Journal of Oral Microbiology

COACTION PUBLISHING

ORIGINAL ARTICLE

Host-associated bacterial taxa from Chlorobi, Chloroflexi, GN02, Synergistetes, SR1, TM7, and WPS-2 Phyla/candidate divisions

Anuj Camanocha[1] and Floyd E. Dewhirst[1,2]*

[1]Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Boston, MA, USA; [2]Department of Microbiology, The Forsyth Institute, Cambridge, MA, USA

Background and objective: In addition to the well-known phyla Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, Spirochaetes, Fusobacteria, Tenericutes, and Chylamydiae, the oral microbiomes of mammals contain species from the lesser-known phyla or candidate divisions, including Synergistetes, TM7, Chlorobi, Chloroflexi, GN02, SR1, and WPS-2. The objectives of this study were to create phyla-selective 16S rDNA PCR primer pairs, create selective 16S rDNA clone libraries, identify novel oral taxa, and update canine and human oral microbiome databases.

Design: 16S rRNA gene sequences for members of the lesser-known phyla were downloaded from GenBank and Greengenes databases and aligned with sequences in our RNA databases. Primers with potential phylum level selectivity were designed heuristically with the goal of producing nearly full-length 16S rDNA amplicons. The specificity of primer pairs was examined by making clone libraries from PCR amplicons and determining phyla identity by BLASTN analysis.

Results: Phylum-selective primer pairs were identified that allowed construction of clone libraries with 96–100% specificity for each of the lesser-known phyla. From these clone libraries, seven human and two canine novel oral taxa were identified and added to their respective taxonomic databases. For each phylum, genome sequences closest to human oral taxa were identified and added to the Human Oral Microbiome Database to facilitate metagenomic, transcriptomic, and proteomic studies that involve tiling sequences to the most closely related taxon. While examining ribosomal operons in lesser-known phyla from single-cell genomes and metagenomes, we identified a novel rRNA operon order (23S-5S-16S) in three SR1 genomes and the splitting of the 23S rRNA gene by an I-CeuI-like homing endonuclease in a WPS-2 genome.

Conclusions: This study developed useful primer pairs for making phylum-selective 16S rRNA clone libraries. Phylum-specific libraries were shown to be useful for identifying previously unrecognized taxa in lesser-known phyla and would be useful for future environmental and host-associated studies.

Keywords: *microbiome*; *primers*; *cloning*; *phylogeny*; *16S rRNA*

*Correspondence to: Floyd E. Dewhirst, Department of Microbiology, The Forsyth Institute, 245 First Street, Cambridge, MA, USA, Email: fdewhirst@forsyth.org

To access the supplementary material for this article, please see Supplementary files under Article Tools online

Received: 17 July 2014; Revised: 2 September 2014; Accepted: 4 September 2014; Published: 8 October 2014

The diversity of organisms associated with the human microbiome and the microbiomes of other mammalian species are becoming understood through molecular studies usually involving 16S rRNA analyses. Most of these studies used 'universal' primers, which do a reasonable job of amplifying bacterial 16S rRNA genes in the most common bacterial phyla. However, use of taxa-selective primers is known to allow deeper coverage of specific phyla. For example, in our previous work examining the human oral microbiome (1) and the canine oral microbiome (2), we used Bacteroidetes- and Spirochaeta-selective primers for greater depth of coverage of taxa in these phyla. The majority of bacterial species or phylotypes identified as comprising the human and canine microbiomes are members of the well-known phyla Firmicutes, Proteobacteria, Bacteroidetes, Actinobacteria, Spirochaetes, Fusobacteria, Tenericutes, and Chylamydiae (2, 3). A minority of bacterial species or phylotypes are members of less common phyla, which have at least some cultivated members, including Chlorobi,

Citation: Journal of Oral Microbiology 2014, **6**: 25468 - http://dx.doi.org/10.3402/jom.v6.25468

Chloroflexi, and Synergistetes, or from candidate divisions with no cultivated strains, including GN02, SR1, TM7, and WPS-2. We have previously identified Chlorobi, GN02, and WPS-2 taxa in the canine but not human microbiomes and were interested to see if creation of phyla-selective libraries could lead to identification of novel human-associated members of these lesser-known phyla.

The primary goal of this research was to identify novel human and canine oral taxa by designing selective 16S rDNA polymerase chain reaction (PCR) primer pairs for the lesser-known phyla, and then making and examining clone libraries. The second goal was to place the human and canine taxa within the phylogenetic structure of these phyla/divisions. This was particularly necessary for candidate divisions GN02 and WPS-2 as little has been previously published on phylogenetic structure within these divisions. Finally, recent single-cell and metagenomic studies have produced complete or high coverage genome sequences for several of these as-yet-uncultivated divisions. We describe the addition of key lesser-known genomes to the Human Oral Microbiome Database (www.homd.org) for their utility in taxonomic anchoring of metagenomic, transcriptomic, and proteomic studies of the oral microbiome.

Each of the lesser-known phyla or candidate divisions has been introduced briefly:

Chlorobi. This phylum, previously called the green sulfur bacteria (GSB), has many cultivated species in the class *Chlorobea*, for example, *Chlorobium vibrioforme* (4). At least five other class-level lineages of sequences from uncultivated bacteria have been recognized (5). Recently, the non-photosynthetic organisms, *Ignavibacterium album* (5) and *Melioribacter roseus* (6), were described from a previously uncultivated Chlorobi class. In addition to genomes for photosynthetic *Chlorobea* species, genomes are now available for the non-photosynthetic *I. album* and *M. roseus* species (7, 8). Two canine but no human oral taxa have been previously reported (1, 2) for the Chlorobi phylum.

Chloroflexi. The phylum Chloroflexi, previously called green non-sulfur bacteria (GNS), is found in many environments (9, 10). There are approximately 30 named species in over 20 genera and many unnamed taxa known from 16S rDNA cloning studies. Eight class-level lineages have been recognized, seven with cultivated named members, *Chloroflexi* (10), *Dehalococcoidia*, *Thermomicrobia* (11), *Anaerolineae* (12), *Caldilineae* (12), *Ktedonobacteria* (13), and *Ardenticatenia* (14). An eighth unnamed class is the SAR202 clade (15). Environmental clones also fall into several additional undesignated class-level clades (14). Members of the class *Chloroflexia* are photosynthetic, but characterized members of other classes are not. One human and one canine oral taxon have been described previously (1, 2).

GN02. The candidate division GN02 was first described in a study of the Guerrero Negro hypersaline microbial mat (16). No bacteria from the GN02 candidate division have been cultivated. This division is recognized in the Greengenes taxonomy (17). Recently, single-cell genomes for 201 taxa were described by Rinke et al. (18). The authors proposed the name 'Gracilibacteria' for the GN02 candidate division and suggest placing the phylum in the super phylum 'Patescibacteria'. Three canine but no human oral taxa were recognized in our previous oral microbiome studies (1, 2).

SR1. The SR1 division was named for clones identified in a study of microbial streamers on sediment from the Sulphur River in Parkers Cave, Kentucky, and was previously included as part of candidate division OP11 (19). One of the original clones identified in this division was clone X112 from the human oral cavity (20), which is now known as uncultured SR1 bacterium human oral taxon HOT-345 (1). SR1 clones were seen at multiple human sites including the oral cavity in the Human Microbiome Project (21). Recently, genomes have been deduced for SR1 members of complex environments using various metagenomic techniques (22, 23). One human and three canine oral taxa have been reported previously (1, 2).

Synergistetes. The broad diversity of the phylum Synergistetes is known from 16S rRNA cloning studies, but currently contains at least 14 cultivated species. The phylum was formally named by Jumas-Bilak et al. (24). Clone sequences and the original cultivated species were initially phylogenetically misplaced in the phyla Firmicutes or Deferribacteres causing substantial taxonomic confusion. The status of the phylum is now well-resolved (24, 25). The original cultivated species in the phylum was *Synergistes jonesii*, a rumen bacterium that degrades toxic pyridinediols (26). The phylogeny and taxonomy of strains and clones of human origin have been described (27, 28). Three human-associated oral taxa have been named as follows: *Jonquetella anthropi* (29), *Pyramidobacter piscolens*, (30) and *Fretibacterium fastidiosum* (31). High coverage genome sequences are available for these oral taxa. Thirteen canine and 10 human Synergistetes oral taxa have been previously described (1, 2).

