An uncultured human-associated bacterium model

Authors

Jorge Dinis[1], David Barton[1], Jamsheed Ghadiri[1], Deepa Surendar[1], Kavitha Reddy[1], Fernando Velasquez[1], Carol Chaffee[2], Wendy Lee[3], Helen Gavrilova[1], Hazel Ozuna[4], Sam Smits[1], and Cleber Ouverney[1*]

[1]San José State University, Department of Biological Sciences, One Washington Square, San Jose, CA 95192-0100

[2]University of Florida, Department of Zoology, PO Box 118525, Gainesville, FL 32611-8525

[3]University of California Santa Cruz, Department of Microbiology and Environmental Toxicology, Santa Cruz, CA 95064

[4]Universidad Metropolitana, Science and Technology Department, San Juan, Puerto Rico, 00928-1150

*Corresponding Author (cleber.ouverney@sjsu.edu)

An uncultured human-associated bacterium model

The Bacteria Domain experienced an explosion of novel lineages identified within the last decade, especially of lineages made entirely of uncultured members[1]. Since numerous cultivable bacteria have been shown to be instrumental in human development[2], health[3] and diseases[4,5], it is reasonable to speculate that strains from uncultured groups, which comprise nearly 80% of the human gut[6] and 68% of human oral[7] microbial consortia, participate in similar functions. The study of human-associated uncultured bacteria, however, has many practical limitations, such as access to patient samples, unpredictable microbial composition, and low relative abundance, all of which challenge experimental promptness and reproducibility. We propose that uncultured bacteria from environmental sources can serve as a model to better understand the roles their counterparts play in humans. In this study, we illustrate this concept using an environmental TM7 bacterium with ≥98.5% 16S rDNA gene homology to a group of TM7 bacteria found in both the human oral cavity and skin. Our TM7 model was readily detectable with molecular techniques as viable cells in sludge and in quantities greater than its human-associated relatives. Our approach circumvents difficulties imposed by sampling humans until either a TM7 strain of interest is cultured or an alternative method is proposed.

The Bacterial Candidate Division TM7 was first reported in 2001 in diverse environmental sources such as soil, freshwater, seawater, hot springs, mouse feces, and termite guts[8]. Recently, TM7 has been detected in various human body sites[7,9-11] and associated with the diseases periodontitis[12], vaginosis[13], and inflammatory bowel disease[14]. However, nothing is known about the direct role TM7 bacteria play in human health.

Although the 16S rDNA gene does not perfectly predict genomic wide phylogenetic homology, nor does it account for possible genomic differences caused by horizontal gene transfer, it is the most reliable and widely applied gene to predict basic genomic similarities[15,16]. Valuable insight into core cell functions such as metabolic processes[17] may therefore be derived. In this study, we assume 16S rDNA gene similarities between phylotypes correspond to similarities in genomic content.

We applied PCR targeting the 16S rDNA gene to screen various accessible environmental sites until establishing sludge as a promising source of a TM7 human homologue. From seven 16S rDNA gene clone libraries generated from sludge samples collected between January 2007 and December 2009 we identified 153 clones in the Candidate Division TM7, of which 103 (67.3%) shared species-level homology (≥ 98.5% similarity) with a human-associated TM7. Six of those sequences were deposited in GenBank (HM208132-37) based on their novelty and relevance to this study.

Activated wastewater sludge clone AAWS56C (accession number HM208134) showed 99.7% similarity to the human skin clone HM269723[9] and 98.6% similarity to the human oral TM7 clone TM7a (AY144355)[12] based on BLAST analysis. To test for potential contamination of our sludge samples, ten lab members handling samples for this project were PCR screened for TM7 from oral and skin sites. TM7 was detected in 4/10 oral and in 0/50 skin samples. The four TM7 positive oral samples generated 150 clones belonging to TM7 (based on BLAST) of which 13 (8.7%) clones had species-level (≥98.5%) 16S rDNA similarity to our sludge clone AAWS56C. Four representative oral TM7 sequences were deposited in GenBank (accession numbers HM215439, HM215442, HM215443, and HM215447). A nucleotide-base comparison among sludge and human homologous sequences (Supplementary Fig. 4) revealed conserved

base mutations and various single-nucleotide polymorphic sites that differentiated the sludge sequences from the human-counterparts, dismissing potential sample contamination.

We established the evolutionary relationships among 16 of our TM7 sequences (5 from sludge and 11 from oral samples) against 239 publicly available TM7 reference sequences from a wide-range of environmental, animal, and human sources (Fig. 1). With support of bootstrap resampling, two monophyletic TM7 subdivisions emerged (Fig. 1, regions labeled 1 and 2). Subdivision 1 was characterized by phylotypes predominantly isolated from environmental sources (soil, rhizosphere, marine, and freshwater), whereas subdivision 2 included phylotypes from environmental, activated wastewater, animal, and human sources. All human-associated TM7 phylotypes clustered within subdivision 2.

Our sludge clone AAWS56C formed a monophyletic group with other human-associated phylotypes from oral plaque and skin samples (Fig. 1, enclosure *TM7a Group*). To our surprise, three other environmental TM7 phylotypes from independent studies (EF515301 "microbial fuel cell," GQ264495 "waste site," and FJ671754 "beef cattle feedlot") shared class-level (92.2-93.7%) homology with human-associated phylotypes and a fourth (AJ318200 "waste-gas biofilter") shared genus-level (95.2%) homology (Fig.1, enclosures I, II, IV, and III), suggesting that additional environmental and human-associated relationships within the TM7 Phylum can be explored.

For an uncultured bacterial model - as the one we propose - to be practical, the bacterium must be prevalent and easily detectable. Analysis from seven independent clone libraries temporally distributed suggests that our sludge TM7 model (henceforth called either "TM7a model" or "environmental TM7a") is commonly found in activated sludge. Each library used

freshly collected samples and all contained the TM7a model 16S rDNA gene as confirmed through DNA sequencing.

Quantitative PCR (qPCR) assay, also targeting 16S rDNA, provided independent evidence supporting the prevalence of the TM7a model (Table 1). Through qPCR we determined the concentration of the three phylogenetic group levels Bacteria, the TM7, and our TM7a model in three sludge sites (S1, S2, and S3) from eight collection dates between August 2007 and January 2010 (Fig. 2A). While Bacteria and TM7 were detected in all 23 (100%) sludge samples, TM7a model was detected in 4/23 (17.4%) samples. Undetermined quantities of the TM7a model in 19/23 (83%) samples could be explained by known limitations of qPCR assays[18]. This was supported by sequences of the TM7a model recovered by cloning DNA amplicons from undetermined sludge samples (data not shown). qPCR TM7a analyses shown included only values with 3 qPCR cycles above the negative controls. TM7a model was also detected in ten additional sludge samples, but because values were only 1-2 qPCR cycles above controls, those counts were omitted from our conservative analysis.

