

Invited Mini Review

Probing the diversity of healthy oral microbiome with bioinformatics approaches

Ji-Hoi Moon* & Jae-Hyung Lee*

Department of Maxillofacial Biomedical Engineering, School of Dentistry, and Department of Life and Nanopharmaceutical Sciences, Kyung Hee University, Seoul 02447, Korea

The human oral cavity contains a highly personalized microbiome essential to maintaining health, but capable of causing oral and systemic diseases. Thus, an in-depth definition of "healthy oral microbiome" is critical to understanding variations in disease states from preclinical conditions, and disease onset through progressive states of disease. With rapid advances in DNA sequencing and analytical technologies, population-based studies have documented the range and diversity of both taxonomic compositions and functional potentials observed in the oral microbiome in healthy individuals. Besides factors specific to the host, such as age and race/ethnicity, environmental factors also appear to contribute to the variability of the healthy oral microbiome. Here, we review bioinformatic techniques for metagenomic datasets, including their strengths and limitations. In addition, we summarize the interpersonal and intrapersonal diversity of the oral microbiome, taking into consideration the recent large-scale and longitudinal studies, including the Human Microbiome Project. [BMB Reports 2016; 49(12): 662-670]

INTRODUCTION

The human microbiota (the collection of microbes that live on and inside us) consists of a wide range of microorganisms whose aggregate membership exceeds the human somatic and germ cells by at least an order of magnitude (1, 2). The collection of genes in the microbiota is called the human microbiome (2); however, "microbiota" and "microbiome" are often used interchangeably (3). One of the most clinically relevant microbial habitats, the human oral cavity is colonized

by a personalized set of microorganisms, including bacteria, archaea, fungi, and viruses (4). Under healthy conditions, the oral microbiota lives in harmony with the host, similar to other body sites. The host provides an environment wherein the microbiome flourish, in turn keeping their host healthy (5). Conversely, the oral microbiome is also considered a key cause of oral diseases, including dental caries and periodontal diseases, as well as many systemic diseases such as diabetes and cardiovascular diseases (5, 6). Because of its crucial role in oral and systemic health, the oral microbiome has become an essential part of microbiomics.

An in-depth definition of a healthy microbiome is an indispensable step toward detecting significant variations in disease states and pre-clinical conditions, as well as understanding the disease onset and progression (7). The advent of next generation sequencing (NGS) or high-throughput sequencing has revolutionized the field of microbiome analysis, providing the tools necessary to address the issue (8). This prompted the launch of the NIH's Human Microbiome Project (HMP), constructed as a large, genome-scale community research project (NIH HMP Working Group, 2009). Over 200 healthy adults were enrolled, and samples were collected from 15 to 18 body sites, including oral, stool, skin, nasal, and vaginal areas, over a period of 1 to 3 visits (9). Besides two major scientific reports (9, 10), several companion papers analyzed the HMP oral datasets (7, 11-13), revealing great variability of the oral microbiome among and within healthy individuals. Furthermore, other recent large-scale and longitudinal studies have augmented our view of the oral microbiome, beyond that of the HMP.

In this paper, we review bioinformatic techniques for metagenomic datasets, including microbial community profiling, and highlight the strengths and weaknesses of the experimental approaches. We also summarize important findings that lead to the current understanding of the the range of healthy microbial diversity. Although viruses, fungi, archaea and protozoa form a part of the normal microbiome (4), majority of the research is concentrated on the domain Bacteria. Therefore, we will focus exclusively on oral bacteria in this review.

*Corresponding authors. Jae-Hyung Lee, Tel: +82-2-961-9290; Fax: +82-2-962-0598; E-mail: jaehlee@khu.ac.kr, Ji-Hoi Moon, Tel: +82-2-961-0795; Fax: +82-2-962-0598; E-mail: prudence75@khu.ac.kr

<https://doi.org/10.5483/BMBRep.2016.49.12.164>

Received 23 September 2016

Keywords: Bioinformatics, Diversity, Human Microbiome Project (HMP), Next-generation Sequencing (NGS), Oral microbiome

BIOINFORMATIC ANALYSIS OF MICROBIOME SEQUENCE DATA

Two distinct metagenomics approaches are commonly used: marker gene metagenomics and full shotgun metagenomics. Marker gene metagenomics is a fast and cost-effective way to obtain a taxonomic distribution profile. In this approach, specific regions of evolutionarily conserved marker genes are first amplified by PCR, and subsequently sequenced (14). In the case of bacterial (and/or archaeal) community analysis, the target region usually contains the 16S ribosomal RNA (rRNA) gene (15). Hence, the approach is referred to as 16S rRNA profiling. Full shotgun metagenomics, also referred as metagenomic whole genome sequencing (WGS), does not target a specific locus or marker gene, but instead breaks the isolated metagenomic DNA into smaller fragments, and subsequently sequences the individual pieces (14). The sequenced small fragments (i.e., sequencing raw reads) can be used for taxonomy profiling (who is there?) as well as for functional profiling (what are they doing?) (14). In this section, we briefly describe the scheme of the techniques involved and the bioinformatic pipelines, to analyze microbiome sequence data obtained from the above two methods.