TM7. The candidate division TM7 ('Torf, Mittlere Schicht' = peat, middle layer) is a commonly encountered lineage of bacteria with no cultured representatives (32). TM7 16S rDNA clones have been identified in studies of many habitats, including soils, fresh groundwater, seawater, and mammalian clinical samples (32). They have been detected in the human oral cavity (1, 20, 28, 33–35), the canine oral cavity (2), the human distal esophagus (36), and in several human body sites in the Human Microbiome Project (3). Three single-cell human oral TM7 genomes are available, but they contain significant contamination from *Leptotrichia wadei* and other species

Citation: Journal of Oral Microbiology 2014, **6**: 25468 - http://dx.doi.org/10.3402/jom.v6.25468

and should be considered a multispecies metagenomes. Seven canine and 12 human oral TM7 taxa have been described previously (1, 2).

WPS-2. The candidate division WPS-2 (Writtenberg Polluted Soil) was first described in a study of polychlorinated biphenyl-polluted soil in Germany (37). A single clone from a dog has been given the designation COT-220 (JN713383) (2).

Materials and methods

DNA purification from clinical samples

Purified DNAs from previous studies of human (1) and canine (2) subgingival plaque were pooled separately. The human pool had approximately equal amounts of bacterial DNA from 10 human subjects (approximately 50 ng each) and the canine pool had similar amounts of bacterial DNA from 10 dogs. In brief, DNA was previously purified from plaque samples as follows: Dental plaque from subgingival periodontal pockets was collected using sterile Gracey curettes. Plaque from the curette was transferred into 100 µl of TE buffer (50 mM Tris-HCl, pH 7.6; 1 mM EDTA). DNA extraction was performed using the DNeasy Tissue Kit (Qiagen, Valencia, CA) following the manufacturer's instructions for isolation of genomic DNA from gram-positive bacteria.

Phyla-selective primer design

Previously obtained 16S rRNA reference sequences for lesser-known taxa in human and canine oral microbiomes (1, 2) were used for BLASTN searches of Greengenes and NCBI databases to identify similar sequences. These sequences plus those included in key papers relevant for each taxon were downloaded and manually aligned in our curated secondary structure-based rRNA database (1). The primary design goal was to obtain primer pairs that would specifically amplify 16S rDNA for lineages including known oral taxa. The secondary design goal was to design primer pairs with broad coverage to amplify the entire phylum or candidate division. Because we wanted taxa-selective primer pairs that amplified nearly the entire 16S rDNA gene, primers were designed close to each end of the 16S rDNA molecule: 5′ within base positions 9–46 (*Escherichia coli* numbering) and 3′ within base positions 1464–1509. Target regions were selected by inspection of the alignments and candidate primers analyzed for melting temperature, hairpin formation and primer dimer formation using OligoAnalyzer (Integrated DNA Technologies).

16S rRNA gene amplification

Purified DNA sample pools were amplified with multiple primer sets designed to be selective for Chlorobi, Chloroflexi, GN02, SR1, and WPS-2 16S rDNA genes (Table 1). PCR amplification was performed as previously described at an annealing temperature of 60°C (1).

Cloning procedures

Following the manufacturer's instructions, TOPO TA cloning kits (Invitrogen, Carlsbad, CA) were used to construct clone libraries. The provided TOP10 competent *E. coli* cells were used for transformation and plated onto kanamycin (50 µg/ml) containing Luria-Bertani agar plates, which were incubated overnight at 37°C.

Library screening

Approximately 64 colonies from each library were used directly as a template for colony PCR to amplify the cloned insert with Invitrogen vector M13 (−21) forward and M13 reverse primers. Colonies were picked with a sterile toothpick and dipped into 50-µl reaction mixtures containing 10 pmol of each primer, 40 nmol of deoxynucleotide triphosphates (dNTPs), 1.25 units of DNA Taq Polymerase (NEB, Beverly, MA), and 5-µl 10x PCR buffer (200 mM Tris-HCl, 100 mM $(NH4)_2SO_4$, 20 mM $MgSO_4$, 1% Triton X-100, 100 mM KCl, pH 8.8). We refer to this method as Colony Touch PCR. Electrophoresis on a 1% agarose gel was used to verify the correct amplicon size for representative clones from each library.

16S rDNA data analysis

Preliminary sequence information was determined using reverse primer AE50 (512–533; Supplementary Table 1). Nearly full sequences of approximately 1,500 base sequences were obtained using five additional primers shown in Supplementary Table 1. Sequencing was performed by Macrogen Corp. (Cambridge, MA). The sequencing primers in Supplementary Table 1 were designed for specificity at an annealing temperature of 50°C, which is used by many commercial sequencing companies. Sequences were assembled from the ABI electropherogram files using Sequencher (Gene Codes Corporation, Ann Arbor, MI).

Phylogenetic tree construction

The evolutionary history was inferred using the neighbor-joining method (38). Bootstrap test (500 replicates) was performed using the method of Felsenstein (39). The evolutionary distances were computed using the Jukes–Cantor method (40). Evolutionary analyses were conducted in MEGA6 (41).

16S-IVS-23S rDNA amplification and sequencing of Chloroflexi

Primers for attempted amplification from the 16S into the 23S genes of Chloroflexi rRNA are presented in Supplementary Table 1. A human oral Chloroflexi-specific primer was designed by visual inspection of Chloroflexi sequences aligned with representative diverse human oral taxa. The proposed primer sequence was checked for mispriming by BLASTN analysis against

