2010

Building the genomic base-layer of the oral “omic” world

The Forsyth Metagenomic Support Consortium

Jacques Izard

Building the genomic base-layer of the oral "omic" world

The Forsyth Metagenomic Support Consortium[1] and Jacques Izard[2]

1 The Forsyth Metagenomic Support Consortium included Oxana V. Baranova, Derek Spencer, Tsute Chen, Wenhan Yu, Alvin R. Plummer, William G. Wade, and Floyd E. Dewhirst at The Forsyth Institute, Boston, MA, USA; Emmanuel F. Mongodin, Derrick E. Fouts, and Karen Nelson at the J. Craig Venter Institute, Rockville, MD, USA.

2 Department of Molecular Genetics, The Forsyth Institute, Boston, MA 02135, USA

Abstract

With the shift of molecular technologies directed toward the understanding of greater biological complexity of the oral cavity, a knowledge gap was created by the lack of genomic data from the diverse oral microorganisms. To facilitate and enable the interpretation of metagenomic, transcriptomic, and proteomic data generated or soon to be generated from oral biofilms, we are providing reference genomic information from phylogenetically diverse oral bacterial isolates. This work, initiated by the National Institute of Dental and Craniofacial Research as an isolated effort, is now part of the Human Microbiome Project. The goal of this effort is the public release of genomic data in support of functional and phylogenetic analyses of the complex oral microbiome. The genomic information acquired will be a key component in understanding the interaction of the oral biofilms with the human host and in developing novel healthcare strategies to prevent and treat oral diseases.

Keywords: oral microbiome, oral biofilm, bacterial phylogeny, metagenome, bacterial diversity

Published in *Interface Oral Health Science 2009: Proceedings of the 3rd International Symposium for Interface Oral Health Science, Held in Sendai, Japan, Between January 15 and 16, 2009 and the 1st Tohoku-Forsyth Symposium, Held in Boston, MA, USA, Between March 10 and 11, 2009*. Editors: Takashi Sasano & Osamu Suzuki, editors.
DOI 10.1007/978-4-431-99644-6_112

Introduction

The human oral biofilms are readily accessible complex bacterial communities. Although it is one of the best-characterized microbiomes at the phylogenetic level, its internal dynamics and relationship to the host remain a mystery. This abundant self-renewable biofilm is responsible for oral health as well as diseases of both hard and soft tissues of the oral cavity. In addition, evidence is accumulating to indicate the microbiome's significant influence on overall health, through a direct interactive exchange between the host response and the members of the oral microbiome [1–3]. Bacterial complexes are involved in those processes, and no individual pathogen can be singled out [4–7].

Bacterial diversity is a key characteristic of dental sub and supragingival plaque as well as the biofilms on other oral surfaces including the gingiva, tongue, hard palate, and cheeks. To date, over 600 bacterial species and a single *archaea* species have been identified as members of the oral community [5, 8, 9]. This number is increasing as more studies are performed all over the world. The added diversity may be related to disease status, regional environmental factors, or diet. For example, one might expect that the oral flora of a subject with a diet rich in raw fish with high fatty acid content will differ from the flora of a vegan subject. Such population changes have been well demonstrated in subjects with periodontitis compared with healthy subjects [4, 6, 10]. This underscores the interplay between the bacterial members of the oral biofilms, the host's physiological reactions and the fact that the mouth is an open system to the environment. Foods and fluids that we ingest as well as air that we breathe are significant sources of new bacterial challenges on a daily basis.

The extent of the morphological diversity of the bacteria in the oral cavity is astonishing. The highly motile spiral of the treponemes might coexist with the gliding multicellular-filamentous species of the genus *Simonsiella*, the corncob arrangements of *Corynebacterium matruchotii* or *Fusobacterium nucleatum* with *Streptococcus sanguis*, as well as cocci, short and long rods [11–14]. The phylogenetic diversity encompasses at least 12 phyla and over 170 genera [8]. The genera include cultivable named species (47%), cultivable yet-to-be-named species (18%), and yet-to-be-cultivated phylotypes, which often wear the inaccurate label of uncultivable in the literature (35%).

These numbers are always fluctuating as the naming of organisms is an ongoing process [14, 15], and novel culture methods are being developed.

The variety of organisms, mentioned above, results in a diverse genetic potential that is mostly untapped and unknown. How those bacteria produce pathogenic factors, resist antibiotics, evade host immune response, communicate with each others, or simply use nutrients for their own energy production is mostly a mystery. The genetic characteristics of these oral bacteria are unknown because most of them are considered commensal and have not been the focus of many studies. Other factors detrimental to their study are their fastidious growth characteristics compared to *Escherichia coli* or *Bacillus subtilis* and their lack of established genetic systems.