16S rRNA profiling

The 16S rRNA gene was introduced as a marker for bacterial phylogeny by Woese *et al.* (16). Ever since, it has been the gold standard for phylogenetic analysis of microbial communities and bacterial taxonomy (11). Bacterial 16S rRNA genes usually include nine hypervariable regions (V1-V9) that exhibit substantial sequence diversity among different bacterial species (17). Numerous studies have assessed the 16S rRNA gene and selected the most appropriate conserved regions to generate amplicons using universal primers, as well as identified the most effective hypervariable regions to be targeted (17-23). Unfortunately, no single hypervariable region is sufficiently different to define all bacterial species and a bias can be introduced by primer specificity as well as efficiency. Basically, the 16S rRNA profiling can be summarized into three steps: (1) preprocessing and denoising of raw reads, (2) taxonomic assignment, and (3) evaluation of microbial diversity.

(1) Preprocessing and denoising of raw sequencing reads:

Although there are standard operations and protocols to generate the NGS sequencing libraries, stochastic errors in the biological processes for the library creation, and/or incomplete chemical reactions in sequencing, could affect the overall quality of the sequencing library and sequencing datasets. Therefore, raw sequencing reads generated should be carefully checked for the successful downstream analysis in the pre-processing step. A number of computational tools have been used for the preprocessing: FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) provides a quick quality check by running a modular set of analyses such as “per base sequence quality”, “per sequence quality score”, “sequence length distribution”, “adapter content”, etc.; FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) allows detecting and trimming the low quality region of the individual read (especially 3'-end of the reads); DUST is used to remove low-complexity regions in the sequencing read (24). Intrinsically, NGS techniques can harbor various errors in the sequencing reads, such as imprecise signals from longer homopolymer runs and chimera sequences. In the denoising step, these errors were identified and corrected for the accurate taxonomic assignments of the sequencing reads. Many popular software, such as QIIME (25) and mothur (26), have implemented the denoising algorithms. In particular, UCHIME is designed to detect chimeric sequences by comparing reference sequences to a database, or by performing *de novo* classification (clustering) (27). Preprocessed and denoised raw sequencing reads are subsequently subject to the process of taxonomic assignment.

(2) Taxonomic assignment: NGS allows investigators to detect and identify novel bacteria that have previously gone undetected. Subsequently, the assignment of 16S rRNA read sequences originated from uncultured bacterial genome into a specific bacterial taxonomy is even more difficult. In two frequently used methods, the reads are assigned into bins, according to either homology between the reads and known reference sequences (i.e., phylotyping) or homology between the reads (i.e., operational taxonomic units [OTUs]) (28). The former method relies upon aligning reads with the reference 16S rRNA database using sequence alignment algorithms, such as BLAST (29). Besides NCBI Genbank, a number of rRNA databases have been constructed and used for the taxonomic assignment (Table 1). Each database has its own

Table 1. A list of 16S ribosomal RNA database

Name	16S rRNA coverage	Database URL (reference)
CORE	Human Oral Bacteria	http://microbiome.osu.edu/ (32)
RDP	Archaea and Bacteria	https://rdp.cme.msu.edu/ (33)
Human Oral Microbiome Database	Human Oral Bacteria	http://www.homd.org/index.php (65)
EzTaxon-e	Archaea and Bacteria	http://www.ezbiocloud.net/eztaxon (95)
SILVA	Archaea and Bacteria	https://www.arb-silva.de/ (96)
Greengenes	Archaea and Bacteria	http://greengenes.secondgenome.com/ (97)

criteria for the curation of data from the original resources. For example, the Human Oral Microbiome Database (HOMD) (30) and CORE (31) have been constructed using 16S rRNA sequences exclusively from human oral bacteria. The second approach is to group 16S rRNA sequencing reads into bins called OTUs with distance-based agglomerative clustering methods, such as CD-HIT (32) and UCLUST (33). Defining species by 97% identity in 16S rRNA gene sequence is a commonly used criterion, but these distinctions are still controversial (11, 34).

NGS platforms generate massively greater number of reads compared to the classical Sanger sequencing, while the reads are relatively much shorter. Unfortunately, current databases and methods are not able to assign all species names or provide enough phylogenetic information for the billions of sequence reads (11). For example, the most commonly used tool for assigning taxonomy, the Ribosomal Database Project (RDP) Classifier (35), does not assign taxonomic names below the genus level (11, 36). Moreover, as revealed in our previous study, the RDP shows insufficient resolution for classifying the GN02 and *Synergistetes* (37). To supplement the limited phylogenetic information, 16S rRNA sequences are usually grouped using the OTU approach described above. Huse *et al.* (11) explored the HMP oral microbiota from over 200 individuals, and identified between 857 and 4,216 OTUs (Table 2). In terms of community membership, oral microbiota was especially diverse, showing the highest estimate of total richness after stool microbiota. Notably, richness as measured by the V1-V3 primers, was consistently higher than that measured by V3-V5 (11). In addition, some taxa (e.g. Lactobacilli OTUs) resolved better with V1-V3 while others (e.g. Bifidobacteriaceae OTUs) were better resolved with

V3-V5 (11). These differences may be due to a mismatch of the primers for amplification, or an inability to differentiate the taxon in that region of the rRNA gene (11). Therefore, as with all 16S rRNA sequencing projects, diversity measurements should be compared with other results using the same 16S rRNA gene region, and the presence of primer bias should not be ignored (11). Additionally, sequencing errors will also affect the taxonomic assignments of the sequence reads, possibly resulting in spurious OTUs and exaggerated diversity estimates, thus making direct comparisons between studies problematic (12).