Table 1. Phylum-selective primers

Primer ID	Phylum/division selectivity	Orientation	Positions	Target sequence[a]	Primer sequence
	E. coli	**Forward**	**9–54**	**GAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCTAACAC**	
F24	'universal'	Forward	9–27	GAGTTTGATYMTGGCTCAG	GAGTTTGATYMTGGCTCAG
AF34	Chlorobi	Forward	22–38	GCTCAGGACGAACGTTG	GCTCAGGACGAACGTTG
AF34-X1	Chlorobi	Forward	22–38	GCTCAGGACGAACGCTG	GCTCAGGACGAACGCTG
AF34-X2	Chlorobi	Forward	22–38	GCTCAGGACGAACGYTG	GCTCAGGACGAACGYTG
AF22	ChloroflexiFamilyAnaerolineae	Forward	38–54	GGCGGCGTGCCTAATAC	GGCGGCGTGCCTAATAC
AF22-X1	ChloroflexiFamilyAnaerolineae	Forward	38–54	GGCGGCGYGCYTAATAC	GGCGGCGYGCYTAATAC
AF25	GN02[L-1]	Forward	19–34	CTGGCTCAGGGTGAAC	CTGGCTCAGGGTGAAC
AF29	GN02[L-1,2,4]	Forward	30–45	TGAACGCTAGCGGTGC	TGAACGCTAGCGGTGC
AF29-X1	GN02[L-3]	Forward	28–45	GATTAAYGCTAGCTGTGC	GATTAAYGCTAGCTGTGC
AF31	SR1	Forward	28–46	GATGAACGCTAGCGAAATG	GATGAACGCTAGCGAAATG
AF31-X1	SR1	Forward	28–46	GATGAACGCTAGCGRAAYG	GATGAACGCTAGCGRAAYG
SYN-X1	SynergistetesnotSpirochaetes	Forward	31–29	TTTGATCCTGGCTCAGGA	TTTGATCCTGGCTCAGGA
TM7-X1	TM7	Forward	22–38	GGCTCAGGATKAAYGCTG	GGCTCAGGATKAAYGCTG
AD81	WPS-2	Forward	18–34	CCTGGCTCAGGACTAAC	CCTGGCTCAGGACTAAC
AD81-X1	WPS-2	Forward	18–34	CCTGGCTCAGGACAAAC	CCTGGCTCAGGACAAAC
AD81-X2	WPS-2	Forward	18–34	CCTGGCTCAGGACWAAC	CCTGGCTCAGGACWAA
AD81-X3	WPS-2	Forward	18–34	CCTGGCTCAGGATTAAC	CCTGGCTCAGGATTAAC
AD81-X4	WPS-2	Forward	18–34	CCTGGCTCAGGAYTAAC	CCTGGCTCAGGAYTAA
AD81-X5	WPS-2	Forward	18–34	CCTGGCTCAGGAYKAAC	CCTGGCTCAGGAYKAAC
AD81-X6	WPS-2	Forward	18–34	CCTGGCTCAGGGCTAAC	CCTGGCTCAGGGCTAAC
	E. coli	**Reverse**	**1464–1509**	**TACCACTTTGTGATTCATGACTGG GGTG AAGTCGTAACAAGGTAAC**	
AD43	'universal'	Reverse	1492–1509	AAGTCGTAACAAGGTADC	GHTACCTTGTTACGACTT
AF35	Chlorobi	Reverse	1476–1493	CTTAGTGACTGG GGCT AA	TTAGCCCCAGTCACTAAG
AF36	Chlorobi/Bacteroidetes	Reverse	1488–1505	GGCT AAGTCGTAACAAGG	CCTTGTTACGACTTAGCC
AF23	Chloroflexi	Reverse	1464–1481	GTCAAAGGTGGGGTTGAT	ATCAACCCCACCTTTGAC
AF24	Chloroflexi/WPS-2/Actinobacteria	Reverse	1488–1505	GACG AAGTCGTAACAAGG	CCTTGTTACGACTTCGTC
AF26	GN02[L-1]	Reverse	1467–1484	CATGGTAGGACTACTGAC	GTCAGTAGTCCTACCATG
AF27	GN02[L-1]	Reverse	1488–1505	GCTT AAGTCGTAACAAGG	CCTTGTTACGACTTAAGC
AF28	GN02[L-1]	Reverse	1488–1505	GCTG AAGTCGTAACAAGG	CCTTGTTACGACTTCAGC
AF33	GN02[L-B]/SR1/TM7	Reverse	1488–1505	GGTT AAGTCGTAACAAGG	CCTTGTTACGACTTAACC
AF32	SR1	Reverse	1475–1494	AATCAGTGACTGG GGTT AAG	CTTAACCCCAGTCACTGATT
M98	Synergistetes/Spirochaetes	Reverse	1483–1501	AGGRG GGTG AAGTCGTAAC	GTTACGACTTCACCCYCCT
M98-X1	Synergistetes/Spirochaetes	Reverse	1483–1501	AGRRG GGTG AAGTCGTAAC	GTTACGACTTCACCCYYCT
AF59	WPS-2	Reverse	1477–1494	CTGATGATTGG GACG AAG	CTTCGTCCCAATCATCAG
AF24	WPS-2/Chloroflexi/Actinobacteria	Reverse	1488–1505	GACG AAGTCGTAACAAGG	CCTTGTTACGACTTCGTC

[a]The *E. coli* sequence in the forward and reverse target sites are shown in bold. The positions in the aligned primers which differ from *E. coli* are underlined.

Citation: Journal of Oral Microbiology 2014, **6**: 25468 - http://dx.doi.org/10.3402/jom.v6.25468

GenBank and examined for Tm, hairpin, and primer dimer formation in OligoAnalyzer. The 23S rRNA primer sequences were based on those previously used to walk the IVS region and sequence the 23S rRNA of helicobacters (42). It is common for different rRNA operons within a bacterium to have distinct 16S-IVS-23S rRNA sequences. Amplicons across this region may produce ambiguous direct sequencing reads, necessitating amplicon cloning as described above to disambiguate the operons.

Additions to the Human Oral Microbiome Database

Full-length 16S rRNA sequences representing novel taxa (based on a full-length 98.5% similarity cutoff definition of phylotype) were assigned human or canine oral taxon numbers (1, 2). Novel human oral taxa found in this study were added to the Human Oral Microbiome Database (http://www.homd.org). Novel canine taxa identified in this study, which expand the canine taxonomy table and 16S rRNA reference set previously described (2), are available upon request from the author (F.E. Dewhirst). Complete or high coverage genome sequences for non-oral taxa that have utility for taxonomic anchoring of metagenomic, transcriptomic and proteomic oral microbiome studies were identified and added as non-oral references to the Human Oral Microbiome Database.

Nucleotide sequences

Sequences for clones from novel taxa identified in this study and for PCR walking were deposited in GenBank with accession numbers: JX294352-JX294356 and KM 018303-KM018332.

Results and discussion

Lesser-known phyla/candidate division primers

Primers designed to selectively amplify the lesser-known taxa are presented in Table 1. The forward primers target the region between *E. coli* positions 18–46 and reverse primers target the region between positions 1464 and 1505. The primers are shown aligned with *E. coli* sequence over the target region in the Target Sequence column. Positions in the primer sequences that differ from *E. coli* are shown underlined to demonstrate how the lesser-known taxa primers differ from the well-known reference Proteobacteria. Several reverse primers take advantage of the variability in the four bases 5′ to the common 'universal' primer 1492–1509 (AD43). Since bases at *E. coli* positions 1488–1491 pair with those at 1409–1412, bases at 1488–1491 could be tentatively inferred for those sequences lacking critical 16S rRNA 3′ sequence data. The region between 1464 and 1487 is quite variable across taxa and primers targeting this region were designed to match the sequences of known oral taxa, but are unlikely to be useful as inclusive primers for all taxa in the phylum/candidate division.

PCR and cloning studies

For the taxa Chlorobi, Chloroflexi, GN02, SR1, and WPS-2, four to 10 primer pairs were tested for their ability to produce 16S rDNA amplicons for human and canine oral sample pools. Amplicons from human DNA pool/primer pair combinations producing strong bands migrating at appropriate molecular weight by gel electrophoresis (data not shown) were cloned and the results are shown in Table 2. Where human samples failed to produce appropriate amplicons, canine amplicons were cloned to allow determination of primer specificity. Libraries were screened by sequencing the 5'-end with reverse primer AE50 (533–521, Supplementary Table 1) and performing BLASTN analysis against the human oral taxon reference set at the Human Oral Microbiome Database (www.homd.org), or against our previously released canine oral taxon reference set (2). For each of the lesser-known taxa examined, at least two primer pairs produced clone libraries with 96–100% of the clone sequences specific for the appropriate phyla/division by BLASTN analysis. Results for a particular library (a cell in Table 2) showing this high specificity are shown in a green font and those with lesser specificity are shown in a brown font. The utility and specificity of the primer pairs for making clone libraries will be discussed below for each lesser-known phylum/division.

Phylogenetic trees

Figures 1–7 show phylogenetic trees created for each lesser-known taxa. Evolutionary history was inferred using the neighbor-joining method (38). In each tree, sequences from this study are shown in bold. Sequences associated with mammalian, insect or other animal sources are marked with a '[H]' for host-associated. Sequences for which there are associated genomes are marked with a '[G]' for genome. The full alignment of sequences used to construct the tree for each phylum is available upon request from the author (F.E. Dewhirst) as a concatenated FASTA file. The figures are described below by phylum/candidate division.

