.** Comparison of the microbial diversity in the oropharynx of bronchial asthma versus non-asthmatic persons.



**Figure 4.** UniFrac analysis of V3 16S sequences of the microbial community of oropharynx samples from bronchial asthma and non-asthmatic persons. (A) A jackknifed clustering of the environments in the UniFrac dataset. The numbers next to the nodes represent the number of times that particular node was observed in a random sampling from the whole dataset. (B) Principal component analysis scatter plot of the microbiota of individuals with bronchial asthma and healthy individuals (non-asthmatic persons, red; bronchial asthma, blue).

noted, however, that this analysis was coverage-limited. To correlate the variation in bacteria between swab samples, the Blastn hits were classified according to genera. Fig. 3 shows that the bronchial asthma samples were dominated by *Streptococcus* spp., *Prevotella* spp., *Neisseria* spp., and *Veillonella* species. The estimated diversity in the affected patients was 72.6%, whereas in the control group it was

46.6%. There were a higher number of bacterial genera represented in the normal control than in the bronchial asthma samples, including *Granulicatella* spp. (3.6%), *Campylobacter* spp. (2.6%), *Capnocytophaga* spp. (2.6%), *Xylanibacter* spp. (2.6%), *Sphingomonas* spp. (1.6%), *Phocoenobacter* spp. (1.0%), *Schlegelella* spp. (1.0%), *Atopobium* spp. (0.5%), *Catonella* spp. (0.5%), *Dialister*

spp. (0.5%), *Enterobacter* spp. (0.5%), *Gemella* spp. (0.5%), *Kaistella* spp. (0.5%), *Microvirgula* spp. (0.5%), *Novosphingobium* spp. (0.5%), *Oscillibacter* spp. (0.5%), and *Parvimonas* spp. (0.5%). The presence of several genera in the bronchial asthma oropharynx samples separated these samples from the non-asthmatic control samples. For instance, *Bergeriella* spp., *Megasphaera* spp., *Paraprevotella* spp., *Phycicola* spp., *Pilibacter* spp. and *Streptobacillus* spp. were detected in the bronchial asthma oropharynx samples but were absent from the non-asthmatic control samples.

### 2-1-3. UniFrac and statistical analysis

Jackknife clustering of environments (Fig. 4A), which is similar to principle component analysis (PCA), showed fairly robust clustering (75% bootstrap support for all nodes but one) in the bronchial asthma and non-asthmatic samples. Comparison of individual samples using Fast UniFrac PCA (Fig. 4B) showed a distinct clustering by dietary treatment when DNA samples from both groups were included in the analysis. Statistical analysis showed significant differences in the frequencies of phyla between the bronchial asthma and non-asthmatic samples (Table 1). At the genus level, *Neisseria* spp. ( $p = 0.0168$ ), *Alcaligenes* spp., *Atopobium* spp., *Campylobacter* spp., *Catonella* spp., *Dialister* spp., *Enterobacter* spp., *Fusobacterium* spp., *Gemella* spp., *Granulicatella* spp., *Haemophilus* spp., *Hallella* spp., *Microvirgula* spp., *Novosphingobium* spp., *Oscillibacter* spp., *Parvimonas* spp., *Porphyromonas* spp., *Pseudomonas* spp., *Rothia* spp., *Schlegelella* spp., *Sphingomonas* spp., *Xylanibacter* spp., and TM7 were abundant in the non-asthmatic individuals. In contrast, *Streptococcus* spp. and *Anaerosporebacter* spp. were predominant in the asthmatic patients at a significant level ( $p < 0.0001$ ).

## 2-2. Diversity of fungal communities (Mycobiome) using next-generation sequencing

This study has been published in the PLOS ONE (36).