Understanding the model organism population abundances and distributions may help one maximize retrieval from samples. The relative abundance of TM7 and TM7a from qPCR counts over the entire course of the study, for instance, provided insights for enriched samples with these phylotypes (Fig. 2A caption). The average relative abundance of TM7 (average TM7/average Bacteria) was 3.02% ± 3.80% (n= 23) with the highest value detected in S1 in May 2008 at 6.87% ± 3.72% (n=9). For TM7a model, the average relative abundance to Bacteria (average TM7a model/average Bacteria) was 0.01% ± 0.01% (n=4) and to TM7 (average TM7a model/average TM7) was 0.44% ± 0.18% (n=4). The highest relative abundance of TM7a model

to Bacteria was detected in S2 in August 2007 at 0.017% ± 0.025% (n=4) and to TM7 in S2 in April 2008 at 1.07% ± 0.15% (n=3).

Statistical analysis of qPCR values indicated that S1 had the overall highest concentrations of TM7. Two-way ANOVA comparisons of TM7 concentration (cell ml^{-1}, qPCR) for each time point between each of the three sampling sites at 95% confidence limit showed that S1 was statistically significantly higher than S2 (p=0.0001, t=4.640, n=18) in August 2007, than S3 (p=0.01, t=3.643, n=18) in May 2008, and than S2 (p=0.0001, t= 8.178, n= 14) and S3 (p = 0.01, t=3.860, n=16) in October 2009 (Fig.2.A2).

Because many qPCR measurements for TM7a model were below detectable levels, we used One-Way ANOVA with Bonferroni multiple comparisons to determine if the average qPCR values of the TM7a phylotype was statistically higher 1) at any particular time of the year for each of the three sites, and 2) in any one of the three sampling sites as compared to the other two sites. Environmental TM7a populations in S2 in August 2007 were statistically significantly higher than all other S2 counts (March 2008, n=6, p= 0.05, t=7.538 and April 2008, n=3, p=0.05, t=7.400) and S3 counts (October 2009, n = 6, t=5.389, p = 0.05) at 95% confidence.

Whole-cell quantification through Fluorescent *In Situ* Hybridization (FISH) detected TM7 and environmental TM7a in all analyzed sludge samples. FISH cell counts were performed in nine samples collected in June, August, and September 2010 at the same three sites as qPCR samples (Fig. 2B). Statistical analysis of FISH counts showed that in September 2010, TM7 populations were significantly higher in S1 (n=10, t = 3.259 p= 0.05, Two-Way ANOVA) and S2 (n=10, t = 4.457 p= 0.001, Two-Way ANOVA) than S3 (Fig. 2B2). No other significant differences in mean populations were determined. In contrast, the relative abundances of TM7 and environmental TM7a of the Bacteria were 1.05% ± 0.23% (n=9) and 0.034 % ± 0.01%

(n=9), respectively, while TM7a relative abundance of TM7 was 7.76% ± 3.07% (n=9). Even though the overall relative abundance of TM7 cells in our sludge samples represented a small fraction of the total Bacteria, it was 5 times greater than TM7 in healthy human oral sites (0.21% ± 0.05%) and ~2 times as the diseased sites with mild periodontitis (0.54% ± 0.10%) based on FISH counts[12], making sludge more amenable for sensitive experimentation with TM7. None of the human-associated TM7a group phylotypes have yet been quantified for comparison to our environmental model.

Microscopic FISH visualization was also used to characterize morphological features of TM7 and TM7a model cells in sludge samples (Fig. 3). TM7 cells ranged from short rods (2.5 x 0.5 μm) and cocco-bacilli (2.0 x 0.7 μm) to long (up to 40.0 μm x 1.2 μm) filaments, similar to previous studies[8,19] (Fig. 3A). TM7a model cells, however, consisted mostly of diplococci and short rods (Fig. 3B), found either isolated or within cell aggregates. FISH results combined with Substrate-Tracking Autoradiography Fluorescent *In Situ* Hybridization (STARFISH)[20] assays strongly suggested TM7 cells were metabolically active prior to fixation (Fig. 3C-D). TM7a model cells have not yet been detected in our STARFISH experiments.

Our survey identified an environmental source for a TM7 bacterium highly homologous to a human-associated phylotype. In fact, our TM7a model has higher sequence similarity to the human skin-associated TM7 than any TM7 found elsewhere in the human body hitherto. Phylogeny unveiled additional prospective environmental sources of TM7 models to different human oral and vaginal phylotypes. Future characterization of the TM7a model is needed including single-cell genomic amplification[21], nutrient uptake requirements[22], and probe protein-coding RNA[23] to confirm functional homology to human-associated TM7. Extrapolation from

this TM7 study can potentially render many other human-associated models for uncultured Bacteria and Archaea.

METHODS SUMMARY

Samples were collected from activated wastewater tanks (August 2007 - September 2010). FISH samples were fixed using 50% methanol and added to 24-teflon well slides with ADD-Cell coating. A modified bead-beater DNA extraction[12] was performed on un-fixed pelleted samples. PCR amplification used 100X diluted sludge sample genomic DNA primers and cycling conditions targeting either TM7[8,12] or ENV-TM7a-1112R (TCAACTATTCACAAGGG, this work, annealing temperature of 57°C). Sequences were cleaned with CodonCode Aligner software, confirmed identity by BLAST, checked for Chimera (Bellerophon v 3.0), aligned before being imported into ARB software package[24], and compared to 541 TM7 reference sequences from the SILVA rRNA project[25] and GenBank[26]. Sequences were selected based on coverage, length ≥1,200 bp, and relevance to this study. Sequence alignments were manually refined and a 1,052 unambiguous column filter was generated in ARB. A neighbor-joining tree with Felsenstein correction was generated using PHYLIP[27] and 1,000 replications bootstrap. Tree sequence metadata were manually curated and annotated[28]. Quantitative PCR (qPCR) was performed with an ABI 7300 with published primers, probes, and cycling conditions to target total Bacteria or TM7 communities[12]. We designed and validated qPCR primers ENV-TM7a-1112R and TM7a-997F (TCCCGAGAAGATTTACG), and probe TM7a-1033 (ATCTGTCACCGAGTTCCA) to target TM7a group. FISH used published probes and protocols for TM7[8,29] and our probe TM7a-1033 was catFISH[29] validated at 30% formamide. FISH imaging and counts used a Zeiss Axioscope-A.1 microscope, Hammamatsu camera, and AxioVision 4.7 software. Zeiss filters #49, #43, and #50 detected DAPI, Cy3, and Cy5, respectively. Substrate-Tracking Autoradiography FISH used previously published protocols[19].

1 Schloss, P. D. & Handelsman, J. Status of the microbial census. *Microbiology and molecular biology reviews : MMBR* **68**, 686-691 (2004).
2 Wilks, M. Bacteria and early human development. *Early Hum Dev* **83**, 165-170 (2007).
3 Round, J. L. & Mazmanian, S. K. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* **9**, 313-323 (2009).
4 Awano, S. *et al.* Oral health and mortality risk from pneumonia in the elderly. *Journal of dental research* **87**, 334-339 (2008).
5 Genco, R. J., Grossi, S. G., Ho, A., Nishimura, F. & Murayama, Y. A proposed model linking inflammation to obesity, diabetes, and periodontal infections. *Journal of periodontology* **76**, 2075-2084 (2005).
6 Kovatcheva-Datchary, P., Zoetendal, E. G., Venema, K., de Vos, W. M. & Smidt, H. Tools for the tract: understanding the functionality of the gastrointestinal tract. *Therap Adv Gastroenterol* **2**, 9-22 (2009).