(3) Evaluation of microbial diversity: To understand the structure and dynamics of microbial community, the measurement of diversity is essential. Two diversity measurements are frequently used to assess and compare microbial communities: alpha (or within-sample) diversity and beta (or between-sample) diversity. Alpha diversity is usually characterized using the total number of organisms within a sample (richness, which may be measured as the number of OTUs), the relative abundances of the organisms (evenness), or indices that combine these two dimensions. In contrast, beta diversity, is often characterized using the number of species (or OTUs) shared between two communities. In particular, UniFrac, a robust method for comparing the differences between microbial communities between samples, measures the proportion of shared branch lengths on a phylogenetic tree between samples (3, 38). Principal Coordinates Analysis (PCoA) can summarize and visualize the UniFrac distances between samples in a scatterplot where points (representing samples) that are more distant from one another are more dissimilar.

Table 2. Counts of patients included, OTUs and estimated richness (number of species) found for both the V1-V3 and the V3-V5 regions (11)

Body Site	V1-V3			V3-V5		
	Patients	OTUs	Estimated richness ^a	Patients	OTUs	Estimated richness ^a
Buccal mucosa	114	2,025	6,635	198	898	4,650
Hard palate	112	1,741	3,793	190	912	3,125
Keratinized gingiva	117	1,545	4,387	206	857	3,352
Palatine Tonsils	119	3,683	10,023	204	1,633	9,020
Saliva	99	2,341	6,546	181	1,399	6,801
Subgingival plaque	119	4,216	14,410	204	1,672	11,501
Supragingival plaque	121	3,851	11,154	205	1,587	8,254
Throat	110	2,343	5,601	192	1,136	4,154
Tongue dorsum	119	3,651	7,910	205	1,503	7,947
Posterior fornix ^b	59	428	1,151	95	400	1,466
Stool ^b	118	6,050	23,665	209	5,391	33,627

^aUpper and lower confidence limits are not included in this table. ^bExample of extraoral sites. The stool samples have the highest estimate of total richness, followed by the oral samples, particularly the plaque and tonsils. The skin sites, such as posterior fornix, have the lowest estimated richness.

Metagenomic WGS data analysis

Although the 16S rRNA profiling is a powerful, effective and straightforward technique to study microbial communities, it only provides the taxonomic composition. The metagenomic WGS data can provide not only taxonomy, but also the biological functional profiles for the microbial communities. The principles of taxonomy profiling processes that employ WGS data are similar to those described above. This section will therefore focus on the functional profiling of the microbial community. The analysis pipeline can be divided into four stages: (1) preprocessing, (2) reconstruction of raw sequencing reads (assembly), (3) gene prediction, and (4) functional annotations.

(1) Preprocessing: Preprocessing assesses the overall quality of WGS data, and most steps are similar to 16S rRNA profiling. Additionally, raw metagenomic NGS reads associated with a host (e.g. human) are checked for host DNA contamination, and the contaminated sequencing reads are removed. Fast short read mapping tools, such as BWA (39) and Bowtie 2 (40), are used to detect the contaminated sequencing reads by aligning raw sequencing reads against the host genome (e.g. human genome).

(2) Reconstruction of raw sequencing reads (assembly): The metagenomic WGS technique generates raw sequencing reads from the whole microbial genomes in the microbial community. Thus, to identify the specific genomes and/or complete protein coding genes in the genomes accurately, it is helpful to reconstruct the microbial genomes from raw sequencing reads. However, obtaining complete genomes has been challenging not only because of the highly repetitive DNA sequences abundant in a broad range of species (from bacteria to mammals), but also because of short reads and high data volumes produced by NGS technology. Therefore, an assembly of shorter reads into genomic contigs and their orientation into scaffolds is often performed. Most of the metagenomic WGS read assembly tools are designed and implemented based on the de Bruijn graph theory algorithm. Initially, all sequencing reads are fragmented into k-mers, following which they are used as the edges in the de Bruijn graph. The nodes of (k-1)-mer prefix and suffix are linked by the edges of k-mers for the graph. Finally, the assembler identifies the Eulerian paths that go across all edges just once in the graph (41). Velvet (42), ABySS (41) and SOAPdenovo (44) use the de Bruijn graph to assemble whole metagenomes from raw sequencing reads. In the HMP, the raw sequencing reads from 749 metagenomic samples were successfully used to assemble contigs using an optimized SOAPdenovo protocol (8). Recently, more sophisticated algorithms have been developed and applied to the next-generation assemblers, such as Meta-IDBA (45), MetaVelvet-SL (46) and IDBA-UD (47).