### 2-2-1. Background

Unlike classical seborrheic dermatitis, dandruff is a non-inflammatory condition of the scalp that is characterized by scaling and is considered to be a form of mild seborrheic dermatitis. Dandruff is a common scalp disorder that affects

almost half of the postpubertal population, regardless of ethnicity and gender, and has several putative causes including non-microbial and microbial factors (37, 38). Potential non-microbial causes of dandruff are excessive exposure to sunlight, irritation of the scalp due to over-shampooing, frequent combing, use of certain cosmetic products, and exposure to dust and dirt, although experimental evidence is lacking (37, 39, 40). The microbial etiopathology that is most widely accepted is the presence of a lipophilic yeast belonging to the genus *Malassezia* (37, 41). The scalp has a biotic community of which the known components are *Staphylococcus* spp., *Propionibacterium* spp., and *Malassezia* spp. (42). Dandruff-afflicted scalps have been reported to have 1.5-fold to 2-fold higher levels of *Malassezia* than non dandruff-afflicted scalps (43). *Malassezia* species are known to be the causative pathogens of pityriasis versicolor, seborrheic dermatitis (SD), and atopic dermatitis (AD) (44, 45). The genus *Malassezia* comprises at least seven species, each of which has specific ecological, biochemical, and genetic characteristics. Some *Malassezia* species (*M. furfur*, *M. gobosa*, and *M. restricta*) are associated with various human infections, but the pathological role of individual species is not yet fully understood (45~48). To investigate the fungal biota associated with dandruff, 26S rRNA gene sequences from fungal species collected from basal (healthy) scalp samples and dandruff-afflicted scalp samples were analyzed using the GS-FLX Titanium sequencer (Roche) for massively parallel ribosomal RNA gene amplicon pyrosequencing.

### 2-2-2. Comparison of fungal microbiota in the human scalp between dandruff-afflicted and non-dandruff-afflicted individuals

A total of 74,811 sequence reads of the 26S rRNA gene were obtained using the Genome Sequencer FLX Titanium platform (Table 2). Statistical analysis revealed that three phyla and 11 genera identified by sequencing and BLAST analysis had significantly different frequencies in healthy scalps and dandruff-afflicted scalps (Table 3). At the phylum level, Basidiomycota were isolated from dandruff-afflicted scalps twice as often as from healthy scalps. Sequences derived from fungi in the genera *Acremonium*, *Coniochaeta*,

