Forsyth
Pioneering Discoveries
Profound Change

L. YANG[1], T. CHEN[1,2], J. IZARD[1,2], A.C. TANNER[1,2], W. WADE[2], B.J. PASTER[1,2], and F.E. DEWHIRST[1,2]

[1] The Forsyth Institute, Cambridge, Massachusetts
[2] Department of Oral Medicine, Infection & Immunity, Harvard School of Dental Medicine, Boston, Massachusetts
[3] Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, England

Scan to download
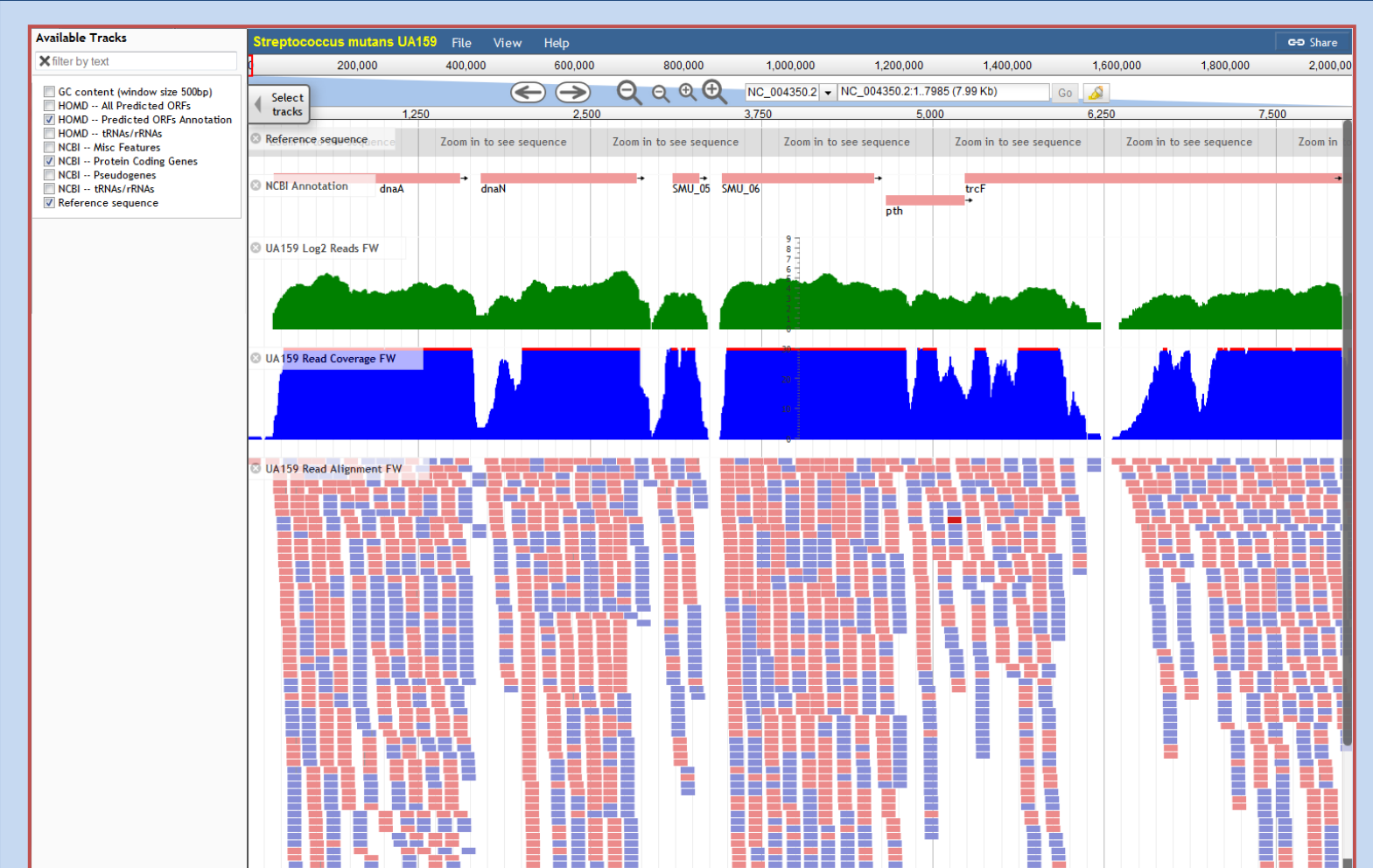a copy of this poster

Abstract

Objective: The Human Oral Microbiome Database (HOMD) is a public database providing comprehensive information on the prokaryotic species present in the human oral cavity. Our team maintains and curates the HOMD with up-to-date information on genomics, phylogeny, and taxonomy for prevalent human oral microbial species. We have also been developing bioinformatics tools that are integrated into the database for use by the research community. This report summarizes the latest content updates as well as newly available tools and features.