(3) Gene prediction: Following reconstruction, next stage is to identify genes in the reads or assembled contigs and/or scaffolds. The prediction of genes in metagenomic contents is still a fairly difficult problem, although several gene prediction

algorithms have been successfully employed for prokaryotic genomes. To predict genes in metagenomic studies, especially for *de novo* genes, several computational methods have been developed, including MetaGeneMark (48), MetaProdigal (49), Glimmer-MG (50), and FragGeneScan (51). Notably, the performance of gene-predicting tools varies considerably: in a comparison of five widely used *ab initio* gene-calling algorithms including FragGeneScan and MetaGeneMark, FragGeneScan is rather accurate for predicting reading frames on short raw reads (75-1,000 bp) while other tools, such as MetaGeneMark, are better suited for higher-quality sequences such as assembled contigs (52). Moreover, it has been reported that combining predictions from various programs can improve the accuracy of prediction and annotation of metagenomic reads (53). Accordingly, researchers should carefully decide what tools to use in their metagenomic study, which potentially impacts the results and conclusion.

(4) Functional annotations: After gene prediction, the identified genes are functionally annotated by comparing the known genes in the functional annotation databases such as PFAM (54), IMG/M (55), COG (56) and MetaRef (57). Further analysis of the relationship between the microbiome and the host phenotype is performed using metabolic pathway information database, i.e. KEGG (58), eggNOG (59) and MinPath (60). In the part of the HMP, Abubucker *et al.* devised HMP Unified Metabolic Analysis Network (HUMANn) to construct metabolic networks of the microbial communities (61). In this study, raw sequencing reads were searched against protein sequence databases, and HUMANn recovers the abundances of individual orthologues gene families and pathway. More specifically, MBLASTX, KEGG orthology and MinPath have been used to assign genes and available pathways. Recently, several metagenomic analysis pipeline software have been developed, such as MG-RAST (62) and IMG/MER (<https://img.jgi.doe.gov/cgi-bin/mer/main.cgi>). The pipelines provide the functional annotation modules in their fully automated pipeline web-server and thus, researchers can easily perform functional annotation tasks using their own data in the web (15).

COMPOSITION AND DIVERSITY OF ORAL MICROBIOME

The HMP assessed oral microbiome composition of seven intra oral sites (buccal mucosa, hard palate, keratinized gingiva, saliva, sub- and supra gingival plaque, and tongue dorsum) and two oropharyngeal sites (throat and palatine tonsils) from 182-206 healthy subjects (18 to 40 years old). A total of 185-322 genera belonging to 13-19 bacterial phyla were discovered (13). The dominant phyla were *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Fusobacteria* and *Actinobacteria*, accounting for over 95% of the entire oral microbiome. An individual sample from a single site of a single subject contained 23-50 genera from 6-9 phyla (13). Among all body

habitats, the oral habitats have the highest alpha diversity showing the highest OTU level richness, after the stool samples (Table 2), while the skin and vaginal microbiota show lower alpha diversity (11, 13). Comparing samples from the same sites among subjects (beta diversity), the oral sites have the lowest beta diversities, signifying that members of the population shared relatively similar organisms in oral sites than in other body sites (10). However, the HMP oral datasets also emphasize the high variance across individuals, especially at the sub-genus level: even OTUs existing in almost every subject, or dominating in some samples, showed variation in relative abundance (11). In the following sections, we discuss in detail about the specific factors that contribute to the variability of the healthy oral microbiome.

Different oral biogeographic niches

The oral cavity is a humid environment which is kept at a fairly consistent temperature (34°C to 36°C) and a relatively neutral pH in most areas, and thus provide great condition for the growth of various microorganisms (63). The oral cavity is composed of diverse habitats with different anatomical structures and physicochemical factors. The oral mucosa covers the cheek, tongue, gingiva, palate, and floor of the mouth and allows rapid elimination of adhering bacteria due to a continuous shedding of its surface epithelial cells (63). On the other hand, papillary surface of the tongue provides shelter for adhering bacteria and protects these bacteria from mechanical cleaning. The hard surface of teeth offers many sites for bacterial colonization, in both supra- and subgingival areas. The gingival crevice (area between the junctional epithelium of the gingiva and teeth), provides a distinctive microbial colonization site, consisting of both hard and soft tissues (63). The epithelium may be keratinized (palate) or nonkeratinized (gingival crevice). Hence, the oral cavity is not considered a uniform environment.

The HMP revealed a substantial divergence in the species richness and evenness among different oral habitats, and also identified microorganisms with specific niche preferences. Hard palate showed the lowest total richness, however the gingival plaque showed the highest total richness (11) (Table 2). Oral sites, especially saliva, have the highest evenness, while buccal mucosa and keratinized gingiva have lower alpha diversity than the other oral sites (10, 13). Each oral habitat in almost every subject was characterized by one or a few signature taxa making up the plurality of the community with highly variable relative abundance among both the individuals and the oral habitats. Most oral habitats are dominated by *Streptococcus*, followed in abundance by *Haemophilus* in the buccal mucosa, *Actinomyces* in the supragingival plaque, and *Prevotella* in the subgingival plaque (10, 13). There is an overlap of species detected in almost all oral sites, such as certain species of *Streptococcus* (OTUs #2, 5 and 6), *Gemella* (OTUs #7 and 8), *Granulicatella* (OTU #13), *Fusobacterium* (OTUs #9 and 27), and *Veillonella*

(OTUs #4 and 7) (11). However, several abundant genera had multiple OTUs with distinct preferences for often only one or two oral sites, such as *Bacteroides*, *Prevotella*, *Corynebacterium*, *Fusobacterium*, *Pasteurella*, and *Neisseria* (11). For example, *Corynebacterium matruchotii* (OTU #15) existed almost only in the supragingival plaque, while *Corynebacterium argentoratense* (OTU #188) was mostly found in saliva and to a lesser extent on the hard palate (11). This may be due to the desquamation of the epithelial cells and the shear forces that are generated during chewing in the buccal fold and the hard palate (64). Analysis of oral samples collected from the elderly (range 73-93), *Lautropia mirabilis* was significantly associated with the supragingival plaque, while *Treponema socranskii* was found only in the subgingival plaque (65). This may be explained by the low oxidation-reduction potential of the subgingival plaque. In the oropharynx, the distribution of *Firmicutes*, *Proteobacteria*, and *Bacteroidetes* was similar to that in saliva, but had more *Proteobacteria* than found in the mouth (66).