**Table 1.** Statistics analysis of differences in the microbes present in bronchial asthma and non-asthmatic persons

Phylum	Genus	Average (%)		Estimate	Standard Error	Chi-square	p-value
		Asthma	Normal				
Firmicutes	<i>Streptococcus</i> spp.	29.5	9.3	1.5626	0.2801	31.12	<.0001
Bacteroidetes	<i>Prevotella</i> spp.	22.4	7.8	–	–	–	–
Firmicutes	<i>Veillonella</i> spp.	13.7	5.2	–	–	–	–
Fusobacteria	<i>Leptotrichia</i> spp.	7.7	3.6	–	–	–	–
Proteobacteria	<i>Neisseria</i> spp.	6	23.3	-0.8193	0.3428	5.71	0.0168
Actinobacteria	<i>Actinomyces</i> spp.	3.8	3.1	–	–	–	–
Actinobacteria	<i>Rothia</i> spp.	2.7	4.2	-0.9276	0.7105	1.7	0.1917
Proteobacteria	<i>Haemophilus</i> spp.	2.2	9.3	-1.0978	0.5758	3.63	0.0566
Firmicutes	<i>Oribacterium</i> spp.	1.6	1	–	–	–	–
Fusobacteria	<i>Fusobacterium</i> spp.	1.1	3.6	-0.5628	0.8079	0.49	0.486
Proteobacteria	<i>Pseudomonas</i> spp.	1.1	3.1	-0.4131	0.8235	0.25	0.6159
Proteobacteria	<i>Alcaligenes</i> spp.	1.1	1.6	-0.08	0.9787	0.01	0.9349
Bacteroidetes	<i>Hallella</i> spp.	1.1	1.6	0.3347	0.9153	0.13	0.7146
TM7	Unclassified genus	1.1	1.6	0.2567	0.925	0.08	0.7814
Bacteroidetes	<i>Porphyromonas</i> spp.	0.6	1	0.0454	1.2281	0	0.9705
Firmicutes	<i>Anaerosporebacter</i> spp.	0.6	0.5	0.7422	1.4176	0.27	0.6006
Proteobacteria	<i>Bergeriella</i> spp.	0.6	0	–	–	–	–
Firmicutes	<i>Megasphaera</i> spp.	0.6	0	–	–	–	–
Bacteroidetes	<i>Paraprevotella</i> spp.	0.6	0	–	–	–	–
Actinobacteria	<i>Phycicola</i> spp.	0.6	0	–	–	–	–
Firmicutes	<i>Pilibacter</i> spp.	0.6	0	–	–	–	–
Fusobacteria	<i>Streptobacillus</i> spp.	0.6	0	–	–	–	–
Firmicutes	<i>Granulicatella</i> spp.	0	3.6	–	–	–	–
Proteobacteria	<i>Campylobacter</i> spp.	0	2.6	–	–	–	–
Bacteroidetes	<i>Capnocytophaga</i> spp.	0	2.6	–	–	–	–
Bacteroidetes	<i>Xylanibacter</i> spp.	0	2.6	–	–	–	–
Proteobacteria	<i>Sphingomonas</i> spp.	0	1.6	–	–	–	–
Proteobacteria	<i>Phocoenobacter</i> spp.	0	1	–	–	–	–
Proteobacteria	<i>Schlegelella</i> spp.	0	1	–	–	–	–
Actinobacteria	<i>Atopobium</i> spp.	0	0.5	–	–	–	–
Firmicutes	<i>Catonella</i> spp.	0	0.5	–	–	–	–
Firmicutes	<i>Dialister</i> spp.	0	0.5	–	–	–	–
Proteobacteria	<i>Enterobacter</i> spp.	0	0.5	–	–	–	–
Firmicutes	<i>Gemella</i> spp.	0	0.5	–	–	–	–
Bacteroidetes	<i>Kaistella</i> spp.	0	0.5	–	–	–	–
Proteobacteria	<i>Microvirgula</i> spp.	0	0.5	–	–	–	–
Proteobacteria	<i>Novosphingobium</i> spp.	0	0.5	–	–	–	–
Proteobacteria	<i>Oscillibacter</i> spp.	0	0.5	–	–	–	–
Firmicutes	<i>Parvimonas</i> spp.	0	0.5	–	–	–	–

**Table 2.** Numbers of sequence reads generated by the GS-FLX Titanium sequencer

Sample	Dandruff Status <sup>a</sup>	No. of sequence reads				No. of reads used for analysis
		Raw seq	Seq with blast hit	(+) strain <sup>b</sup>	(-) strain <sup>b</sup>	
G1_B1	Normal	12,353	12,257	6,552	5,705	10,735
G1_B2	Normal	11,238	11,132	5,353	5,779	10,029
G1_B3	Normal	13,473	13,418	7,144	6,274	12,119
G3_B6	Normal	11,281	11,222	5,787	5,435	9,614
G4_B4	Normal	10,280	10,183	5,593	4,590	9,092
G5_B5	Patient	8,373	8,344	4,496	3,848	7,589
G5_B7	Patient	7,813	7,782	3,869	3,913	7,034
Total		74,811	74,338	38,794	35,544	66,212

<sup>a</sup>"Normal" status refers to a healthy scalp while "patient" refers to an individual with a dandruff-afflicted scalp.

<sup>b</sup>(+), positive strand; (-), reverse complementary strand.