7 Dewhirst, F. E. *et al.* The human oral microbiome. *J Bacteriol* **192**, 5002-5017 (2010).
8 Hugenholtz, P., Tyson, G. W., Webb, R. I., Wagner, A. M. & Blackall, L. L. Investigation of candidate division TM7, a recently recognized major lineage of the domain *Bacteria* with no known pure-culture representatives. *Applied and Environmental Microbiology* **67**, 411-419 (2001).
9 Gao, Z., Tseng, C. H., Pei, Z. & Blaser, M. J. Molecular analysis of human forearm superficial skin bacterial biota. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2927-2932 (2007).
10 Bik, E. M. *et al.* Bacterial diversity in the oral cavity of 10 healthy individuals. *The ISME journal* **4**, 962-974 (2010).
11 Eckburg, P. B. *et al.* Diversity of the human intestinal microbial flora. *Science* **308**, 1635-1638 (2005).
12 Brinig, M. M., Lepp, P. W., Ouverney, C. C., Armitage, G. C. & Relman, D. A. Prevalence of bacteria of division TM7 in human subgingival plaque and their association with disease. *Appl Environ Microbiol* **69**, 1687-1694 (2003).
13 Fredricks, D. N., Fiedler, T. L. & Marrazzo, J. M. Molecular identification of bacteria associated with bacterial vaginosis. *N Engl J Med* **353**, 1899-1911 (2005).
14 Kuehbacher, T. *et al.* Intestinal TM7 bacterial phylogenies in active inflammatory bowel disease. *J Med Microbiol* **57**, 1569-1576 (2008).
15 Wu, D. *et al.* A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* **462**, 1056-1060 (2009).
16 Konstantinidis, K. T. & Tiedje, J. M. Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol* **10**, 504-509 (2007).
17 Weng, L., Rubin, E. M. & Bristow, J. Application of sequence-based methods in human microbial ecology. *Genome Res* **16**, 316-322 (2006).
18 Bissonnette, L., Arnason, J. T. & Smith, M. L. Real-time fluorescence-based detection of furanocoumarin photoadducts of DNA. *Phytochem Anal* **19**, 342-347 (2008).
19 Ouverney, C. C. & Fuhrman, J. A. in *Molecular Microbial Ecology Manual, Second Edition* (eds G. A. Kowalchuk *et al.*) 1689-1710 (Kluwer Academic Publishers, 2004).
20 Ouverney, C. C. & Fuhrman, J. A. Combined microautoradiography-16S rRNA probe technique for determination of radioisotope uptake by specific microbial cell types in situ. *Applied and Environmental Microbiology* **65**, 1746-1752 (1999).
21 Marcy, Y. *et al.* Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *Proc Natl Acad Sci U S A* **104**, 11889-11894 (2007).
22 Ouverney, C. C. & Fuhrman, J. A. Marine planktonic archaea take up amino acids. *Appl Environ Microbiol* **66**, 4829-4833 (2000).
23 Amann, R. & Fuchs, B. M. Single-cell identification in microbial communities by improved fluorescence in situ hybridization techniques. *Nat Rev Microbiol* **6**, 339-348 (2008).
24 Ludwig, W. *et al.* ARB: a software environment for sequence data. *Nucleic Acids Res* **32**, 1363-1371 (2004).
25 Pruesse, E. *et al.* SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35** (2007).

26 Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Res* **33**, D34-38 (2005).
27 Felsenstein, J. Mathematics vs. Evolution: Mathematical Evolutionary Theory. *Science* **246**, 941-942 (1989).
28 Smits, S. A. & Ouverney, C. C. jsPhyloSVG: a javascript library for visualizing interactive and vector-based phylogenetic trees on the web. *PLoS One* **5**, e12267 (2010).
29 Ouverney, C. C., Armitage, G. C. & Relman, D. A. Single-cell enumeration of an uncultivated TM7 subgroup in the human subgingival crevice. *Appl Environ Microbiol* **69**, 6294-6298 (2003).
30 Schloss, P. D. & Handelsman, J. Introducing SONS, a tool for operational taxonomic unit-based comparisons of microbial community memberships and structures. *Applied and environmental microbiology* **72**, 6773-6779 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements. We thank the various funding agencies who supported this work: NIH SC3GM082291, Howard Hughes Medical Institution 52006312, NIH Minority Access to Research Career (MARC) 5T34GM008253, NIH RISE 1R25GM071381, and NSF REU-RUMBA DBI-0354149.

Author Contributions. **J.D.** qPCR assays in sludge samples, FISH counts, DNA sequence alignment, and phylogenetic tree in ARB. **D.B**. sample collection, genomic DNA extraction, qPCR assays and FISH assays. **J.G.** Human samples genomic DNA extraction, PCR, cloning, and qPCR. **D.S.**, **K.R.**, **F.V.,** and **C.C.** initial screening for environmental sites with TM7. **W.L.** designed TM7a probe. **H.G.** and **H.O.** TM7a probe validation through catFISH. **S.E.S.** FISH and STARFISH assays, jsPhyloSVG-based interactive phylogenetic tree online. **C.O.** principal investigator, managing, and data analysis. All authors contributed to the final version of the manuscript.

Author Information. Relevant and unique 16S rDNA sequences were deposited in GenBank with accession numbers for 6 sludge (HM208132-37) and 14 human oral (HM215440-53) clones. IRB protocol #F0904009 used in the human screening for TM7 and TM7a group phylotypes. Correspondence should be addressed to C.O. at cleber.ouverney@sjsu.edu.

FIGURE CAPTIONS

- Note: Figures attached separately:
    - Fig. 1, Fig. 2, and Fig. 3
    - Supplementary Information: Fig 4

Figure 1. **Neighbor-joining phylogenetic tree of Candidate Division TM7** with 255 TM7 phylotypes, where 160 are identified as being environmental, 42 as animal-associated, and 53 as human-associated. This tree contains 208 unique OTUs by using 99% identity threshold by DOTUR analysis[30]. The bar chart above the outer ring represents the clone counts for each of the 16 phylotypes in our study. Subdivisions 1 and 2 are marked with a gray banner on the inner border of ring. Bootstrap values of major branches are indicated according to legend. An interactive version of this tree (http://www.phylotouch.com/tm7), developed with jsPhyloSVG[28], includes meta-analysis data such as distance matrices and links to sequences and publications.