Genome sequences

While high coverage genomes are available for 356 of the human oral taxa at HOMD (as of July 2014), few are available for oral taxa in the lesser-known phyla, particularly those with no cultivated members. Very recently, single-cell and metagenomic sequences have become available for a few uncultivated oral taxa and for several non-oral members of these lesser-known phyla. Supplementary Table 2 provides information on 40 relevant genomes. Because metagenomic, transcriptomic, and proteomic studies require reference DNA or protein sequences to provide a taxonomic anchor, phyla without

Table 2. PCR and clone library screening results

Phylum	Forward primer	Forward primer target	Reverse primer	Reverse primer target	Clones seen in canine clone library	Specificity[a]	Clones seen in human clone library	Specificity[a]
Chlorobi	AF34	Chlorobi	AF35	Chlorobi	Chlorobi	100% (48/48)	ND	ND
Chlorobi	AF34	Chlorobi	AF36	Chlorobi/ bacteroidetes	Chlorobi	100% (46/46)	ND	ND
Chlorobi	F24	'Universal'	AF35	Chlorobi	Chlorobi	100% (48/48)	ND	ND
Chlorobi	F24	'Universal'	AF36	Chlorobi/ bacteroidetes	Chlorobi Bacteroidetes	13% (6/45) 87% (39/45)	ND	ND
Chloroflexi	AF22	Chloroflexi	AF23	Chloroflexi	ND[b]	ND	Chloroflexi	100% (28/28)
Chloroflexi	AF22	Chloroflexi	AF24	Chloroflexi/WPS-2/ Actinobacteria	ND	ND	Chloroflexi	100% (23/23)
Chloroflexi	F24	'Universal'	AF23	Chloroflexi	ND	ND	Actinobacteria Fusobacteria Firmicutes	90% (54/60) 8% (5/60) 2% (1/60)
Chloroflexi	F24	'Universal'	AF24	Chloroflexi/WPS-2/ Actinobacteria	Actinobacteria Chloroflexi WPS-2	79% (34/43) 14% (6/43) 7% (3/43)	Actinobacteria	100% (49/49)
GN02	AF25	GN02[L-1]	AF26	GN02[L-1]	ND	ND	GN02[L-1]	100% (63/63)
GN02	AF25	GN02[L-1]	AF27	GN02[L-1]	ND	ND	GN02[L-1] TM7	96% (55/57) 4% (2/57)
GN02	AF25	GN02[L-1]	AF28	GN02[L-1]	ND	ND	Fusobacteria Firmicutes Ac tinobacteria	83% (15/18) 11% (2/18) 6% (1/18)
GN02	AF29	GN02[All]	AF26	GN02[L-1]	ND	ND	GN02[L-1]	100% (63/63)
GN02	AF29	GN02[All]	AF27	GN02[L-1]	ND	ND	GN02[L-1]	100% (63/63)
GN02	AF29	GN02[All]	AF28	GN02[L-1]	ND	ND	GN02[L-1]	100% (38/38)
GN02	AF29	GN02[All]	AF33	GN02[L-2]/ SR1[L-1,L-2]/TM7	ND	ND	GN02[L-2]	100% (37/37)
GN02	F24	'Universal'	AF26	GN02[L-1]	ND	ND	GN02[L-1] Fusobacteria	98% (62/63) 2% (1/63)
GN02	F24	'Universal'	AF27	GN02[L-1]	ND	ND	GN02[L-1] TM7 Fusobacteria Proteobacteria	62% (40/64) 31% (20/64) 5% (3/64) 2% (1/64)
GN02	F24	'Universal'	AF28	GN02[L-1]	ND	ND	Firmicutes Fusobacteria	72% (43/60) 28% (17/60)
SR1	AF31	SR1[L-1]	AF32	SR1[L-1]	ND	ND	SR1[L-1]	100% (59/59)

Table 2 (*Continued*)

Phylum	Forward primer	Forward primer target	Reverse primer	Reverse primer target	Clones seen in canine clone library	Specificity[a]	Clones seen in human clone library	Specificity[a]
SR1	AF31	SR1[L-1]	AF33	SR1[L-1,L-2]/ GN02[L-2]/TM7	ND	ND	SR1[L-1]	100% (60/60)
SR1	F24	'Universal'	AF32	SR1[L-1]	ND	ND	SR1[L-1]	90% (54/60)
							Campylobacter	10% (6/60)
SR1	F24	'Universal'	AF33	SR1[L-1,L-2]/ GN02[L-2]/TM7	ND	ND	TM7	70% (45/64)
							Fusobacteria	8% (5/64)
							Firmicutes	6% (4/64)
							Bacteroidetes	6% (4/64)
							Proteobacteria	5% (3/63)
							SR1	3% (2/64)
							Actinobacteria	2% (1/64)
WPS-2	AD81	WPS-2	AF59	WPS-2	WPS-2	98% (46/47)	ND	ND
					Actinomyces	2% (1/47)		
WPS-2	AD81	WPS-2	AF24	Chloroflexi/WPS-2/ Actinobacteria	WPS-2	65% (30/46)	Actinobacteria	100% (59/59)
					Actinomyces	35% (16/46)		
WPS-2	F24	'Universal'	AF59	WPS-2	WPS-2	96% (46/48)	ND	ND
					Firmicutes	4% (2/48)		
WPS-2	F24	'Universal'	AF24	Chloroflexi/WPS-2/ actinobacteria	Actinobacteria	79% (34/43)	Actinobacteria	100% (49/49)
					Chloroflexi WPS-2	14% (6/43)		
						7% (3/43)		

[a]Cells for libraries with $<96\%$ specificity are indicated by green font, cells for libraries with lesser specificity are indicated in brown font.
[b]ND indicates Not Done. Clone libraries were not constructed for these primer pair DNA source combinations.

Citation: Journal of Oral Microbiology 2014, **6**: 25468 - http://dx.doi.org/10.3402/jom.v6.25468

reference genomic information become un-linkable and their data are usually discarded. Therefore, we identified genomes from bacterial taxa that are most closely related (by 16S rRNA phylogeny) to human oral taxa in each lesser-known phyla. As indicated in Supplementary Table 2, these genomes have been added to the HOMD database and have been given special non-oral human oral taxon (HOT) numbers. These genomes are discussed below by phylum/candidate division.

Chlorobi

Amplicons were produced from canine samples but not human samples with Chlorobi-selective primer pairs listed in Table 1. Four canine amplicons were cloned and their sequences analyzed to determine specificity. As shown in Table 2, primers pair AF34 and AF35 were highly specific in amplifying canine oral samples, where 100% of clone sequences matched canine oral taxa COT-046 or COT-312 ($>$98.5% similarity). A clone from a feline oral library was $>$98.5% similar to the reference sequence for COT-046 (unpublished data). Thus, Chlorobi taxa have been identified in the oral cavities of at least two mammalian species. The reverse primer AF35 also amplifies 16S rDNA from Bacteroidetes phylum members when paired with the 'universal' forward primer F24.

A 16S rRNA neighbor-joining phylogenetic tree for the Chlorobi phylum is shown in Fig. 1 with Bacteroidetes sequences as an out-group. Six major class-level lineages are marked '1' through '6'. The lineages are numbered as in the publication by Iino et al. (5). Lineage 1 is the class *Ignavibacteria* and lineage 6 is the class *Chlorobia*. Full sequence clones from this study and five clones from the human skin cluster tightly with the two previously identified canine oral taxa, COT-046 and COT-312 (2) in a genus-level clade marked 'A' in lineage 5. In addition to sequences from the mammalian oral cavity, lineage 5 contains sequences from termite hindgut, penguin dropping, and *Paralvinella* mucus. Lineage 5 is designated c_OPB56/otu_1308 in Greengenes (43). We suspect the human skin sample sequences, which fell into clusters of canine oral taxa represent human subjects licked by dogs since the cloning series for those specific human subjects contained sequences for up to 50 canine oral taxa found in dog saliva (2) (unpublished analysis).

Until the description of *Ignavibacterium album* and the class *Ignavibacteria*, only species in the class *Chlorobea* (anaerobic photoautotrophic; GSB) had been grown axenically and named (5). The complete genome of *I. album* has been determined (8). The organism is non-photosynthetic, motile, and capable of organoheterotrophic growth under both oxic and anoxic conditions. A second organism from the class *Ignavibacteria*, *Melioribacter roseus*, has been named and its genome analyzed (6, 7). These authors suggest promoting the class *Ignavibacteria* to the phylum level as *Ignavibacteriae*. Unaddressed by these authors are the fate of the other 5 *Chlorobi* class-level lineages. We do not believe changing the taxonomic level of the class *Ignavibacteria* is justified. Genome information for the three Chlorobi taxa with genomes in Fig. 1 is presented in Supplementary Table 2.

