# Comparative Microbial Genomics

Margaret J. Duncan

*Department of Molecular Genetics,The Forsyth Institute, Boston, USA*

**Summary:** A brief overview of microbial genome sequencing is followed by an update on new directions in the field. Interesting facets of individual oral bacterial genomes are described, together with a discussion of how genome information is being exploited to understand gene expression and regulation. The field of comparative genomics is reviewed and illustrated with examples of the mechanisms that drive the evolution of bacteria and their virulence factors.

**Key words:** genomics, oral bacteria

*Oral Biosci Med 2005; 2/3: 191-194*

## INTRODUCTION

Since the publication of the first bacterial genome sequence ten years ago (Fleischmann et al, 1995) the field of microbial genomics has passed from infancy through adolescence to adulthood. This event is marked with high enthusiasm for the tremendous amount of new information generated, and sober maturity with the realization of how much we still do not know. While we may be approaching late logarithmic growth phase with regard to the number of single genome sequences published and in progress, 266 and 730 respectively as of June, 2005 (Gold Genomes Online database), this sequence data will be analyzed, compared, and reanalyzed for many years to come.

To date, most of the sequenced genomes are from medically or environmentally important bacteria, but now there is an effort, financed by the US Department of Energy through the Joint Genome Institute, to sequence genomes of under-represented or new prokaryotes in order to 'fill in the gaps' in the evolutionary 'Tree of Life Project'. New scientific frontiers are being forged with metagenomes that have the advantage of containing sequence information from all the organisms in a defined environment, both culturable and non-culturable, since the latter organisms can represent 60-99% of the bacterial component. While metagenomes from specific natural environments were the first to be sequenced, they were followed by projects to sequence the microbiomes of the human intestine and the oral cavity. Sequence information generated by these projects adds significantly to the large amount of genetic information that we do not, as yet, understand. However, from the comparison of large amounts of sequence data we can begin to make intelligent guesses and testable hypotheses about gene function, how genomes evolved, and identify common themes in bacterial pathogenicity.

## OVERVIEW OF GENOME SEQUENCING

The random shot-gun strategy is used to sequence individual prokaryote and eukaryote genomes, as well as metagenomes. Consisting of four phases, in the first a genomic library of small fragments (ca. 2-kb) is prepared from chromosomal DNA isolated from the organism of interest. The fragments are ligated to a high copy number bacterial plasmid, and transformed into *Escherichia coli*. At the same time, larger insert libraries are prepared in plasmid vectors, bacteriophage lambda or fosmids, and these will be used later in the genome assembly phase as backup DNA templates for closing sequence and physical gaps. In the second phase, the ends of all the small fragments are sequenced to obtain approximately eight-fold coverage of the genome with overlapping small random sequences. In the third phase, the random sequences are assembled, usually beginning with 30-base pairs overlaps, building up increasingly larger fragments that ultimately can be joined to yield one contiguous sequence. During the final phase, open reading frames

of at least 100 base pairs are identified, first by automated annotation, and finally by visual inspection.

### Genomes of Oral Bacteria

The genomes of 15 oral bacteria have been sequenced, and these include both putative oral pathogens and beneficial bacteria (Table 1). Annotated sequences for several of these organisms can be found at the websites of The Institute for Genomic Research (Comprehensive Microbial Resource) and Oralgen hosted at the Los Alamos National Laboratory. Once a genome sequence is published it is fair game for re-annotation, and, as anticipated, several other groups have added oral bacterial genomes to their repertoires for comparative genomic purposes (Table 2).

A genome sequence contains both protein coding regions as well as associated promoter elements, and thus information necessary to study the regulation of gene expression. For virulence genes both promoter and structural gene sequences are potential targets in the development of new drugs and vaccines. Research to identify virulence–associated functions of oral bacteria and determine how they are regulated has flourished in the post-genomic era using different approaches.