While metagenomics, transcriptomics, and proteomics are part of the next key steps of understanding complex systems, they all rely on availability of the genomic sequence to decipher the content of the dataset via similarities to better known systems. These similarities provide the original clues toward function and phylogenetic attribution. With the absence of such data for most of the oral microbiome members, the scientific community was likely at risk to miss some of the benefits from the ongoing technological advances in the omic world. Those advances touch all aspects of fields looking at a large-scale analysis of a genome, a proteome, or a metabolome, just to name a few. Prior to joining the Human Microbiome project, we provided genome surveys to the community (**Table 1**). Genome surveying is a low cost approach to provide anchoring genetic data for metagenomic analysis of obscure branches of the phylogenetic tree. Genetic libraries were created, and 12 organisms from 6 phyla were sequenced (Table 1). A minimum of 350,000 bases were recovered and annotated per genome (Table 1). A novel annotation pipeline was created to accommodate the short contigs generated by low sequencing coverage of random genomic libraries (Dewhirst et al. unpublished data) [8, 15]. This took advantage of the lengths of the singlets (single sequencing read) and the contigs (assembly of two or more sequencing reads) that may not cover the full length of the open reading frame but are sufficient to ascertain similarities and to provide matching sequences to pyrosequencing generated short reads. The dynamic annotation, updated periodically, for each genome survey is available online at

Table 1. Genome survey data produced and released in Genbank

Phylum	Genus	Species	Strain	Taxon ID[a]	SEQF ID[b]	No. of contigs & singlets	Combined length (Kbp)	Genbank accession number
Bacteroidetes	*Prevotella*	*sp. oral taxon 302*	F0020	302	1,020	404	362	FI090687-FI091098
Firmicutes	*Bulleidia*	*extructa*	W1219	603	1,088	350	385	ET629122-ET629476
Firmicutes	*Eubacterium*	*infirmum*	ATCC 700433	105	1,108	413	409	ET630587-ET631005
Firmicutes	*Solobacterium*	*moorei*	W5408	678	1,152	338	354	ET631874-ET632213
Firmicutes	*Veillonella*	*parvula*	ATCC 17745	161	1,058	353	372	ET632214-ET632569
Fusobacteria	*Leptotrichia*	*buccalis*	ATCC 14201	563	1,028	345	392	ET631526-ET631873
Proteobacteria	*Campylobacter*	*gracilis*	ATCC 33236	623	1,005	384	402	ET629477-ET629866
Proteobacteria	*Campylobacter*	*rectus*	ATCC 33238	748	1,089	356	353	ET629867-ET630228
Proteobacteria	*Campylobacter*	*showae*	ATCC 51146	763	1,091	352	340	ET630229-ET630586
Spirochaetes	*Treponema*	*lecithinolyticum*	OMZ 684T	653	1,060	1,377	1,469	ET632570-ET633946
Synergistetes	*Jonquetella*	*anthropi*	E3_33 E1	777	1,476	509	618	ET631006-ET631525
Synergistetes	*Pyramidobacter*	*piscolens*	W5455	357	1,541	852	615	DU723013-DU723395
						Total	6,033	

a. Taxon ID refers to a phylotype designation as part of the investigation of phylogenetic diversity of the oral microbiome, http://www.homd.org

b. SEQF ID is a unique identifier of bacterial genomes sequenced from a unique isolate from a specific laboratory http://www.homd.org

the Human Oral Microbiome database (http://www.homd.org) [8]. The manual annotation of *Pyramidobacter piscolens* genome survey data was published when the organism was named [16].

With the announcement of the Human Microbiome Project, another era of bacterial genome sequencing began [17]. Becoming a member of this scientific endeavor allowed the transition from the production of genome surveys to the sequencing of full genomes by pyrosequencing. This is being done in partnership with four genomic centers: the Broad Institute of MIT and Harvard, the J. Craig Venter Institute, the Genome Sequencing Center at Washington University, and the Human Genome Sequencing Center at the Baylor College of Medicine. Now, with the first 50 bacterial genomes provided to the centers for sequencing, we look forward to closing the gap in describing genetic and phylogenetic diversity. The progress done can be monitored at the Human Microbiome Project website (http://www.hmpdacc.org) and at the Human Oral Microbiome database. In addition, the bacterial strains are provided to the Biodefense and Emerging Infections Research Resources Repository at ATCC (http://www.beiresources.org) for availability to the scientific community at large. For phylogenetic studies, the essentially full sequences of the 16S ribosomal RNA gene are also provided to GenBank repository.

How will this benefit the patient? With the acceptance that oral bacteria are part of the equation for good oral health, it becomes important to ensure that day-to-day treatments (toothpastes, oral rinses, etc.) are not damaging the host immune response or suppressing of the most critical group of bacteria involved in oral health. Moving toward prevention requires an unprecedented effort to understand what health is. It is also important to understand what the organisms' proportions are, which bacterial complexes are prevalent, which metabolic pathways are shared, which part of the genomic potential is expressed (transcriptomics), what the composition of a bacteria in the biofilm (proteomics) is, and consequently which abundant potential peptide targets are available for drug design and which metabolites are released (metabolomics) and could be targets for inhibitors. Genetic studies of both the normal and the pathogenic flora will be, as they have been in the past, a key element to develop drug therapies [18, 19]. Discovering the components contributing to health at the microbiome level can lead to clinical strategies to maintain a healthy balance between the microbiome and the host response.