Since our initial Chlorobi 16S rRNA primer designs, we downloaded and aligned many additional relevant sequences for constructing the Chlorobi phylogenetic tree. Examination of the Chlorobi sequence alignment (available from author F.E. Dewhirst) provides additional information on primer coverage and specificity. Forward primer AF34 is a perfect match for members of the clade marked 'B' in Fig. 1. Other members of Chlorobi have the 'T' at position 36 (*E. coli* numbering) replaced by a 'C' (corresponding to sequence AF34-X1 in Table 1). A better Chlorobi-selective forward primer should contain a 'Y' at position 36: AF34-X2. The AF34 primers select for Chlorobi over Bacteroidetes bacteria as they have two base mismatches with Bacteroidetes sequences. The highly selective reverse primer AF35 is useful only for clade 'B' organisms. The more general reverse primer AF36 amplifies both Chlorobi and Bacteroidetes. The primer pair AF34-X2/AF35 should be quite useful for selective PCR 16S rRNA operons of Chlorobi bacteria.

As no lineage 5 organisms have been cultured or their genomes sequenced, essentially no phenotypic or genotypic information is available for the oral host-associated taxa. While this study did not obtain Chlorobi amplicons from the pooled human sample examined, a broader survey of additional human oral samples from a variety of oral sites using the described primers may well identify human host-associated oral Chlorobi taxa.

Chloroflexi

Amplicons were successfully produced from both human and canine oral samples using the Chloroflexi-selective primers in Table 1. One canine and four human libraries were produced and their clones sequenced and analyzed. The specificities of the five libraries are shown in Table 2. Use of the Chloroflexi forward primer AF22 with either Chloroflexi reverse primers AF23 or AF24 produced libraries with 100% Chloroflexi sequences. However, when paired with the universal forward primer F24, both reverse primers yielded mostly or exclusively Actinobacteria libraries. The partial sequencing of the 51 human library Chloroflexi clones showed all had $>$99.3% similarity to the HOT-439 reference sequence AY331414. Full sequences were obtained for four of these clones, which are included in Fig. 2. The canine oral sample cloned from an amplicon with primers F24/AF24 (actually part of the WPS-2 cloning series) produced six Chloroflexi sequences. Two sequences matched the previously described canine oral taxon COT-306 at $>$98.5% similarity (2), but the remaining four sequences appeared to be from novel canine taxa. Full sequencing revealed that they

Citation: Journal of Oral Microbiology 2014, **6**: 25468 - http://dx.doi.org/10.3402/jom.v6.25468

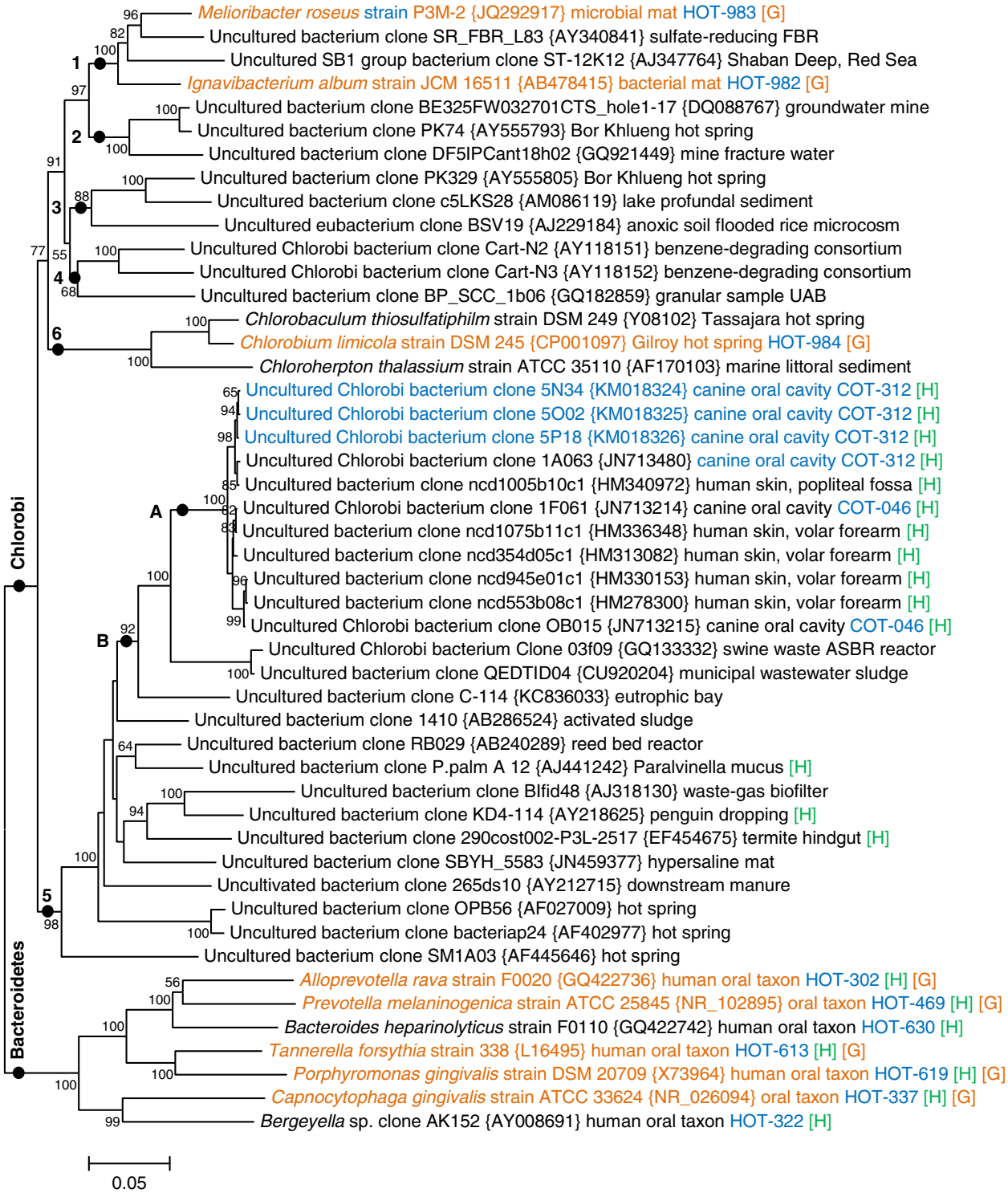

Fig. 1. Neighbor-joining tree for phylum Chlorobi. The evolutionary history was inferred using the neighbor-joining method (38). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (500 replicates) are shown next to the branches (39). The evolutionary distances were computed using the Jukes–Cantor method (40) and are in the units of the number of base substitutions per site. The scale equals 0.05 substitutions per site. Evolutionary analyses were conducted in MEGA6 (41). Taxa in blue text are from this study. Taxa with blue HOT- or COT human or canine oral taxon numbers only are from previous studies (1, 2). Sequences from host-associated samples are designated by a green '[H]'. Taxa in brown text possess genome sequences and are designated in addition by a brown '[G]'. The numbers '1' to '6' adjacent branch points in the tree marked with '●' are major class-level lineages. The letters 'A' and 'B' adjacent branch points in the tree designate clades discussed in text.