Figure 2. Quantification of Bacteria, TM7, and TM7a model by (A) qPCR and (B) FISH. Average number of (A1 and B1) Bacteria cells mL^{-1}, (A2 and B2) Candidate division TM7 cells mL^{-1}, and (A3 and B3) TM7a model cells mL^{-1}. Closed circles indicate undetermined quantification values. Error bars represent standard error of the mean. **qPCR**: (A1) Bacteria concentration ranged from 1.30×10^7 (March 2008) to 1.28×10^9 cells ml^{-1} (September 2009) with an overall average of $3.70 \times 10^8 \pm 8.00 \times 10^7$ (mean ± s.e.m.) (n=23) cells ml^{-1} from all samples. (A2) TM7 counts ranged from 3.30×10^5 (June 2009) to 5.84×10^7 cells ml^{-1} (September 2009) with an average of $1.09 \times 10^7 \pm 3.17 \times 10^6$ cells ml^{-1} (n=23) for all samples. (A3) TM7a model counts ranged from 4.02×10^3 (March 2008) to 7.84×10^4 cells ml^{-1} (August 2007) with an average of $2.84 \times 10^4 \pm 1.74 \times 10^4$ cells ml^{-1} (n=4) for all samples. **FISH**: (B2) Concentration of FISH-labeled TM7 cells ranged from 6.94×10^5 to 7.22×10^6 cells ml^{-1} (September 2010) and averaged $2.16 \times 10^6 \pm 7.65 \times 10^5$ (n=9) cells ml^{-1} for all sample sites combined. (B3) TM7a model cell concentrations ranged from 1.08×10^4 (June 2010) to 1.84×10^5 cells mL^{-1} (September 2010), and averaged $6.63 \times 10^4 \pm 1.97 \times 10^4$ cells mL^{-1} (n=9) for all samples.

Figure 3. Micrographs depicting TM7 (red) and TM7a model (blue) in the total microbial community (green) of activated sludge samples through (A-B) FISH and STARFISH (C1-C2). (A) TM7a model short filamentous cells with three segments labeled with both TM7 and TM7a probes. (B) TM7 and TM7a model cells (latter marked with arrows) as coccobacilli and cocci; two insets show TM7a diplo-bacillus morphologies commonly found in sludge. (C) TM7 long filamentous cells (C1, red-orange) taking up a mixture of dissolved tritiated amino acids (C2, bright field) through micro-autoradiography (STARFISH) assay, suggesting TM7 are metabolic active in wastewater. Scale bars = 5 μm.

TABLE

Table 1. 16S rDNA gene amplification parameters for qPCR assay from activated wastewater.

qPCR Assay Targets[a]	Amplification Parameters				Standard Range
	Amplification efficiency ± SD	Slope ± SD	y intercept ± SD	R^2	
General Bacteria	1.02 ± 0.025	-3.29 ± 0.056	37.7 ± 0.80	>.98	10^3-10^8
Division TM7	0.96 ± . 0.019	-3.44 ± 0.051	36.0 ± 0.48	>.99	10^2-10^7
TM7a model	0.92 ± 0.020	-3.53 ± 0.056	35.5 ± 0.65	>.99	10^1-10^6

[a]Ech assay target consisted of three triplicate runs, total sample population (n=9).

SUPPLEMENTARY INFORMATION

Includes Figure 4 (file attached separately), Table 2 and Table 3 (below).

CAPTIONS

Figure 4. Nucleotide-base comparison of aligned 16S rDNA gene sequences obtained from this study and reference sequences in the (*) TM7a Group in Fig.1. Symbols above some bases indicate single nucleotide polymorphism (SNP) between the Activated Wastewater HM208134 (TM7a model) clone and (+) Antecubital fossa HM269723 and Subgingival crevice SBG3 sequences, (Δ) all other members of the TM7a Group, (◊) and all analyzed sequences. This comparison was performed using CLC Main Workbench 5.

Table 2. Primers and probes for Quantitative PCR (qPCR) and Fluorescence *In Situ* Hybridization (FISH) targeting the 16S rDNA gene.

Name	Function	Target	Sequence (5'-3')	T_m (°C) 50 mM Na^+	% GC	Label	Reference
Primers							
BAC-8F	Cloning, qPCR	Bacteria	AGAGTTTGATCMTGGCTCAG	49.7-51.8	45.0-50.0	----------	14
BAC-515R	qPCR	Bacteria	KACCGCGGCKGCTGGCA	54.3-59.1	71.0-82.0	----------	14
TM7-910F	qPCR	TM7	CATAAAGGAATTGACGGGGAC	52.4	48.0	----------	14
TM7-1177R	Cloning, qPCR	TM7	GACCTGACATCATCCCCTCCTTCC	60.8	58.0	----------	14
TM7a-997F	Cloning, qPCR	TM7a Group	TCCCGAGAAGATTTACG	44.6	47.0	----------	This work
HUM-TM7a-1112R	Cloning	Human TM7a	ACAACTAGACACAAGGG	44.6	47.0	----------	This work
ENV-TM7a-1112R	Cloning, qPCR	TM7a model	TCAACTATTCACAAGGG	42.2	41.0	----------	This work
Probes							
TM7-1093	qPCR	TM7	AGTCCATCAACGAGCGCAACC	56.3	57.0	VIC-BHQ	14
TM7a-1033	qPCR	TM7a Group	ATCTGTCACCGAGTTCCA	48.0	50.0	FAM- BHQ	This work
BAC-338I	qPCR	Bacteria	GCTGCCTCCCGTAGGAGT	54.9	67.0	FAM-TAMRA	14
BAC-338II	qPCR	Bacteria	GCAGCCACCCGTAGGTGT	54.9	67.0	FAM- BHQ	14
BAC-338III	qPCR	Bacteria	GCTGCCACCCGTAGGTGT	54.9	67.0	FAM- BHQ	14
TM7-905	FISH	TM7	CCGTCAATTCCTTTATGTTTTA	47.4	32.0	CY5	8
TM7a-1033	FISH	TM7a Group	ATCTGTCACCGAGTTCCA	48.0	50.0	CY3	This work

Table 3. qPCR primer pairs used in assays targeting Bacteria, TM7, and TM7a model with respective amplicon size and annealing temperature(s).

Target Group	Forward Primer	Reverse Primer	Amplicon Size (bp)	Annealing temp. (°C)
Bacteria	BAC-8F	[a]BAC-515R	507	55, 60, 64[a]
Division TM7	TM7-910F	TM7-1177R	267	61
TM7a Group	TM7a-997F	ENV-TM7a-1112R	115	57

[a]Primer optimized to these three annealing temperatures due to degenerative bases as previously described[12].

METHODS

Environmental Site Survey. We surveyed numerous environmental sites with similar characteristics to environments where TM7 had been previously detected[8] to locate a TM7 reservoir. Sites with old forest and garden soils, rhizosphere, seawater, and sludge were analyzed through 16S rDNA pCR, cloning, and sequencing. Aerobic sludge samples were selected due to the presence of clones similar to a human-associated TM7, among other highly diverse TM7 phylotypes. All procedures described below apply to activated wastewater sludge samples only.

Sample Collection, Preparation, and DNA Extraction. A total of 23 samples, each with 200ml of activated sludge, were collected in sterile Whirl-Pak bags (Nasco, Fort Atkinson, WI) from three sites at the San Jose-Santa Clara County Water Pollution Control Plant between August 2007 and January 2010. Samples were homogenized by shaking and for a few of the sludge samples, 40 ml were aliquoted to a sterile 50ml conical tube and fixed in 50% methanol. All sample aliquots were immediately stored in ice and transported to the lab (20 min transit time) for processing. Once in the lab, live samples were immediately processed for genomic DNA extraction and STARFISH[20], whereas fixed samples were prepared for DAPI and FISH counts as explained below. Genomic DNA was extracted from all 23 sludge samples following a bead beater half-lysis protocol[31,32]. First, 30 g of neat sample weighed in sterile 50ml conical tubes were centrifuged at 9,000 g for 15 min at 4°C. The supernatant was discarded and 0.25 g of the pellet was used for genomic DNA extraction.