**Table 3.** Fungal classifications and differences between normal and dandruff-afflicted scalps

Classification level	Name	% average		Estimate	Standard error	<i>p</i> -value
		Normal	Patient			
Phylum	Ascomycota	82.57	61.30	0.2979	0.0093	<.0001
	Basidiomycota	13.12	35.70	-1.0011	0.0178	<.0001
	Uncultured_fungus	4.31	3.00	0.3627	0.0413	<.0001
Genus	<i>Acremonium_spp.</i>	61.77	57.91	0.042	0.0102	<.0001
	<i>Coniochaeta_velutina</i>	0.49	0.00	5.0654	1.0032	<.0001
	<i>Cryptococcus_spp.</i>	12.31	0.26	3.8312	0.109	<.0001
	<i>Didymella_spp.</i>	20.22	0.05	6.0094	0.2503	<.0001
	<i>Eupenicillium_spp.</i>	0.01	0.04	-1.8498	0.7596	0.0149
	<i>Filobasidium_floriforme</i>	0.20	33.85	-5.1703	0.1274	<.0001
	<i>Malassezia_spp.</i>	0.07	1.91	-3.376	0.2218	<.0001
	<i>Penicillium_spp.</i>	0.05	3.44	-4.3043	0.2599	<.0001
	<i>Rhodotorula_spp.</i>	0.84	0.05	2.8899	0.2654	<.0001
	Uncultured ascomycete	1.92	0.02	4.3478	0.3559	<.0001
	Uncultured soil fungus	2.14	2.47	-0.1667	0.0521	0.0014

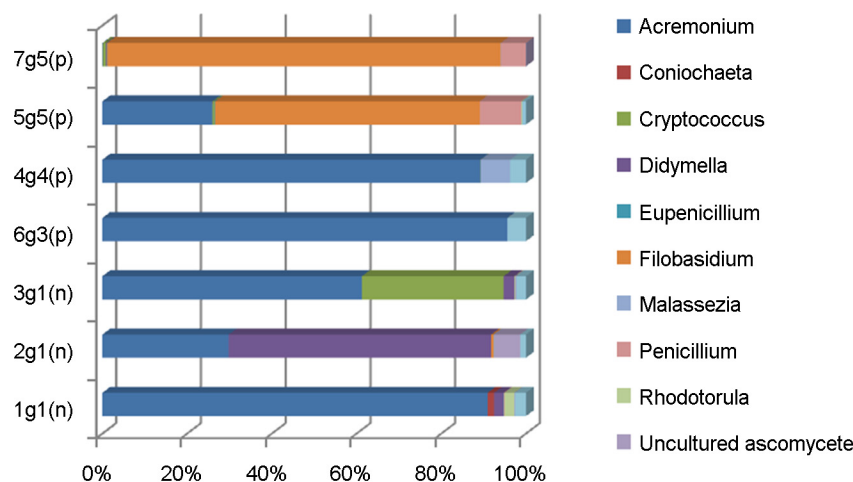
*Cryptococcus*, *Didymella*, *Rhodotorula*, and an uncultured *Ascomycete* were found in greater frequency in libraries generated from healthy scalp samples. In contrast, *Eupenicillium*, *Filobasidium*, *Malassezia*, and *Penicillium* were found significantly more often in samples from dandruff-afflicted scalps ( $p < 0.0001$ ) along with an uncultured soil fungus ( $p = 0.0014$ ). The most dominant fungal phylum in

samples from the dandruff-afflicted scalps was Ascomycota, which accounted for 94.31%, 91.23%, 65.00%, 95.01%, 88.32%, 37.21%, and 6.28% of the fungal population of the seven swab samples, respectively, while Basidiomycota accounted for 2.32%, 1.55%, 32.26%, 0.02%, 7.06%, 61.66%, and 93.49% of the fungal population of each of the seven swab samples, respectively (Fig. 5A). A total of





**Figure 5. Diagram of mycobiome communities at the phylum level.** (A) Distribution of mycobiomes between normal scalps and dandruff-afflicted scalps. (B) Distribution of genera belonging to the phylum Ascomycota in healthy and dandruff-afflicted scalps. (C) Distribution of genera belonging to the phylum Basidiomycota in healthy and dandruff-afflicted scalps.



**Figure 6.** Diagram of the fungal communities at the generic level. Comparison of genus frequencies in the mycobiomes of normal and dandruff-afflicted scalps.