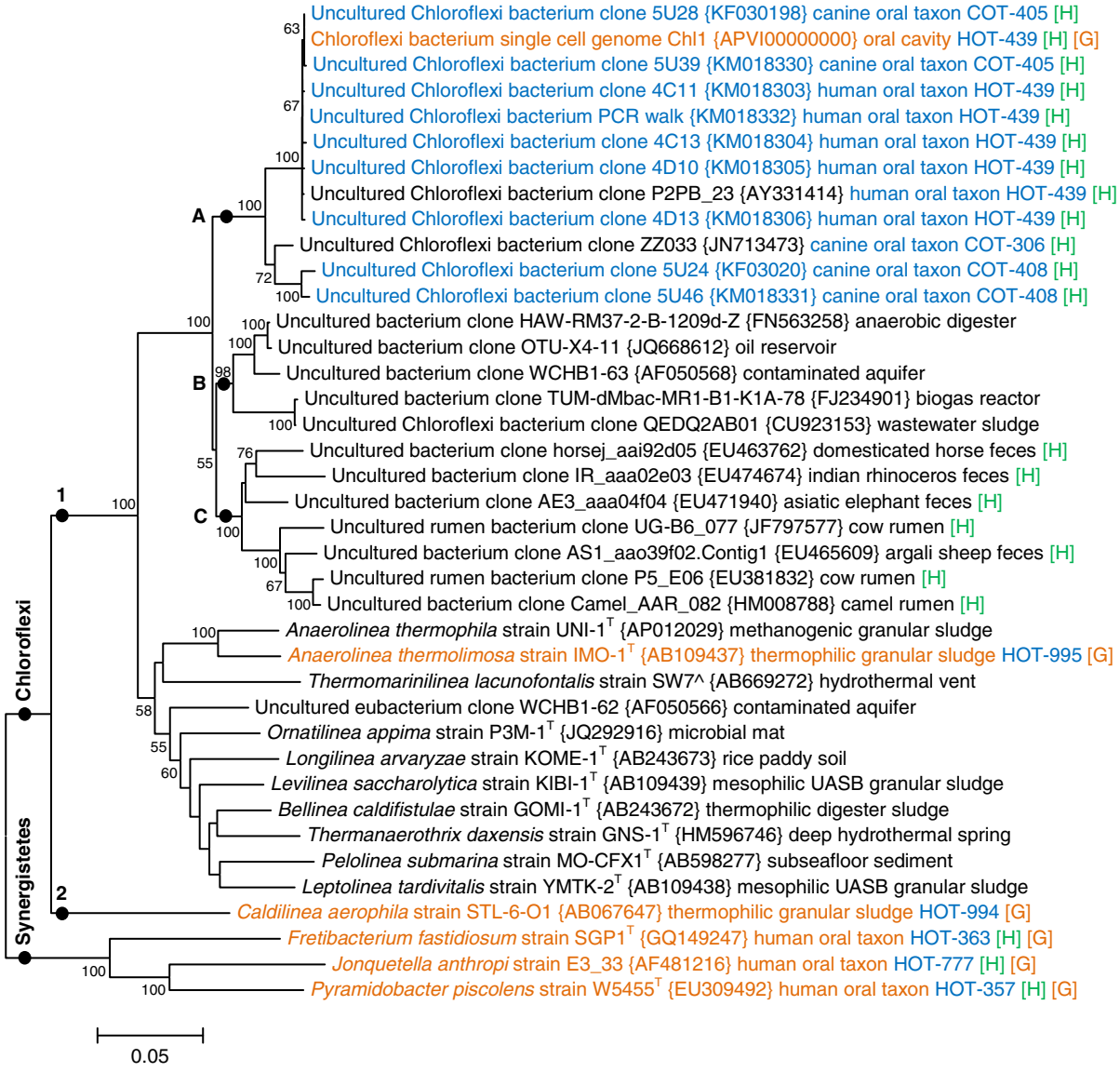

Fig. 2. Neighbor-joining tree for phylum Chloroflexi. Tree construction, text color, and symbols are the same as in Fig. 1. The numbers '1' and '2' denote the classes Anaerolineae and Caldilineae, respectively. The letters 'A' through 'C' designate genus-level clades discussed in text.

represent two additional canine oral taxa that we have designated uncultivated Chloroflexi bacterium COT-405 and COT-408. From a human oral sample used to construct the libraries that yielded HOT-439 clones, we were able to successfully amplify a portion of the rRNA operon from the 16S rRNA gene into the 23S rRNA gene. A single operon sequence was obtained by directly sequencing the PCR amplicon and the intervening sequence contains tRNA Ile with anticodon GAT and tRNA Ala anticodon TGC. The primers used for this study are given in Supplementary Table 1. Sequencing across the IVS region allowed completion of the 3'-end of the 16S rRNA sequence for HOT-439 including the regions under and adjacent to typical 1492 and 1525 'universal' reverse primers. The three 23S rRNA reverse primers described in Supplementary Table 1 appear to have wide specificity and may be useful in amplifying across 16S-23S rRNA IVS region of other lesser-known phyla that lack 3′-end 16S rRNA sequence information or information linking 16S rRNA and 23S rRNA sequences from environmental surveys. Subsequent to completion of this work, single-cell partial genomes for human oral taxon HOT-439 were released in GenBank. Our Chloroflexi 16S-23S fragment sequence (KM018332) and that in Chloroflexi genome assembly Chl1-2 (APVK00000000) are essentially identical (99.8% similarity).

Shown in Fig. 2 is a neighbor-joining tree for the classes *Anaerolineae* (marked '1') and *Caldilineae*

Citation: Journal of Oral Microbiology 2014, **6**: 25468 - http://dx.doi.org/10.3402/jom.v6.25468