Method: The HOMD was created with PHP, Perl, MySQL and Apache and is hosted on a cluster of computers and housed in the Forsyth Institute Data Center. The HOMD Genomic BLAST was constructed based on the NCBI BLAST+. The new HOMD Genome Browser was implemented based on the "JBrowse" system. Novel human oral taxa were created based on 16S rRNA reference sequences of recently described oral species or recognized phylotypes. New genomic sequences were obtained through the collaboration with the Human Microbiome Project (HMP) sequencing centers and from repositories world-wide.

Result: Currently HOMD contains a total of 688 human oral taxa. Of these, 343 (49.8%) are validly described, 101 (14.7%) unnamed (but cultivated) and 244 (35.5%) known only as uncultivated phylotypes. The current version (13.2) of the 16S rRNA reference sequences contains 831 full length sequences representing prevalent human oral taxa. The HOMD Genomics database currently contains 1158 genomes, representing 339 oral taxa, with either static or dynamic annotations. Additional genomic sequences are being added frequently as they appear at NCBI and other sequence repositories. The HOMD Genomic BLAST and the HOMD JBrowse Genome Browser have also been implemented for searching and navigating among all the human oral microbes with available sequences.

Conclusion: HOMD is a comprehensive bioinformatics resource for the scientific community. HOMD is available at http://www.homd.org.

Introduction

The goal of creating the Human Oral Microbiome Database (HOMD) is to provide the scientific community with comprehensive information on the approximately 700 prokaryote species that are present in the human oral cavity. Approximately 49.8% are officially named, 14.7% unnamed (but cultivated) and 35.5% are known only as uncultivated phylotypes. The HOMD presents a provisional naming scheme for the currently unnamed species so that strain, clone, and probe data from any laboratory can be directly linked to a stably named reference scheme. The HOMD links sequence data with phenotypic, phylogenetic, clinical, and bibliographic information. Genome sequences for oral bacteria determined as part of this project, the Human Microbiome Project, and other sequencing projects are being added to the HOMD as they become available. Genomes for 339 oral taxa (49.3% of all HOMD taxa) are currently available on HOMD. The HOMD site offers easy to use tools for viewing all publically available oral bacterial genomes.

The Human Oral Microbiome

Taxonomy
- Total No. of taxa: **688**
- Named Species: **343**
- Unnamed Cultivated: **101**
- Uncultured Phylotypes: **244**

16S rRNA Reference Sequences
- Current Version: **13.2**
- Total No. of Sequences: **831**
- Searchable on the HOMD Web Site
- Downloadable from HOMD Web Site

Genomic Sequences
- Total Genomes: **1158**
- NCBI Annotated: **1096**
- HOMD Annotated: **397**
- Total Taxa with Genomes: **339**

Total No. of	in HOMD	In RefSeq
Domain	2	1
Phylum	14	13
Class	24	23
Order	40	39
Family	83	82
Genus	183	182
Species	688	§686