PCR amplification, Cloning, and Sequencing. Each genomic DNA extraction was first screened for TM7 via PCR amplification of the 16S rDNA gene (~ 1,200 bp) with the broad-range forward primer BAC-8F and the TM7-1177R (Table 2, Supplementary Information). Each PCR reaction consisted of 1x PCR Buffer B (Fisher Scientific, Waltham, Massachusetts), 1.5 mM MgCl2, 0.2 mM of each deoxynucleoside triphosphate, 0.2 pmol of each forward and reverse primer, 1.25x10-2 units of AmpliTaq DNA polymerase (Applied Biosystems, Foster City, California), and 3 to 30 ng of nucleic acid. Screening PCR cycling conditions included a 3 min 96°C hot start, followed by 40 cycles at 94°C for 1 min denaturation, 61°C for 1 min annealing, and 72°C for 1 min elongation, with a final elongation step at 72°C for 3 min. PCR positive samples were amplified *de novo* in a 50 µl PCR reaction volume and 30 PCR cycles to minimize PCR bias[33]. PCR amplicons (1,200 bp) were cleaned in a 2% agarose E-gel and E-gel CloneWell Safe-Imager real-time transilluminator (Invitrogen, Carlsbad, CA), then cloned in Invitrogen pCR2.1-TOPO vector, and transformed into One Shot TOP10 competent cells following Invitrogen protocols for blue-white screening. Plasmids were purified from 303 clones with QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA) and sequenced with M13F primer by Sequetech, Mountain View, CA.

DNA Sequence Analysis. All 16S rDNA partial sequences generated from our clone libraries were cleaned of poor quality bases and PCR primers in CodonCode Aligner v 3.0.2 software (CodonCode Corporation, Dedham, MA), then checked for TM7 similarity through Basic Local Alignment Search Tool (BLAST)[34] in GenBank (27). Sequences with ≥ 98.5% similarity to a human-associated TM7 based on BLAST as well as unique TM7 clones (≤ 98.5% to any clone in GenBank) were fully sequenced. Each consensus sequence (~1,140 bp) was screened for

chimera using Chimera Check in Bellerophon v 3.0 before being aligned using SINA Webaligner (http://www.arb-silva.de/aligner)[25].

Phylogenetic Tree. Aligned sequences were imported into ARB database software package [24] and compared to 541 TM7 reference sequences collected from the SILVA rRNA project[25] and Genebank with length ≥1,200 bp. All sequences were manually aligned based on 16S rDNA primary and secondary structures. A 1,052 unambiguous column filter was generated. A total of 255 Candidate division TM7 sequences were used in the final phylogenetic analysis. A neighbor-joining tree with Felsenstein correction was generated using PHYLIP interference package (28). Tree topography was tested by bootstrap re-sampling analysis of 1,000 replications. Phylotypes were defined using 99% similarity thresholds for approximate species level classification. Bootstrap values ≥ 50% were included. The interactive tree online was generated with jsPhyloSVG[28].

Quantitative PCR (qPCR). Samples with confirmed TM7 sequences were processed through qPCR to quantify the total Bacteria, the total TM7, and the TM7a model 16 rDNA gene copy number in 23 sludge samples. Quantification of each of the three phylogenetic groups was performed in separate qPCR runs, each run with its own standard curve in triplicates, and each full assay was repeated 2-4 times on an ABI 7300 instrument (Applied Biosystems, Foster City, CA). TaqMan primers and probe combinations and assay conditions are listed in Table 3 (Supplementary). TM7a group-specific primers and probes were designed for this work and validated against target plasmids DNA with 1-4 mismatches at primer or probe target site.

The gene copy number from Bacteria, TM7 and TM7a model qPCR reactions were converted to cells ml^{-1} of neat activated wastewater based on these four corrections: i) corrections for nucleic acid dilutions used for PCR amplification, ii) corrections for target DNA concentration, iii) corrections for changes in volume during activated wastewater processing, and iv) an estimated 3.5 average rDNA gene copy number per cell for bacteria quantification [35] or an estimated 2.0 average rDNA gene copy number per cell for TM7 and for TM7a model quantification (22)[21,36].

Each qPCR reaction mixture to quantify total Bacteria, TM7, and TM7a model 16S rDNA gene copy numbers had a final volume of 20 µl made of 10 µl Taqman Universal PCR Master Mix (Applied Biosystems, Foster City, California), an additional $2.5x10^{-2}$ units of AmpliTaq DNA polymerase (Applied Biosystems, Foster City, California), and primer pairs (Table 3) and fluorescent probe (Table 2) (Supplementary). qPCR cycling conditions included an initial denaturation at 95°C for 10 min followed by 50 cycles at 95°C for 30 sec, annealing temperature for Bacteria, TM7, and TM7a model (Table 3) for 30 sec, and 72°C for 30 sec. Fluorescence levels for all qPCR runs were read at the end of each 72°C elongation step.

Fluorescence *In Situ* Hybridization (FISH) and Fluorescence Microscopy. First, we determined sample volumes per well of Teflon-coated slides to generate an average distribution of 100-500 cells per microscopic field of view through DAPI-stained cell counts of the 50% methanol-fixed sludge samples under epifluorescence microscopy as previously explained[29]. Fixed sludge samples were then transferred to Teflon-coated slides with ADD-cell adherence coating (Fisher Scientific, Waltham, Mass). FISH conditions for TM7-905 probe has been published[29]. The TM7a-1033 FISH probe (Table 2) was designed in ARB for this work and

empirically validated at a single nucleotide discrimination via the catFISH[37], a method that relies on expression of heterologous TM7 rRNA targets of varying number of nucleotide mismatches to the TM7a group probe. Hybridization conditions were optimized until only cells with perfect-match rRNA targets were visible through FISH.

Substrate-Tracking AutoRadiography with Fluorescence *In Situ* Hybridization (STARFISH) –STARFISH detects single cell capacity to take up a radioactively labeled nutrient by combining FISH with autoradiography[19,22]. We used STARFISH to determine the metabolic activity of TM7 by adding trace amounts (~5 nM) of a mixture of 15 tritiated amino acids (American Radiolabeled Chemicals, Inc, St Louis, MO) to the sludge and incubating at ambient temperature and aerobic conditions for one hour before fixing the samples in 50% methanol.