Influence of geography, climate and ethnicity

Although the HMP produced a huge volume of data, the resulting 16S rRNA datasets are composed of samples from medical students in the USA, and host information is nearly prohibitive to access, which lead to removal of the potential to observe any systematic patterns and regional or ethnic differences (67). A population-scale study of 120 healthy individuals from 12 worldwide locations showed a significant variation in the saliva microbiome according to the locations (68). Notably, the saliva microbiome of Batwa Pygmies, a former hunter-gatherer group from Africa, was much more diverse than the saliva microbiome of two agricultural African groups, probably owing to their different lifestyle and diet (69). In another study of 3 human groups from different geographic and climatic areas (76 native Alaskans, 10 Germans and 66 Africans) the distinctiveness of the saliva microbiome was seen, the reasons of which (e.g. different lifestyles and/or host genetics and physiology) remain to be clarified (70). In the study, alpha diversity was highest for the German group and lowest for the African group, while the opposite was true for beta diversity. It is intriguing to speculate that the higher population density of Germany may provide more opportunities for bacteria to be spread among individuals (70).

Ethnicity is likely to exert a selection pressure on the oral microbiome. Mason *et al.* (71) analyzed dental plaque and saliva samples collected from 192 subjects belonging to four ethnic affiliations (non-Hispanic blacks, non-Hispanic whites, Chinese, and Latinos) and found obvious ethnicity-specific clustering of microbial communities, thus reinforcing prior observations (72-74). It seems that this selection pressure is most likely genetic rather than environmental, because the two ethnicities that shared a common food, nutritional and lifestyle heritage (Caucasians and African Americans) showed substantial microbial diversity (71). It is known that not only

innate immune responses to infectious agents, but also tooth morphologies, vary according to ethnic affiliation (75-78). Therefore, ethnicity possibly contribute to the bacterial selection by affecting the environment for bacterial colonization (71).

Vertical and horizontal transmission

Vertical transmission from mother to child starts at birth (79). Depending on the delivery mode (vaginal or Caesarian), infants acquire bacterial communities similar to their mother's vaginal microbiota or skin microbiota (80). A study of healthy three-month-old infants delivered vaginally (25 infants) and born by C-section (38 infants) found differences in the infant's oral microbiota owing to the mode of delivery, with vaginally delivered infants having a higher taxonomic diversity (81). The method of feeding also affects the infant's microbiome: oral lactobacilli with antimicrobial properties were found in breast-fed infants but not found in formula-fed infants (82, 83). Horizontal transmission of oral microbiota among siblings and other individuals sharing the same environment also contributes to oral microbiome diversity. In a study, 264 saliva samples were collected from 107 individuals (including 45 twin pairs), at up to three time-points during a 10-year period, spanning adolescence. The twins resembled each other more closely than the whole population at all time-points, but became less similar to each other when they aged and no longer cohabited (84).

Temporal variation

Studies looking at the temporal variation of the oral microbiome have revealed conflicting results: in a longitudinal study of 5 adults at three time-points (from 5 to 29 days), the salivary microbial community appeared to be stable at different time points (85). The HMP consortium (10) and Zhou *et al.* (13) reported that, among 22 HMP habitats of the human body, the oral habitat has the most stable microbiota, showing the highest community similarity between two visits (mean time interval between visits is 212 days), while the skin and vaginal microbiota are less stable. In contrast, a reanalysis of the HMP datasets by a method for quantifying the difference between two cohorts revealed that the relative abundances of core OTUs in an individual sample showed significantly greater changes from 1st to 2nd visit at oral and stool body regions, as compared with the vaginal region (12). More recently, a weekly longitudinal study of 85 adults over 3 months, showed high levels of temporal variability in both diversity and community structure in tongue microbiome, as in other body habitats studied (86). Furthermore, this study found that the composition of an individual's microbiome as well as their degree of temporal variability, are personalized features. Collectively, although intrapersonal variation over time is lower than interpersonal variation, intrapersonal temporal dynamics need to be considered when attempting to link changes in microbiome structure to changes in health status (86).