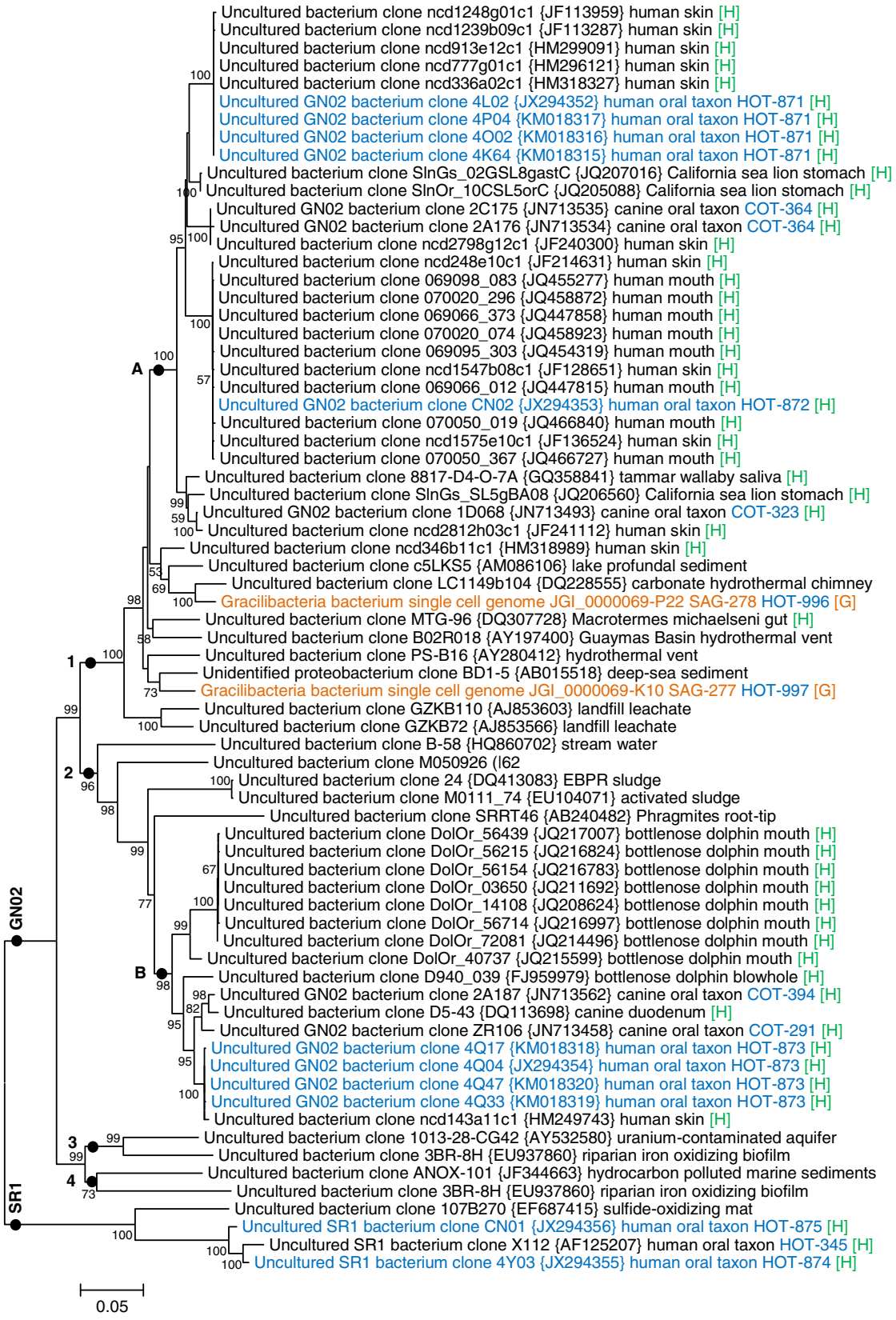

Fig. 3. Neighbor-joining tree for candidate division GN02. Tree construction, text color, and symbols are the same as in Fig. 1. The numbers '1' through '4' denote class-level lineages. The letters 'A' and 'B' are designated two genus level clades of host-associated taxa discussed in text.

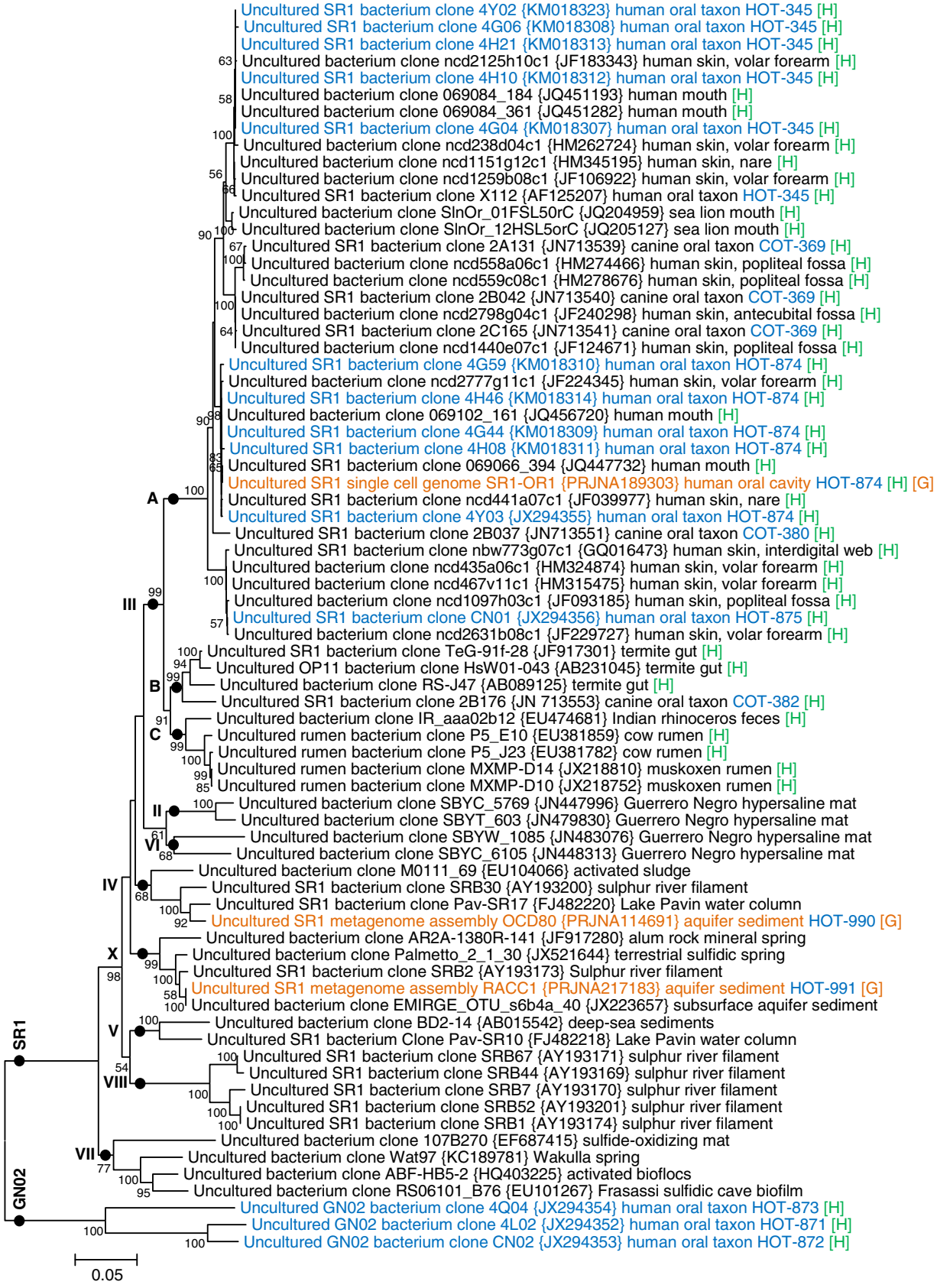

Fig. 4. Neighbor-joining tree for candidate division SR1. Tree construction, text color, and symbols are the same as in Fig. 1. The Roman numerals denote class-level clades as described by Davis et al. (53). Letters 'A' through 'C refer to clades discussed in the text.

Citation: Journal of Oral Microbiology 2014, **6**: 25468 - http://dx.doi.org/10.3402/jom.v6.25468