Lab Member Skin and Oral Screening. Potential contamination of sludge samples with TM7 bacterial DNA associated with our lab members was tested by PCR screening with primers BAC-8F and HUM-TM7a-1112R, cloning, and sequencing 174 clones (IRB protocol F0904009). From each of 10 individuals, we collected seven samples: six skin samples and one oral cavity sample. Skin samples from the arm, palm of hand, ear, umbilicus, and scalp were collected with a sterile cotton swab moist in sterile MilliQ water, whereas tooth scrapings were collected with a sterile plastic tip by scrapping several tooth surfaces. Each sample was placed in 100 ml sterile MilliQ water, pelleted at 5,000 g for 10min at room temperature, and resuspended in 180 µl enzymatic lysis buffer (20 mM Tris-HCl pH 8.0, 2 mM sodium EDTA, 1.2% Triton X-100, 20 mg/ml Lysozyme). Genomic DNA was extracted using the DNeasy Blood and Tissue Kit following the Gram-Positive Bacteria Protocol (Qiagen, Valencia, CA). Conditions for PCR, cloning and sequence analysis were the same as those described for sludge samples above, except that human samples did not require dilution and the PCR annealing temperature was 58°C.

31 Lepp, P. W. *et al.* Methanogenic Archaea and human periodontal disease. *Proc Natl Acad Sci U S A* **101**, 6176-6181 (2004).

32 Roh, C., Villatte, F., Kim, B. G. & Schmid, R. D. Comparative study of methods for extraction and purification of environmental DNA from soil and sludge samples. *Appl Biochem Biotechnol* **134**, 97-112 (2006).

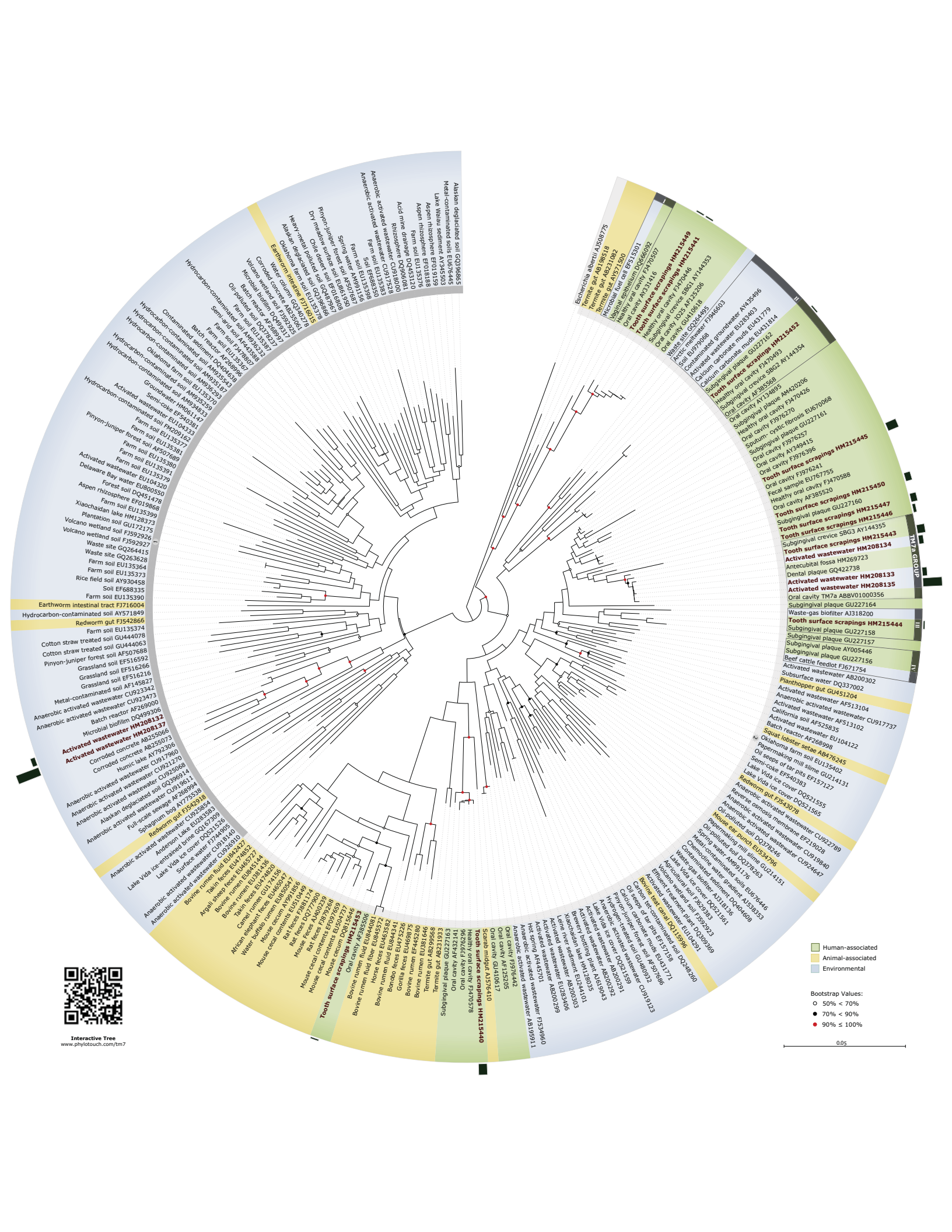
33 Suzuki, M. T. & Giovannoni, S. J. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl Environ Microbiol* **62**, 625-630 (1996).

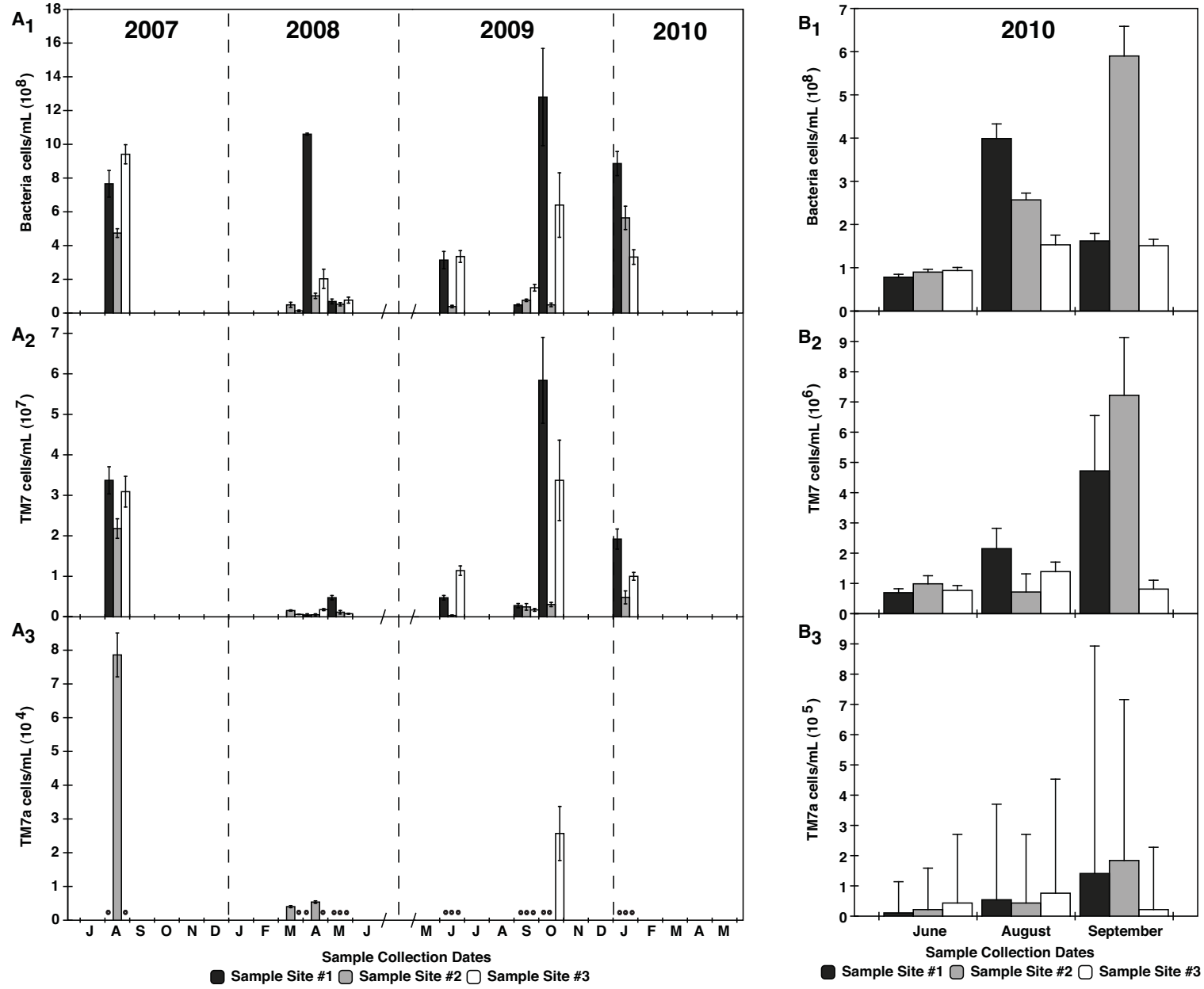
34 McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* **32**, W20-25 (2004).