53 species in the phylum Ascomycota were detected in the samples overall. Thirty-seven species were identified on healthy scalps and 24 were identified on dandruff-afflicted scalps. Eighteen species of *Basidiomycota* were detected on healthy scalps and 11 species were identified on dandruff-afflicted scalps. *Acremonium* spp. (phylum Ascomycota) was the most dominant fungus on both healthy and dandruff-afflicted scalps. On healthy scalps, *Acremonium* spp. and *Didymella bryoniae* comprised over 95% (72% and 23%, respectively) of the Ascomycota, which corresponds to >80% of the fungal community of healthy scalps compared to just over 60% in dandruff-afflicted scalps. In dandruff-afflicted scalps, >90% of the Ascomycota were *Acremonium* spp., and several *Penicillium* spp., including *P. chrysogenum* (2%) and *P. meleagrimum* (3%) were also present (Fig. 5B). In the phylum Basidiomycota, populations of *Cryptococcus* spp. observed on healthy scalps were replaced with *Filobasidium* (94%) on dandruff-afflicted scalps (Fig. 5C).

### 2-2-3. Discovery of an unusual fungus, *Filobasidium floriforme*

*Filobasidium floriforme* was present at a high frequency on the two scalps graded as having the most severe dandruff condition (5g5; 62.5% and 7g5, 92.9%; see Fig. 6). Even though their total frequency was lower, *Penicillium* species (*P. chrysogenum* and *P. meleagrimum*) showed the same tendency. Among the samples from healthy scalps, only 3g1 had a high frequency (32%) of species in the genus

*Cryptococcus*, including *C. albidosiilis* (1.72%), *C. diffluens* (12.81%), and *C. liquefaciens* (14.7%). Several *Malassezia* spp., which are involved in pityriasis versicolor, SD, and AD as discussed previously (49), were detected in 2g1 and 4g4 scalps, including *M. globosa*, *M. restricta*, *M. samuelsii*, and one non-culturable species. *Malassezia* spp. were present in healthy scalp samples at a frequency of 0.07% and in dandruff-afflicted samples at a frequency of 2%. *Filobasidium floriforme* belongs to the phylum Filobasidiales. The Filobasidiales are an order in the fungal class Tremellomycetes. The one family in this order, the *Filobasidiaceae*, contains five species, namely *F. capsuligenum* ATCC 14437<sup>T</sup>, *F. uniguttulatum* ATCC 32047<sup>T</sup>, *F. elegans*, *F. floriforme* and *F. globisporum* (50). *Filobasidium* spp. have not been studied intensively in the context of human disease. Only one species, *F. uniguttulatum*, has been reported to cause cryptococcal meningitis (51). *F. uniguttulatum* is a teleomorphic fungus that was first isolated from an infected human nail in 1934 (52) and named *Eutorulopsis uniguttulata*. In 1952, it was renamed *Cryptococcus neoformans* var. *uniguttulatus* on the basis of morphological and physiological similarities with *C. neoformans*, but with reduced capsule formation<sup>s</sup>. Phylogenetically, *F. uniguttulatum* is more closely related to *C. albidus* than to *C. neoformans* (53, 54). The results of the dandruff study suggested that the major fungus associated with dandruff is *F. uniguttulatum* rather than *Malassezia* spp. Therefore,

control of *Filobasidium* spp. could potentially be an alternative strategy for treating dandruff.

## SUMMARY

Metagenomic analysis is a valuable technique for studying microbial communities. In this review, I highlighted the potential applications of two types of metagenomic approaches to identify and detect causative organisms of unexplained diseases. These approaches involved the genomic analysis of microorganisms by Sanger sequencing and 454 pyrosequencing. These studies revealed the distribution and profile of bacteria and fungi associated with asthma and dandruff, respectively, thereby contributing to defining the microbial community associated with these disease states. The studies discussed above highlight how a metagenomic approach can be used to discover unknown microorganisms that may cause idiopathic infectious diseases.