§ The 16S rRNA for *Eikenella sp.* oral taxon 009 is a chimera and has been dropped from RefSeq; the sequence for *Methanobrevibacter oralis* is only ~900nt and is not included in the current RefSeq.
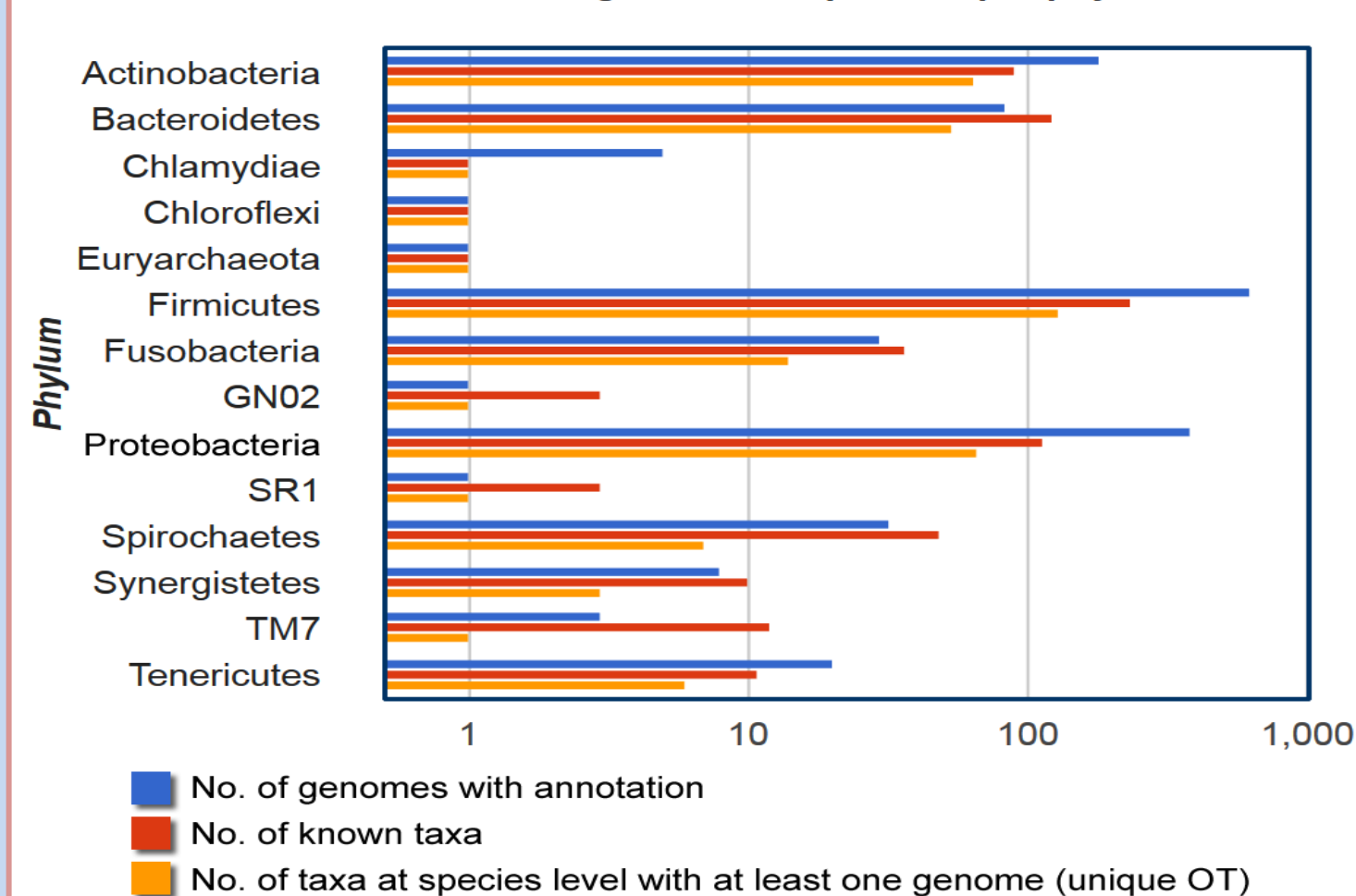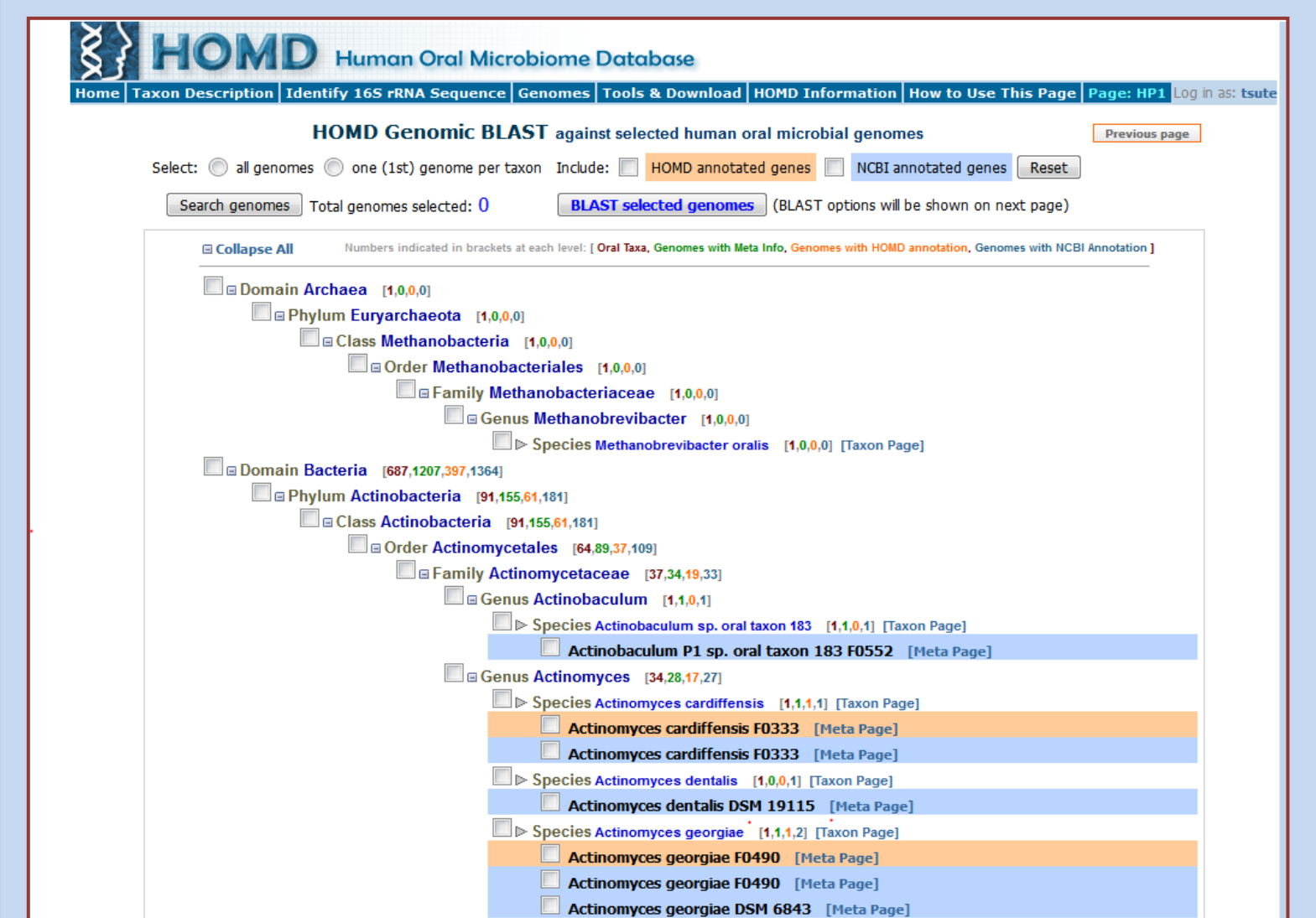
No. of Taxon	No. of RefSeq
*1	6
¥1	5
8	4
17	3
78	2
581	1