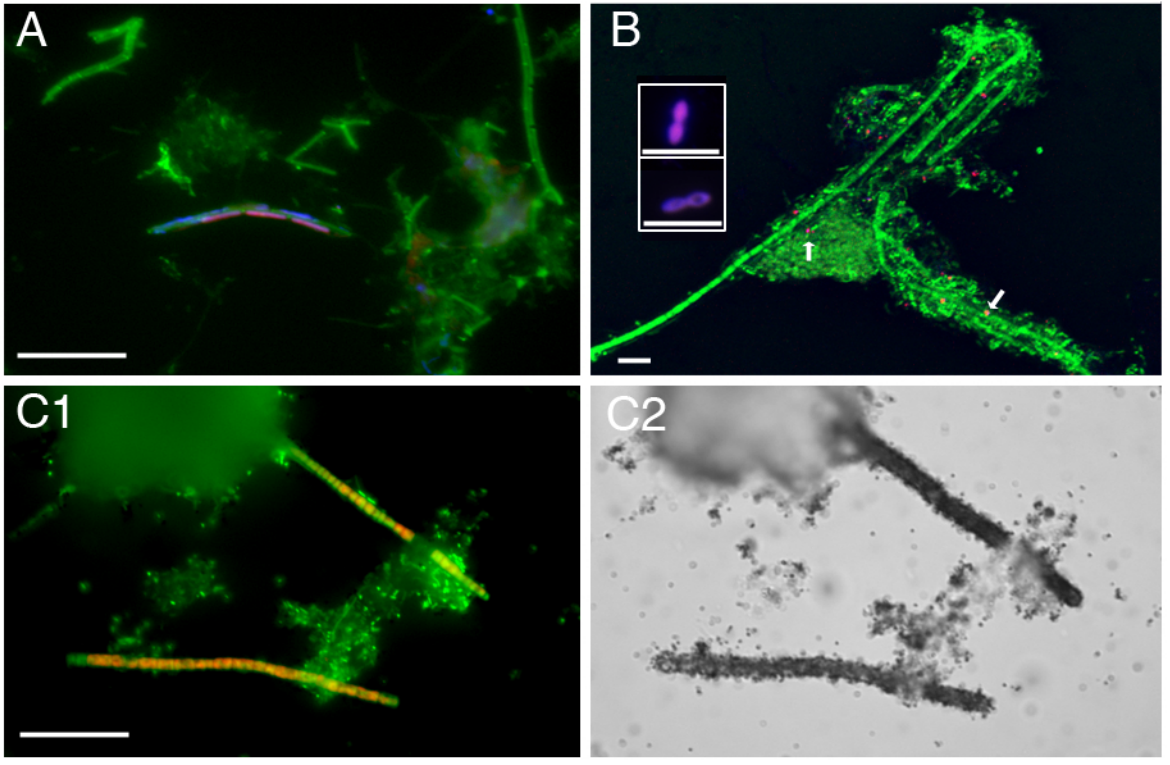
35 Klappenbach, J. A., Saxman, P. R., Cole, J. R. & Schmidt, T. M. rrndb: the Ribosomal RNA Operon Copy Number Database. *Nucleic Acids Research* **29**, 181-184 (2001).

36 Podar, M. *et al.* Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Applied and environmental microbiology* **73**, 3205-3214 (2007).

37 Ouverney, C. C., Armitage, G. C. & Relman, D. A. in *Molecular Microbial Ecology Manual* (eds G. A. Kowalchuk *et al.*) 727-742 (Kluwer Academic Publishers, 2004).