## REFERENCES

- 1) Staley JT, Konopka A. Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* 1985;39:321-46.
- 2) Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 1995;59:143-69.
- 3) Hugenholtz P, Goebel BM, Pace NR. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 1998;180:4765-74.
- 4) Hugenholtz P, Hooper SD, Kyrpides NC. Focus: Synergistetes. *Environ Microbiol* 2009;11:1327-9.
- 5) Auguet JC, Barberan A, Casamayor EO. Global ecological patterns in uncultured Archaea. *ISME J* 2010;4:182-90.
- 6) Giovannoni SJ, Britschgi TB, Moyer CL, Field KG. Genetic diversity in Sargasso Sea bacterioplankton. *Nature* 1990;345:60-3.
- 7) Pace NR. A molecular view of microbial diversity and the biosphere. *Science* 1997;276:734-40.
- 8) Rappé MS, Giovannoni SJ. The uncultured microbial majority. *Annu Rev Microbiol* 2003;57:369-94.
- 9) Zhou X, Bent SJ, Schneider MG, Davis CC, Islam MR, Forney LJ. Characterization of vaginal microbial communities in adult healthy women using cultivation-independent methods. *Microbiology* 2004;150:2565-73.
- 10) Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. Defining the normal bacterial flora of the oral cavity. *J Clin Microbiol* 2005;43:5721-32.
- 11) Bogaert D, Keijsers B, Huse S, Rossen J, Veenhoven R, van Gils E, *et al.* Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PLoS One* 2011;6:e17035.
- 12) Greninger AL, Chen EC, Sittler T, Scheinerman A, Roubinian N, Yu G, *et al.* A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PLoS One* 2010;5:e13381.
- 13) Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, Wang D. Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog* 2008;4:e1000011.
- 14) Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* 2006;312:1355-9.
- 15) Bae JW. Recent Methodological Approaches to Human Microbiome. *J Bacteriol Virol* 2011;41:1-7.
- 16) Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, *et al.* Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 1977;265:687-95.
- 17) Swerdlow H, Wu SL, Harke H, Dovichi NJ. Capillary gel electrophoresis for DNA sequencing. Laser-induced fluorescence detection with the sheath flow cuvette. *J Chromatogr* 1990;516:61-7.
- 18) Hunkapiller T, Kaiser RJ, Koop BF, Hood L. Large-scale and automated DNA sequence determination. *Science* 1991;254:59-67.
- 19) Noguchi H, Park J, Takagi T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* 2006;34:5623-30.
- 20) Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998;8:175-85.
- 21) Keith CS, Hoang DO, Barrett BM, Feigelman B, Nelson