The next steps are for the scientific community to expand the work of isolating new strains and providing genomic sequence information. This process will improve our ability to understand how the oral microbiome self-organizes and interacts with the host. This host interaction is a key component in the dynamic interplay, leading to the maintenance of oral health as well as the population shifts resulting in chronic diseases such as caries and periodontitis. Increased understanding of the behavior of the oral biofilms through transcriptomics, proteomics, interactomics, functomics, and many other "omic" disciplines relies on the available genomic data, and will open new avenues for treatment strategies.

Acknowledgment This work was supported by the grants DE017106 (JI) and DE16937 (Floyd E. Dewhirst) from the National Institute of Dental and Craniofacial Research, National Institutes of Health Bethesda, MD. We would like to thank the collaborators that are now actively involved in producing the pyrosequencing data at the Broad Institute of MIT and Harvard, the J. Craig Venter Institute, the Genome Sequencing Center at Washington University, and the Human Genome Sequencing Center at the Baylor College of Medicine.

References

1. Mealey BL, Oates TW (2006) Diabetes mellitus and periodontal diseases. J Periodontol 77(8):1289–1303
2. Cavrini F, Sambri V, Moter A et al (2005) Molecular detection of *Treponema denticola* and *Porphyromonas gingivalis* in carotid and aortic atheromatous plaques by FISH: report of two cases. J Med Microbiol 54(1):93–96
3. Davenport ES, Williams CE, Sterne JA et al (1998) The East London Study of Maternal Chronic Periodontal Disease and Preterm Low Birth Weight Infants: study design and prevalence data. Ann Periodontol 3(1):213–221
4. Socransky SS, Haffajee AD, Cugini MA et al (1998) Microbial complexes in subgingival plaque. J Clin Periodontol 25(2):134–144
5. Paster BJ, Boches SK, Galvin JL et al (2001) Bacterial diversity in human subgingival plaque. J Bacteriol 183(12):3770–3783
6. Marsh PD (1994) Microbial ecology of dental plaque and its significance in health and disease. Adv Dent Res 8(2):263–271
7. Sakamoto M, Huang Y, Ohnishi M et al (2004) Changes in oral microbial profiles after periodontal treatment as determined by molecular analysis of 16S rRNA genes. J Med Microbiol 53(Pt 6):563–571
8. Dewhirst FE, Izard J, Paster BJ et al (2008) The Human Oral Microbiome Database. http://www.homd.org
9. Lepp PW, Brinig MM, Ouverney CC et al (2004) *Methanogenic archaea* and human periodontal disease. Proc Natl Acad Sci USA 101(16):6176–6181
10. Aas JA, Paster BJ, Stokes LN et al (2005) Defining the normal bacterial flora of the oral cavity. J Clin Microbiol 43(11):5721–5732
11. Izard J, Hsieh CE, Limberger RJ et al (2008) Native cellular architecture of Treponema denticola revealed by cryo-electron tomography. J Struct Biol 163:10–17
12. Saglie R, Newman MG, Carranza FA Jr et al (1982) Bacterial invasion of gingiva in advanced periodontitis in humans. J Periodontol 53(4):217–222
13. Lancy P Jr, Dirienzo JM, Appelbaum B et al (1983) Corncob formation between *Fusobacterium nucleatum* and *Streptococcus sanguis*. Infect Immun 40(1):303–309
14. Lancy P Jr, Appelbaum B, Holt SC et al (1980) Quantitative in vitro assay for "corncob" formation. Infect Immun 29(2):663–670
15. Chen T, Abbey K, Deng W et al (2005) The bioinformatics resource for oral pathogens. Nucleic Acids Res 33:W734–W740
16. Downes J, Vartoukian SR, Dewhirst FE et al (2009) *Pyramidobacter piscolens gen. nov., sp. nov., a member of the phylum 'Synergistetes' isolated from the human oral cavity*. Int J Syst Evol Microbiol 59:972–980
17. Turnbaugh PJ, Ley RE, Hamady M et al (2007) The human microbiome project. Nature 449(7164):804–810

18. Stokes NR, Sievers J, Barker S et al (2005) Novel inhibitors of bacterial cytokinesis identified by a cell-based antibiotic screening assay. J Biol Chem 280(48):39709–39715
19. Hung DT, Shakhnovich EA, Pierson E et al (2005) Small-molecule inhibitor of *Vibrio cholerae* virulence and intestinal colonization. Science 310(5748):670–674