Escherichia albertii AJ508775
Termite gut AB198518
Termite gut AB231082
Termite gut AY571500
Microbial fuel cell EF515301
Vaginal epithelium DQ666092
Healthy oral cavity AY314416
Healthy oral cavity FJ470507
Tooth surface scrapings HM215449
Tooth surface scrapings HM215441
Healthy oral cavity FJ470446
Subgingival crevice SBG1 AY144353
Oral cavity D25 AF126206
Oral cavity GU410618
Waste site GQ264895
Soil EU793068
Arctic meltwater FJ946603
Contaminated groundwater AY435496
Activated wastewater EU283403
Activated wastewater EU431779
Calcium carbonate muds EU431814
Calcium carbonate GU227162
Subgingival plaque GU227160
Tooth surface scrapings HM215452
Subgingival crevice SBG2 AY144354
Healthy oral cavity FJ470493
Subgingival crevice AF385568
Healthy oral cavity AY134895
Subgingival plaque AM420206
Oral cavity AY134896
Subgingival plaque AM420207
Oral cavity FJ470426
Healthy oral cavity FJ976270
Oral cavity EU670068
Sputum- cystic fibrosis EU227161
Subgingival plaque GU227161
Oral cavity FJ976257
Oral cavity AY349415
Oral cavity FJ976396
Tooth surface scrapings HM215445
Oral cavity FJ976241
Fecal sample EU767755
Healthy oral cavity FJ470588
Oral cavity AF385520
Tooth surface scrapings HM215450
Subgingival plaque GU227160
Tooth surface scrapings HM215447
Tooth surface scrapings HM215446
Subgingival crevice SBG3 AY144355
Tooth surface scrapings HM215443
Activated wastewater HM208134
Antecubital fossa HM269723
Dental plaque GQ422738
Activated wastewater HM208133
Activated wastewater HM208135
Oral cavity TM7a ABBV01000356
Subgingival plaque GU227164
Waste-gas biofilter AJ318200
Tooth surface scrapings HM215444
Subgingival plaque GU227158
Subgingival plaque GU227157
Subgingival plaque AY005446
Subgingival plaque GU227156
Beef cattle feedlot FJ671754
Activated wastewater AB200302
Subsurface water DQ337002
Activated wastewater EU104122
Batch reactor AF268998
Oklahoma farm soil EU135402
Papermaking mill slime GU214131
Oil seeps of tar pits EF157127
Lake Vida ice cover DQ521555
Redworm gut FJ543078
Anaerobic activated wastewater CU922789
Reverse osmosis membrane EF219028
Anaerobic activated wastewater CU919840
Mouse ear punch EU538796
Papermaking mill slime GU214151
Metal-contaminated sediment AJ538353
Chemocline water gradient AJ538352
Lake Vida ice cover DQ521561
Agricultural soil DQ369383
Volcanic wetland soil FJ592923
Activated wastewater EU104291
Cow rumen ter-contaminated soil DQ248260
Oil seeps of tar pits EF157158
Activated wastewater CU919123
Calcium carbonate muds EU431771
Pinyon-Juniper forest soil GU480932
Hydrogen-activated wastewater CU921599
Anaerobic activated wastewater AB200231
Lake Vida ice cover DQ521559
Activated wastewater AB200232
Activated wastewater AB200303
Brewery bottling plant EU244101
Activated wastewater EU283406
Xiaochaidan lake HM128035
Leme river sediment AB200399
Activated wastewater AB200302
Anaerobic activated wastewater FJ534960
Acid spring AF445701
Anaerobic activated wastewater AB195911
Oral cavity FJ976442
Oral cavity AF125205
Oral cavity GU410617
Scarab midgut AJ576410
Tooth surface scrapings HM215440
Healthy oral cavity FJ470578
Healthy oral cavity FJ976266
Oral cavity AF432141
Subgingival plaque GU227163
Termite gut AB231933
Termite gut AB299568
Bovine rumen EF445280
Bovine rumen EU381646
Gorilla feces EU468875
Bovine rumen EU475226
Bovine rumen EU463582
Bonobo feces EU844341
Horse feces EU845572
Bovine rumen fluid EU844081
Bovine rumen fluid EF385508
Oral cavity AF385506
Tooth surface scrapings HM215453
Bovine rumen fluid EU815646
Mouse cecal contents EU504737
Mouse cecal contents EF092659
Mouse cecal contents EF092668
Rat feces AJ400239
Rat feces DQ779900
Rat feces FJ881174
Rat feces EU510449
Mouse cecal contents AY991866
Mouse cecal contents EU659547
Water buffalo rumen EU659547
African elephant feces EU474820
Camel rumen EU474836
Takin feces EU841144
Bovine rumen EU381636
Bovine rumen EU474852
Argali sheep feces EU842427
Takin feces EU485577
Bovine rumen fluid EU924140
Anaerobic activated wastewater CU926910
Anaerobic activated wastewater FJ744905
Surface water CU918140
Lake Vida ice-entrained brine DQ521526
Anderson Lake EU285583
Anaerobic activated wastewater CU923854
Redworm gut FJ543918
Sphagnum bog AY775538
Full-scale sewage AF268994
Anaerobic activated wastewater CU919611
Alaskan deglaciated soil GQ396914
Anaerobic activated wastewater CU921270
Anaerobic activated wastewater CU925068
Humic lake AY792306
Corroded concrete AB255066
Corroded concrete AB255073
Activated wastewater HM208137
Activated wastewater HM208132
Microbial biofilm DQ499306
Batch reactor AF269000
Anaerobic activated wastewater CU923473
Anaerobic activated wastewater CU923342
Metal-contaminated soil AF145827
Grassland soil EF516216
Grassland soil EF516266
Grassland soil EF516592
Pinyon-Juniper forest soil AF507688
Cotton straw treated soil GU444063
Cotton straw treated soil GU444078
Farm soil EU135374
Hydrocarbon-contaminated soil AY571849
Redworm gut FJ542866
Earthworm intestinal tract FJ716004
Farm soil EU135390
Soil EF688335
Farm soil EU135373
Farm soil EU135364
Waste site GQ263628
Waste site GQ264415
Rice field soil AY930458
Volcano wetland soil FJ592927
Volcano wetland soil FJ592926
Plantation soil GU172175
Xiaochaidan lake HM128373
Farm soil EU135399
Aspen rhizosphere EF019868
Forest soil DQ451478
Soil EU800550
Delaware Bay water EU104320
Activated wastewater EU104320
Farm soil EU135379
Farm soil EU135391
Farm soil EU135380
Farm soil AF507689
Farm soil EU135377
Pinyon-Juniper forest soil AF507687
Farm soil FM201162
Farm soil EU104333
Hydrocarbon-contaminated soil EU104381
Semi-coke EF540381
Groundwater HM061147
Hydrocarbon-contaminated soil AM935259
Hydrocarbon-contaminated soil AM935837
Hydrocarbon-contaminated soil AM935293
Hydrocarbon-contaminated soil AM935187
Hydrocarbon-contaminated soil AM935543
Oklahoma farm soil EU135398
Hydrocarbon-contaminated soil AM935438
Hydrocarbon-contaminated soil AM935439
Contaminated sediment DQ404308
Hydrocarbon-contaminated soil AM935432
Batch reactor AF268996
Farm soil EU135387
Farm soil FJ478807
Semi-arid soil AF443588
Farm soil EU135387
Oil-polluted soil DQ093927
Microbial biofilm DQ499317
Batch reactor AF268997
Volcano wetland soil FJ592922
Corroded concrete AB255061
Hydrocarbon-contaminated soil DQ234276
Earthworm intestine FJ716015
Oklahoma farm soil EU135378
Alaskan deglaciated soil GQ396896
Heavy-metal polluted soil EU135376
Chile desert soil GQ487966
Dry meadow surface soil EU861950
Pinyon-Juniper forest soil EF076809
Spring water AF507687
Farm soil AM991156
Farm soil EU135398
Soil EF688330
Anaerobic activated wastewater CU917528
Anaerobic activated wastewater EU135383
Rhizosphere DQ918000
Acid mine drainage DQ006081
Aspen rhizosphere DQ453120
Farm soil EU135376
Aspen rhizosphere EF018188
Lake Waiau sediment AY345503
Metal-contaminated soils EU676445
Alaskan deglaciated soil GQ396865
I
II
III
IV
TM7a GROUP
Human-associated
Animal-associated
Environmental
Bootstrap Values:
50% < 70%
70% < 90%
90% ≤ 100%
0.05
Interactive Tree
www.phylotouch.com/tm7


A1
2007
2008
2009
2010
Bacteria cells/mL (10^8)
A2
TM7 cells/mL (10^7)
A3
TM7a cells/mL (10^4)
J A S O N D J F M A M J M J J A S O N D J F M A M
Sample Collection Dates
Sample Site #1 Sample Site #2 Sample Site #3
B1
2010
Bacteria cells/mL (10^8)
B2
TM7 cells/mL (10^6)
B3
TM7a cells/mL (10^5)
June
August
September
Sample Collection Dates
Sample Site #1 Sample Site #2 Sample Site #3


A
B
C1
C2