- MC, Thai H, *et al.* Partial sequence analysis of 130 randomly selected maize cDNA clones. *Plant Physiol* 1993;101:329-32.
- 22) Richter DC, Ott F, Auch AF, Schmid R, Huson DH. MetaSim: a sequencing simulator for genomics and metagenomics. *PLoS One* 2008;3:e3373.
  - 23) Ronaghi M, Uhlén M, Nyrén P. A sequencing method based on real-time pyrophosphate. *Science* 1998;281:363, 365.
  - 24) Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, *et al.* Genome sequencing in micro-fabricated high-density picolitre reactors. *Nature* 2005;437:376-80.
  - 25) Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol* 2007;8:R143.
  - 26) Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol* 2008;26:1135-45.
  - 27) Rusk N, Kiermer V. Primer: Sequencing--the next generation. *Nat Methods* 2008;5:15.
  - 28) Schuster SC. Next-generation sequencing transforms today's biology. *Nat Methods* 2008;5:16-8.
  - 29) Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 1977;74:5463-7.
  - 30) Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A* 2003;100:8817-22.
  - 31) Hall N. Advanced sequencing technologies and their wider impact in microbiology. *J Exp Biol* 2007;210:1518-25.
  - 32) Pettersson E, Lundberg J, Ahmadian A. Generations of sequencing technologies. *Genomics* 2009;93:105-11.
  - 33) Ramsey CD, Gold DR, Litonjua AA, Sredl DL, Ryan L, Celedón JC. Respiratory illnesses in early life and asthma and atopy in childhood. *J Allergy Clin Immunol* 2007;119:150-6.
  - 34) Huffnagle GB. The microbiota and allergies/asthma. *PLoS Pathog* 2010;6:e1000549.
  - 35) Noverr MC, Huffnagle GB. The 'microflora hypothesis' of allergic diseases. *Clin Exp Allergy* 2005;35:1511-20.
  - 36) Park HK, Ha MH, Park SG, Kim MN, Kim BJ, Kim W. Characterization of the fungal microbiota (mycobiome) in healthy and dandruff-afflicted human scalps. *PLoS One* 2012;7:e32847.
  - 37) Piérard-Franchimont C, Xhauflaire-Uhoda E, Piérard GE. Revisiting dandruff. *Int J Cosmet Sci* 2006;28:311-8.
  - 38) Ranganathan S, Mukhopadhyay T. Dandruff: the most commercially exploited skin disease. *Indian J Dermatol* 2010;55:130-4.
  - 39) Jo JH, Jang HS, Ko HC, Kim MB, Oh CK, Kwon YW, *et al.* Pustular psoriasis and the Kobner phenomenon caused by allergic contact dermatitis from zinc pyrithione-containing shampoo. *Contact Dermatitis* 2005;52:142-4.
  - 40) Piérard GE, Piérard-Franchimont C. Squamometry in acute photodamage. *Skin Research and Technology* 1995;1:137-9.
  - 41) Gupta AK, Madzia SE, Batra R. Etiology and Management of Seborrheic Dermatitis. *Dermatology* 2004;208:89-93.
  - 42) Leyden JJ, McGinley KJ, Kligman AM. Role of microorganisms in dandruff. *Arch Dermatol* 1976;112:333-8.
  - 43) McGinley KJ, Leyden JJ, Marples RR, Kligman AM. Quantitative microbiology of the scalp in non-dandruff, dandruff, and seborrheic dermatitis. *J Invest Dermatol* 1975;64:401-5.
  - 44) Guého E, Boekhout T, Ashbee HR, Guillot J, Van Belkum A, Faergemann J. The role of *Malassezia* species in the ecology of human skin and as pathogens. *Med Mycol* 1998;36:220-9.
  - 45) Sugita T, Takashima M, Shinoda T, Suto H, Unno T, Tsuboi R, *et al.* New yeast species, *Malassezia* dermatis, isolated from patients with atopic dermatitis. *J Clin Microbiol* 2002;40:1363-7.
  - 46) Erchiga VC, Florencio VD. *Malassezia* species in skin diseases. *Curr Opin Infect Dis* 2002;15:133-42.
  - 47) Gupta AK, Kohli Y, Faergemann J, Summerbell RC. Epidemiology of *Malassezia* yeasts associated with pityriasis versicolor in Ontario, Canada. *Med Mycol* 2001;39:199-206.
  - 48) Prohić A. Identification of *Malassezia* species isolated from scalp skin of patients with psoriasis and healthy subjects. *Acta Dermatovenerol Croat* 2003;11:10-6.

- 49) Sugita T, Suto H, Unno T, Tsuboi R, Ogawa H, Shinoda T, *et al.* Molecular analysis of *Malassezia* microflora on the skin of atopic dermatitis patients and healthy subjects. *J Clin Microbiol* 2001;39:3486-90.
- 50) Kirk PM, Cannon PF, Minter DW, JA S, editors. *Dictionary of the Fungi*: 10th ed. Wallingford: CABI; 2008.
- 51) Pan W, Liao W, Hagen F, Theelen B, Shi W, Meis JF, *et al.* Meningitis caused by *Filobasidium uniguttulatum*: case report and overview of the literature. *Mycoses* 2012;55:105-9.
- 52) Kwon-Chung KJ. Perfect state of *Cryptococcus uniguttulatus*. *Int J Syst Bacteriol* 1977;27:293-9.
- 53) Kwon-Chung KJ, Boekhout T, Wickes B, Fell J. Systematics of the Genus *Cryptococcus* and its type species *C. neoformans*. In: Heitman J, Kozel TR, Kwon-Chung KJ, Perfect JR, Casadevall A, editors. *Cryptococcus*: From human pathogen to model yeast. Washington, DC: ASM press; 2011. pp. 3-16.
- 54) Fell JW, Boekhout T, Fonseca A, Scorzetti G, Statzell-Tallman A. Biodiversity and systematics of basidiomycetous yeasts as determined by large-subunit rDNA D1/D2 domain sequence analysis. *Int J Syst Evol Microbiol* 2000;50:1351-71.
-