### Comparative Genomics

A field that integrates both *in vitro* and *in silico* technologies, comparative genomics is applicable to multiple strains of a single bacterial species, to multiple species within a genus or phylum, or to the whole bacterial kingdom. Sequencing of multiple bacterial strains is still the gold standard for comparative purposes, and can be cost-effective for species that are genetically highly divergent. However, many species do not diverge significantly, in which case differences between strains can be identified using DNA microarrays. Such 'genomotyping' has been applied to many pathogens, including *Porphyromonas gingivalis* (Chen et al, 2004). The concept of core and flexible genomes was developed from these comparisons, whereby core genes are present in all strains of a species and encode essential 'housekeeping' activities, while the flexible genome contains genes that are only present in some strains, and most likely were acquired by lateral gene transfer.

Another important and useful consequence of having a substantial number of genome sequences is the ability to compare protein sequences across a wide spectrum of bacteria with the goal of assigning function to the large class of predicted open reading frames that are still annotated as hypothetical. By combining data derived from multiple bioinfomatics searches such as identification of active site motifs, genome context, and protein–protein interactions, it may be possible to make testable predictions of function for subsets of hypothetical proteins.

### Lateral Gene Transfer and Pathogenicity Islands

From the comparison of multiple bacterial genomes it is clear that the transfer of genes between species, i.e. lateral gene transfer, has occurred more frequently than previously appreciated. Signs that a gene may have been acquired by this mechanism include different guanosine and cytosine content and codon usage from host genes, encoded antibiotic resistance functions, activities associated with virulence and genetic linkage with known moveable DNA elements. Clusters of genes with these properties are known as pathogenicity islands, because they often contain genes for virulence factors in bacteria that cause disease (Hacker and Kaper, 2000). These islands often carry genes encoding integrases and transposases that are involved in DNA mobility, and they may be associated with transfer RNA genes, favored sites for the integration of foreign DNA. Through comparative genomic studies it is possible to track the spread of transposons, antibiotic resistance genes, pathogenicity islands, and bacteriophage genes between species, and the evolution of associated virulence functions.

The conditions prevailing in dental plaque, i.e. constant cell-to-cell contact, are favorable for DNA transfer by 'mating', and genes associated with conjugative transposons from enteric Bacteroides were found in the genomes of *Porphyromonas gingivalis* and *Prevotella intermedia* (Duncan, 2003). In addition, the *ermF* and *tetQ* resistance genes were detected in both Gram-positive and Gram-negative bacteria found in the colon, as well as oral Porphyromonas species (Salyers et al, 2004). These findings raise the questions of whether and how antibiotic resistance and other genes can be transferred between plaque bacteria, and to other commensals that colonize humans.

Another mechanism of gene transfer that may be operational in dental plaque relates to increased genetic competence of biofilm-grown bacteria, and the subsequent uptake of DNA by transformation. In a recent study two interesting features of biofilm life were established: first, that transformation frequencies in biofilm-grown cells of *Streptococcus mutans* were 10- to 600-fold higher than planktonic cells; and second, working on the premise that DNA in the environment originates from dead bacteria, it was demonstrated that biofilm-grown cells were not only able to take up purified chromosomal DNA, but also DNA released from heat-killed biofilms (Li et al, 2001). So far there is no direct evidence that gene transfer can actually occur in the multi-species biofilm of dental plaque, but these recent reports lend credence to the possibility.