* *Treponema socranskii* (HOT 769)
¥ *Veillonella parvula* (HOT 161)

Distribution of genomes sequenced per phylum

- No. of genomes with annotation
- No. of known taxa
- No. of taxa at species level with at least one genome (unique OT)

Next Generation Genome Viewer

- Based on JBrowse - a fast, embeddable genome browser built completely with JavaScript and HTML5, with optional run-once data formatting tools written in Perl.
- Google map of the genome viewer – convenient viewing options including zooming, mouse navigation, clickable and searchable annotation.
- Multiple annotation sources in different panels: HOMD dynamic annotation, NCBI, LANL Oralgen, etc.
- Whole genome or nucleotide level view with 6 frame translations.
- View custom mapping and annotation data such as gene annotation, NGS sequence mapping, RNAseq mapping using local computer data without upload.
- Upload and view user's reference genome sequences (HOMD unique feature).

HOMD Dynamic Genomic BLAST

- Implemented based on the NCBI BLAST+ .
- Provides a visual taxonomy-based navigation interface for easy and dynamic selection of a set of genomes for sequence homology search.
- The selection can be a combination of individual genomes and/or a group of genomes related at any taxonomic level (species, genus, etc).
- The BLAST parameters are dynamically presented after the genome selection and the results are available on the web and for download in multiple formats.

HOMD Google Group & Mailing List

To join HOMD Google Discussion Group: Send an empty email to: homd-mail+subscribe@googlegroups.com. Or subscribe now by scanning this email address (encoded as a QR code) with your cell phone and then send the email.

HOMD Mailing List: Visit the mailing list web site: https://groups.google.com/forum/#!forum/homd-mail. Or QR code, scan this QR code to visit the mailing list in your cell phone browser.

For more information visit HOMD home page (http://www.homd.org) and look for Google Group and Mailing List option.

Funding