Age-related changes

Along with a variety of physiological changes which accompanies aging, microbial habitats also greatly change in the oral cavity. The eruption of primary teeth and replacement of the primary dentition with permanent dentition may lead to shifts in the microbial community composition at different phases of life (87). Edentulous infants have been found to have lower diversity than their mothers or primary care givers in the oral microbial composition (88). In deciduous dentition, a higher proportion of Proteobacteria (*Gammaproteobacteria*, *Moraxellaceae*) was found than Bacteroidetes. With increasing age, Bacteroidetes (mainly genus *Prevotella*), *Veillonellaceae* family, Spirochaetes, and candidate division TM7 increased (89). Several organisms, including members of the genera *Veillonella*, *Actinomyces* and *Streptococcus*, were reported to have age-specific abundance profiles during adolescence (84). Xu *et al.* (87) analyzed the oral microbiome (saliva, supragingiva and mucosa) across a wide age range (3 days-76 years), in which only a very small overlap of shared OTU was observed. In this study, a distinct temporal shift was observed in the relative abundance of most genera. The average relative abundance of the dominant bacterial phyla, *Actinobacteria*, *Bacteroides*, *Firmicutes*, *Fusobacteria*, *Proteobacteria*, *Spirochetes* and candidate division TM7 varied by age/dentition stage (87).

CONCLUSIONS

We have only begun to understand the tremendous diversity of the oral microbiome and a number of challenges remain, such as the vast uncultivated species and the lack of reference genomes (90). Until recently, about half of all known bacterial phyla were identified only from their 16S rRNA gene sequences (91). In fact, the bacteria that can be grown in the laboratory are only a portion of the total diversity that exists in the oral cavity (92). One method to address this challenge is single-cell genomics, which is a powerful tool for accessing genetic information from uncultivated microorganisms (93). Future work combining metagenomics and single cell genomics, as well as advances in each separate method, should help to overcome these issues, providing new insights into the uncultivated lineages (94).

Rapidly developing sequencing methods and analytical techniques are enhancing our ability to understand the human microbiome, leading to the concept of a 'personal microbiome'. The focus now shifts from characterizing oral microbiota to functional studies encompassing genomics, transcriptomics, and metabolomics of both host and microbes. Future investigations will inevitably be personal omics profiling in order to probe the temporal patterns associated with both molecular changes and related physiological health and disease. This knowledge is vital for the development of efficacious prevention and treatment protocols for oral diseases and, ultimately, contribute to the development of personalized medicine and personalized dental medicine.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Science, ICT & Future Planning (NRF-2015R1C1A2A01054588) and the National Honor Scientist Program (NRF-2012R1A3A1050385). This work was also supported in part by 2016 Sabbatical Research Program of Kyung Hee University.

REFERENCES

1. Hamady M and Knight R (2009) Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res* 19, 1141-1152
2. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R and Gordon JI (2007) The human microbiome project. *Nature* 449, 804-810
3. Ursell LK, Metcalf JL, Parfrey LW and Knight R (2012) Defining the human microbiome. *Nutr Rev Suppl* 1, S38-44
4. Wade WG (2013) The oral microbiome in health and disease. *Pharmacol Res* 69, 137-143
5. Zarco MF, Vess TJ and Ginsburg GS (2012) The oral microbiome in health and disease and the potential impact on personalized dental medicine. *Oral Dis* 18, 109-120
6. He J, Li Y, Cao Y, Xue J and Zhou X (2015) The oral microbiome diversity and its relation to human diseases. *Folia Microbiol (Praha)* 60, 69-80
7. Segata N, Haake SK, Mannon P et al (2012) Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. *Genome Biol* 13, R42
8. Gevers D, Pop M, Schloss PD and Huttenhower C (2012) Bioinformatics for the Human Microbiome Project. *PLoS Comput Biol* 8, e1002779
9. Human Microbiome Project Consortium (2012b) A framework for human microbiome research. *Nature* 486, 215-221
10. Human Microbiome Project Consortium (2012a) Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207-214
11. Huse SM, Ye Y, Zhou Y and Fodor AA (2012) A core human microbiome as viewed through 16S rRNA sequences clusters. *PLoS One* 7, e34242
12. Li K, Bihan M and Methé BA (2013) Analyses of the stability and core taxonomic memberships of the human microbiome. *PLoS One* 8, e63139
13. Zhou Y, Gao H, Mihindukulasuriya KA et al (2013) Biogeography of the ecosystems of the healthy human body. *Genome Biol* 14, R1
14. Sharpton TJ (2014) An introduction to the analysis of shotgun metagenomic data. *Front Plant Sci* 5, 209
15. Oulas A, Pavloudi C, Polymenakou P et al (2015) Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform Biol Insights* 9, 75-88
16. Woese CR and Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74, 5088-5090
17. Mizrahi-Man O, Davenport ER and Gilad Y (2013) Taxonomic classification of bacterial 16S rRNA genes using short sequencing reads: evaluation of effective study designs. *PLoS One* 8, e53608
18. Wang Q, Garrity GM, Tiedje JM and Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73, 5261-5267
19. Huse SM, Dethlefsen L, Huber JA, Mark Welch D, Relman DA and Sogin ML (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet* 4, e1000255
20. Liu Z, DeSantis TZ, Andersen GL and Knight R (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Res* 36, e120
21. Nossa CW, Oberdorf WE, Yang L et al (2010) Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. *World J Gastroenterol* 16, 4135-4144
22. Claesson MJ, Wang Q, O'Sullivan O et al (2010) Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* 38, e200
23. Tremblay J, Singh K, Fern A et al (2015) Primer and platform effects on 16S rRNA tag sequencing. *Front Microbiol* 6, 771
24. Morgulis A, Gertz EM, Schäffer AA and Agarwala R (2006) A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J Comput Biol* 13, 1028-1040
25. Sevinsky JR, Turnbaugh PJ, Walters WA et al (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 7, 335-336
26. Schloss PD, Westcott SL, Ryabin T et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75, 7537-7541
27. Edgar RC, Haas BJ, Clemente JC, Quince C and Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194-2200
28. Schloss PD and Westcott SL (2011) Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol* 77, 3219-3226
29. Altschul SF, Madden TL, Schäffer AA et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-3402
30. Chen T, Yu WH, Izard J, Baranova OV, Lakshmanan A and Dewhirst FE (2010) The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. *Database (Oxford)* 2010:baq013
31. Griffen AL, Beall CJ, Firestone ND et al (2011) CORE: a phylogenetically-curated 16S rDNA database of the core oral microbiome. *PLoS One* 6, e19051
32. Fu L, Niu B, Zhu Z, Wu S and Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing

- data. *Bioinformatics* 28, 3150-3152
33. Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460-2461
 34. Schloss PD and Handelsman J (2005) Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. *Appl Environ Microbiol* 71,1501-1506
 35. Cole JR, Wang Q, Fish JA et al (2013) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res* 42, D633-642
 36. Kotamarti RM, Hahsler M, Raiford D, McGee M and Dunham MH (2010) Analyzing taxonomic classification using extensible Markov models. *Bioinformatics* 26, 2235-2241
 37. Moon JH, Lee JH and Lee JY (2015) Subgingival microbiome in smokers and non-smokers in Korean chronic periodontitis patients. *Mol Oral Microbiol* 30, 227-241
 38. Lozupone C and Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* 71, 8228-8235
 39. Li H and Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760
 40. Langmead B and Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-359
 41. Kim Y, Koh I and Rho M (2014) Deciphering the human microbiome using next-generation sequencing data and bioinformatics approaches. *Methods* 79-80, 52-59
 42. Zerbino DR and Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18, 821-829
 43. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ and Birol I (2009) ABySS: a parallel assembler for short read sequence data. *Genome Res* 19, 1117-1123
 44. Luo R, Liu B, Xie Y et al (2012) SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 1, 18
 45. Peng Y, Leung HC, Yiu SM and Chin FY (2011) Meta-IDBA: a *de Novo* assembler for metagenomic data. *Bioinformatics* 27, i94-101
 46. Afiahayati, Sato K and Sakakibara Y (2015) MetaVelvet-SL: an extension of the Velvet assembler to a *de novo* metagenomic assembler utilizing supervised learning. *DNA Res* 22, 69-77
 47. Peng Y, Leung HC, Yiu SM and Chin FY (2012) IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420-1428
 48. Zhu W, Lomsadze A and Borodovsky M (2010) *Ab initio* gene identification in metagenomic sequences. *Nucleic Acids Res* 38, e132
 49. Hyatt D, LoCascio PF, Hauser LJ and Uberbacher EC (2012) Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* 28, 2223-2230
 50. Kelley DR, Liu B, Delcher AL, Pop M and Salzberg SL (2012) Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. *Nucleic Acids Res* 40, e9
 51. Rho M, Tang H and Ye Y (2010) FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* 38, e191
 52. Trimble WL, Keegan KP, D'Souza M et al (2012) Short-read reading-frame predictors are not created equal: sequence error causes loss of signal. *BMC Bioinformatics* 13, 183
 53. Yok NG and Rosen GL (2011) Combining gene prediction methods to improve metagenomic gene annotation. *BMC Bioinformatics* 12, 20
 54. Finn RD, Coghill P, Eberhardt RY et al (2015) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44, D279-285
 55. Markowitz VM, Chen IM, Chu K et al (2014) IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res* 42, D568-573
 56. Tatusov RL, Galperin MY, Natale DA and Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28, 33-36.
 57. Huang K, Brady A, Mahurkar A et al (2014) MetaRef: a pan-genomic database for comparative and community microbial genomics. *Nucleic Acids Res* 42, D617-624.
 58. Kanehisa M and Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28, 27-30
 59. Huerta-Cepas J, Szklarczyk D and Forslund K (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* 44, D286-293
 60. Ye Y and Doak TG (2009) A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput Biol* 5, e1000465
 61. Abubucker S, Segata N, Goll J et al (2012) Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput Biol* 8, e1002358
 62. Keegan KP, Glass EM and Meyer F (2016) MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function. *Methods Mol Biol* 1399, 207-233
 63. Marcotte H and Lavoie MC (1998) Oral Microbial Ecology and the Role of Salivary Immunoglobulin A. *Microbiol Mol Biol Rev* 62, 71-109
 64. Chen H and Jiang W (2014) Application of high-throughput sequencing in understanding human oral microbiome related with health and disease. *Front Microbiol* 5, 508
 65. Preza D, Olsen I, Willumsen T, Grinde B and Paster BJ (2009) Diversity and site-specificity of the oral microflora in the elderly. *Eur J Clin Microbiol Infect Dis* 28, 1033-1040
 66. Lemon KP, Klepac-Ceraj V, Schiffer HK, Brodie EL, Lynch SV and Kolter R (2014) Comparative analyses of the bacterial microbiota of the human nostril and oropharynx. *MBio* 1, e00129-10
 67. McDonald D, Birmingham A and Knight R (2015) Context and the human microbiome. *Microbiome* 3, 52
 68. Nasidze I, Li J, Quinque D, Tang K and Stoneking M (2009) Global diversity in the human salivary microbiome. *Genome Res* 19, 636-643
 69. Nasidze I, Li J, Schroeder R, Creasey JL, Li M and Stoneking M (2011) High Diversity of the Saliva Microbiome in Batwa Pygmies. *PLoS One* 6, e23352
 70. Li J, Quinque D, Horz HP et al (2014) Comparative