**Table 1 Genome projects for oral bacteria**

| Genome | Strain | Genome size (Mb) | Collaborating institutions | Funding | Websites for searches/annotation |
|---|---|---|---|---|---|
| *Actinobacillus actinomy-cetemcomitans* | HK 1651 | 2.105 | Univ. of Oklahoma | NIDCR[a] | www.genome.ou.edu<br>www.oralgen.lanl.gov |
| *Actinomyces naeslundii* | MG1 | 3.0 | TIGR | NIDCR | www.tigr.org |
| *Fusobacterium nucleatum* | ATCC 25586 | 2.17 | Integrated Genomics Inc. | Integrated Genomics Inc. | www.genome.org/cgi/doi |
| *Fusobacterium nucleatum* | ATCC 10953 | 2.4 | BCM-HGSC[b]UCLA[c] | NIDCR | www.hgsc.tmc.edu/ |
| *Fusobacterium nucleatum subspecies vincentii* | ATCC 49256 | 2.118 | Integrated Genomics Inc. | Integrated Genomics Inc. | www.genome.org/cgi/doi |
| *Porphyromonas gingivalis* | W83 | 2.343 | TIGR<br>The Forsyth Institute | NIDCR | www.tigr.org<br>www.oralgen.lanl.gov |
| *Prevotella intermedia* | 17 | 2.8 | TIGR | NIDCR<br>LANL | www.tigr.org |
| *Streptococcus gordonii* | NTCT 7868 | 2.2 | TIGR | NIDCR | www.tigr.org |
| *Streptococcus mitis* | | | | Univ. of Wurzburg | |
| *Streptococcus mitis* | NTCT 12261 | 2.2 | TIGR | NIDCR | www.tigr.org |
| *Streptococcus mutans* | UA159 | 2.03 | Univ. of Oklahoma<br>Ohio State Univ. | NIDCR | www.genome.ou.edu<br>www.oralgen.lanl.gov |
| *Streptococcus sanguinis* | SK36 | | Commonwealth Biotechnologies Inc., Virginia Commonwealth Univ. | NIDCR | www.sangius.mic.vcu.edu |
| *Streptococcus sobrinus* | 6715 | 2.2 | TIGR | NIDCR | www.tigr.org |
| *Tanerella forsythensis* | ATCC 43037 | | TIGR | NIDCR | www.tigr.org |
| *Treponema denticola* | ATCC 35405 | 2.843 | TIGRBCM-HGSC | NIDCR | www.tigr.org<br>www.oralgen.lanl.gov |

[a]National Institute of Dental and Craniofacial Research; [b]Baylor College of Medicine Human Genome Sequencing Center; [c]University of California Los Angeles

**Table 2 Databases for genomic tools and annotations**

| Database | Description | Website |
|---|---|---|
| BROP | Bioinformatics Resource for Oral Pathogens | www.brop.org |
| MIPS/PEDANT | Munich Information Center for Protein Sequences; Protein Extraction, Description, and Analysis Tool | www.mips.gsf.de |
| STRING | Search Tool for the Retrieval of Interacting Genes/Proteins | www.string.embl.de |
| BIND | Biomolecular Interaction Network Database | www.bind.ca |
| DIP | Database of Interacting Proteins | www.dip.doe-mbi.ucla.edu |
| Gold Genomes Online | Re-annotation | www.genomesonline.org |
| Integrated Microbial Genomics | Re-annotation of Oral Bacterial Genomes | www.Img.jgi.doe.gov |
| The Seed | Gene Analysis and Annotation with Emphasis on Metabolic Systems | www.theseed.uchicago.edu/FIG/index.cgi |

## CONCLUSIONS

Oral microbiology has been revitalized with the advent of genome sequences for its constituent organisms. Eventually this new information will provide insights on pathogenesis, gene regulation, and metabolism. Genome comparisons have yielded evidence of past genetic exchanges, and stimulated the development of new methods for enhancing gene transfer. New projects to determine the oral microbiome will push our horizons still further towards identifying unculturable components of the oral cavity.

## ACKNOWLEDGEMENT

Supported by NIH-NIDCR R01-DE-10510 and R01-DE-015931

## REFERENCES

Chen T, Hosogi Y, Nishikawa K, Abbey K, Fleischmann RD, Walling J, Duncan MJ. Comparative whole genome analysis of virulent and avirulent strains of *Porphyromonas gingivalis*. J Bacteriol 2004;186:5473-5479.

Duncan MJ. Genomics of oral bacteria. Crit Rev Oral Biol Med 2003;14:175-187.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 1995;269:496-512.

Hacker J, Kaper JB. Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol 2000;54:641-679.

Li YH, Lau PC, Lee JH, Ellen RP, Cvitkovitch DG. Natural genetic transformation of *Streptococcus mutans* growing in biofilms. J Bacteriol 2001;83:897-908.

Salyers AA, Gupta A, Wang Y. Human intestinal bacteria as reservoirs for antibiotic resistance genes. Trends Microbiol 2004;12:412-416.

***Reprint requests:***
Margaret J. Duncan, Ph.D.
Department of Molecular Genetics
The Forsyth Institute
140 Fenway
Boston, MA 02115, USA
E-mail: mduncan@forsyth.org