- analysis of the human saliva microbiome from different climate zones: Alaska, Germany, and Africa. *BMC Microbiol* 14, 316
71. Mason MR, Nagaraja HN, Camerlengo T, Joshi V and Kumar PS (2013) Deep sequencing identifies ethnicity-specific bacterial signatures in the oral microbiome. *PLoS One* 8, e77287
 72. Rylev M and Kilian M (2008) Prevalence and distribution of principal periodontal pathogens worldwide. *J Clin Periodontol* 35(8 Suppl), 346-361
 73. Haffajee AD, Bogren A, Hasturk H, Feres M, Lopez NJ and Socransky SS (2004) Subgingival microbiota of chronic periodontitis subjects from different geographic locations. *J Clin Periodontol* 31, 996-1002
 74. Kim TS, Kang NW, Lee SB, Eickholz P, Pretzl B and Kim CK (2009) Differences in subgingival microflora of Korean and German periodontal patients. *Arch Oral Biol* 54, 223-229
 75. Miller MA and Cappuccio FP (2007) Ethnicity and inflammatory pathways-implications for vascular disease, vascular risk and therapeutic intervention. *Curr Med Chem* 14, 1409-1425
 76. Nguyen DP, Genc M, Vardhana S, Babula O, Onderdonk A and Witkin SS (2004) Ethnic differences of polymorphisms in cytokine and innate immune system genes in pregnant women. *Obstetrics and Gynecology* 104, 293-300
 77. Lavelle CL (1970) Crowding and spacing within the human dental arch of different racial groups. *Arch Oral Biol* 15, 1101-1103
 78. Lavelle CL (1971) Mandibular molar tooth configurations in different racial groups. *J Dent Res* 50, 1353
 79. Zaura E, Nicu EA, Krom BP and Keijsers BJ (2014) Acquiring and maintaining a normal oral microbiome: current perspective. *Front Cell Infect Microbiol* 4, 85
 80. Dominguez-Bello MG, Costello EK, Contreras M et al (2010) Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A* 107, 11971-11975
 81. Lif Holgerson P, Harnevik L, Hernell O, Tanner AC and Johansson I (2011) Mode of birth delivery affects oral microbiota in infants. *J Dent Res* 90, 1183-1188
 82. Holgerson PL, Vestman NR, Claesson R et al (2013) Oral microbial profile discriminates breast-fed from formula-fed infants. *J Pediatr Gastroenterol Nutr* 56, 127-136
 83. Vestman NR, Timby N, Holgerson PL et al (2013) Characterization and in vitro properties of oral lactobacilli in breastfed infants. *BMC Microbiol* 13, 193
 84. Stahringer SS, Clemente JC, Corley RP et al (2012) Nurture trumps nature in a longitudinal survey of salivary bacterial communities in twins from early adolescence to early adulthood. *Genome Res* 22, 2146-2152
 85. Lazarevic V, Whiteson K, Hernandez D, François P and Schrenzel J (2010) Study of inter-and intra-individual variations in the salivary microbiota. *BMC Genomics* 11, 523
 86. Flores GE, Caporaso JG, Henley JB et al (2014) Temporal variability is a personalized feature of the human microbiome. *Genome Biol* 15, 531
 87. Xu X, He J, Xue J et al (2015) Oral cavity contains distinct niches with dynamic microbial communities. *Environ Microbiol* 17, 699-710
 88. Cephas KD, Kim J, Mathai RA et al (2011) Comparative analysis of salivary bacterial microbiome diversity in edentulous infants and their mothers or primary care givers using pyrosequencing. *PLoS One* 6, e23503
 89. Crielaard W, Zaura E, Schuller AA, Huse SM, Montijn RC and Keijsers BJ (2011) Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health. *BMC Med Genomics* 4, 22
 90. McLean JS (2014) Advancements toward a systems level understanding of the human oral microbiome. *Front Cell Infect Microbiol* 4, 98
 91. Lasken RS and McLean JS (2014) Recent advances in genomic DNA sequencing of microbial species from single cells. *Nat Rev Genet* 15, 577-584
 92. Dewhirst FE, Chen T, Izard J et al (2010) The human oral microbiome. *J Bacteriol* 192, 5002-5017
 93. Clingenpeel S, Schwientek P, Hugenholtz P and Woyke T (2014) Effects of sample treatments on genome recovery via single-cell genomics. *ISME J* 8, 2546-9
 94. Solden L, Lloyd K and Wrighton K (2016) The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Curr Opin Microbiol* 31, 217-226
 95. Kim OS, Cho YJ, Lee K et al (2012) Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int J Syst Evol Microbiol* 62, 716-721
 96. Quast C, Pruesse E, Yilmaz P et al (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41, D590-596
 97. DeSantis TZ, Hugenholtz P, Larsen N et al (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72, 5069-5072