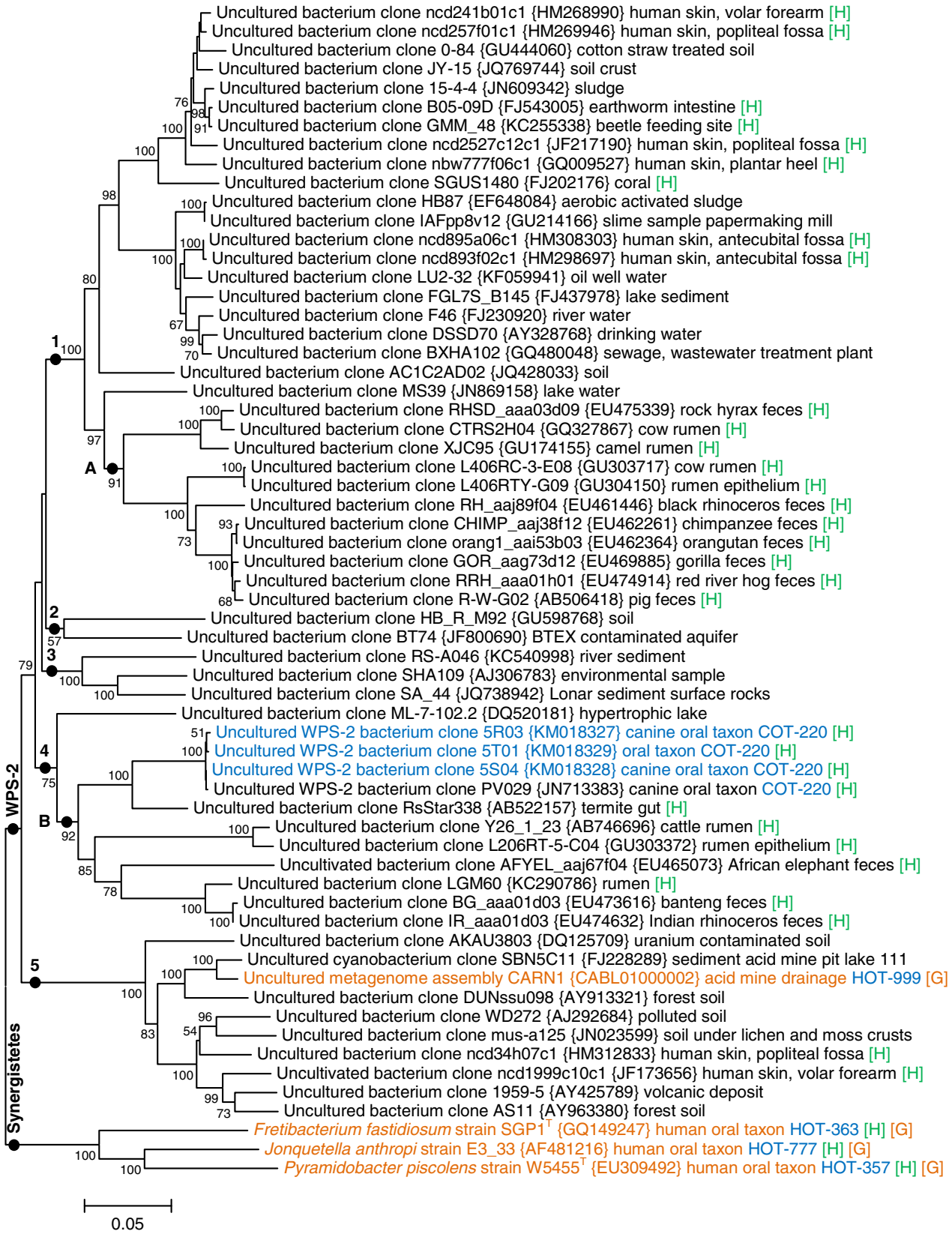

Fig. 5. Neighbor-joining tree for candidate division WPS-2. Tree construction, text color, and symbols are the same as in Fig. 1. The numbers '1' through '5' denote class-level clades. The letters 'A' and 'B designate clades rich in host-associated taxa are discussed in text.

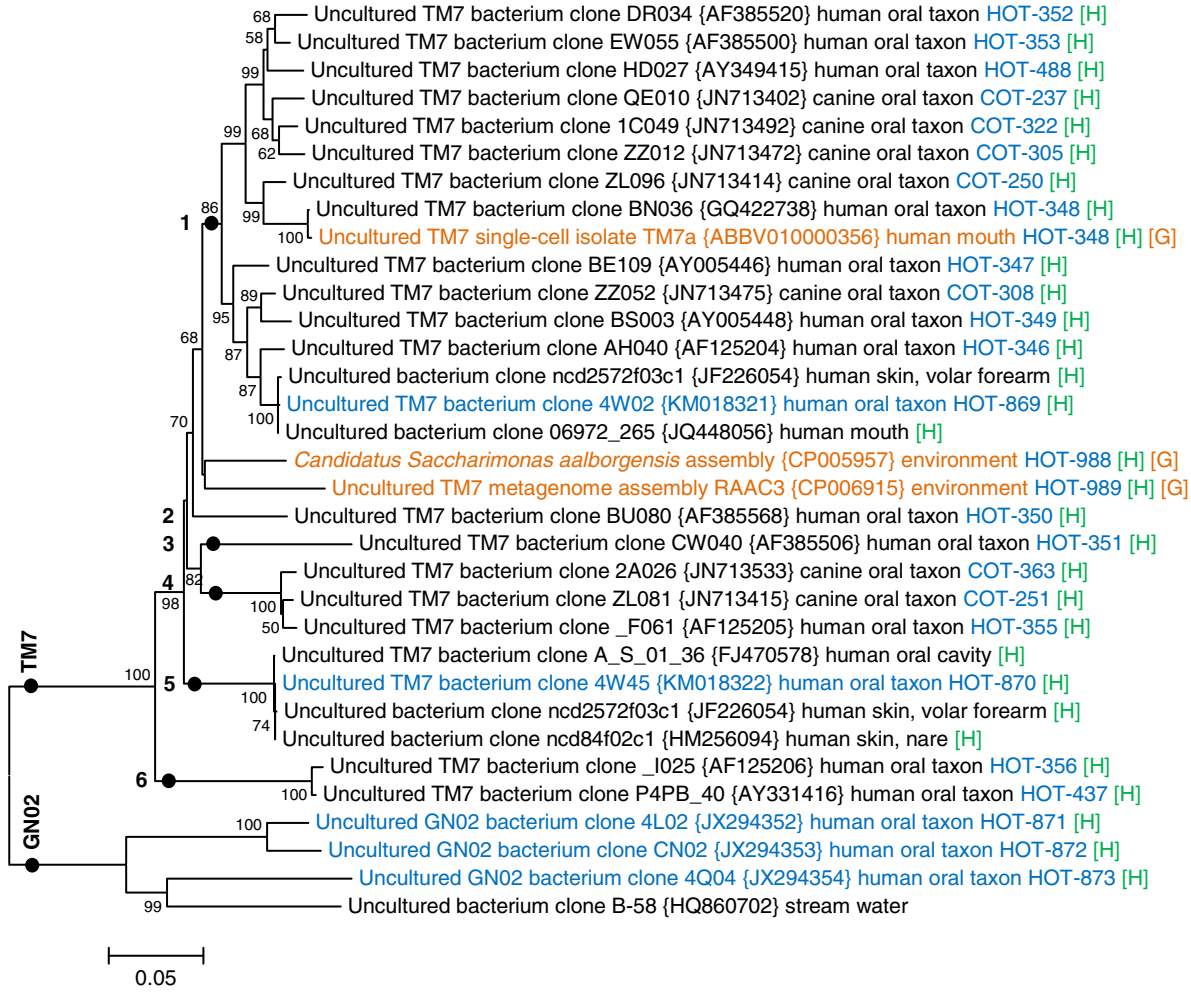

Fig. 6. Neighbor-joining tree for candidate division TM7. Tree construction, text color, and symbols are the same as in Fig. 1. The numbers '1' through '6' designate class-level clades.

(marked '2') of the phylum Chloroflexi with sequences from the phylum Synergistetes as out-group. Additional phylogenetic trees showing all the major clades within the Chloroflexi can be seen in referenced articles (14, 44, 45). Figure 2 includes the 10 named species in the class *Anaerolineae*, selected clones from GenBank with similarity greater than 90% to human oral taxon HOT-439, and the canine and human oral clones from this and previous studies. The oral clones are closest to the cultivated species *Anaerolinea thermophila* (AP012029) with 87% similarity. The dog and human oral clones are closest to a clade of environmental isolates (marked 'B' in Fig. 2.) including the uncultured bacterium clone WCHB1-63 from a contaminated aquifer. At about 92% similarity is a cluster of clone sequences from the feces or rumen of several mammalian species (46) (marked 'C'). It thus appears that the class *Anaerolineae* of phylum Chloroflexi contains two genus-level clades of mammalian host-associated organisms, while the remainder of the phylum is populated by environmental taxa.

It should be noted that human oral taxon HOT-439 and canine oral taxon COT-405 appear to be the same 16S rRNA phylotype ($>$99% similarity). Whether the Chloroflexi bacteria from these two hosts represent one or two species will require comparing phenotypic and genotypic characteristics of isolates from both hosts. Further research obtaining multiple independent oral and fecal samples from various different mammals is needed to determine if host-associated Chloroflexi taxa demonstrate consistent host specificity.

Oak Ridge National Laboratory investigators M. Podar, A. Campbell, and P. Schwientek have obtained two single-cell genomes from cells of Chloroflexi bacterium human oral taxon HOT-439, Chl1 and Chl2, (APVI00000000 and APVJ00000000). Each genome contains about 1,100,000 bp of partial genome information. The two genomes were co-assembled as Chl1-2 (APVK00000000) with 1,766,700 bp and 1693 protein sequences. No paper is currently published on this work, thus there is no estimate of genome completeness.

Citation: Journal of Oral Microbiology 2014, **6**: 25468 - http://dx.doi.org/10.3402/jom.v6.25468

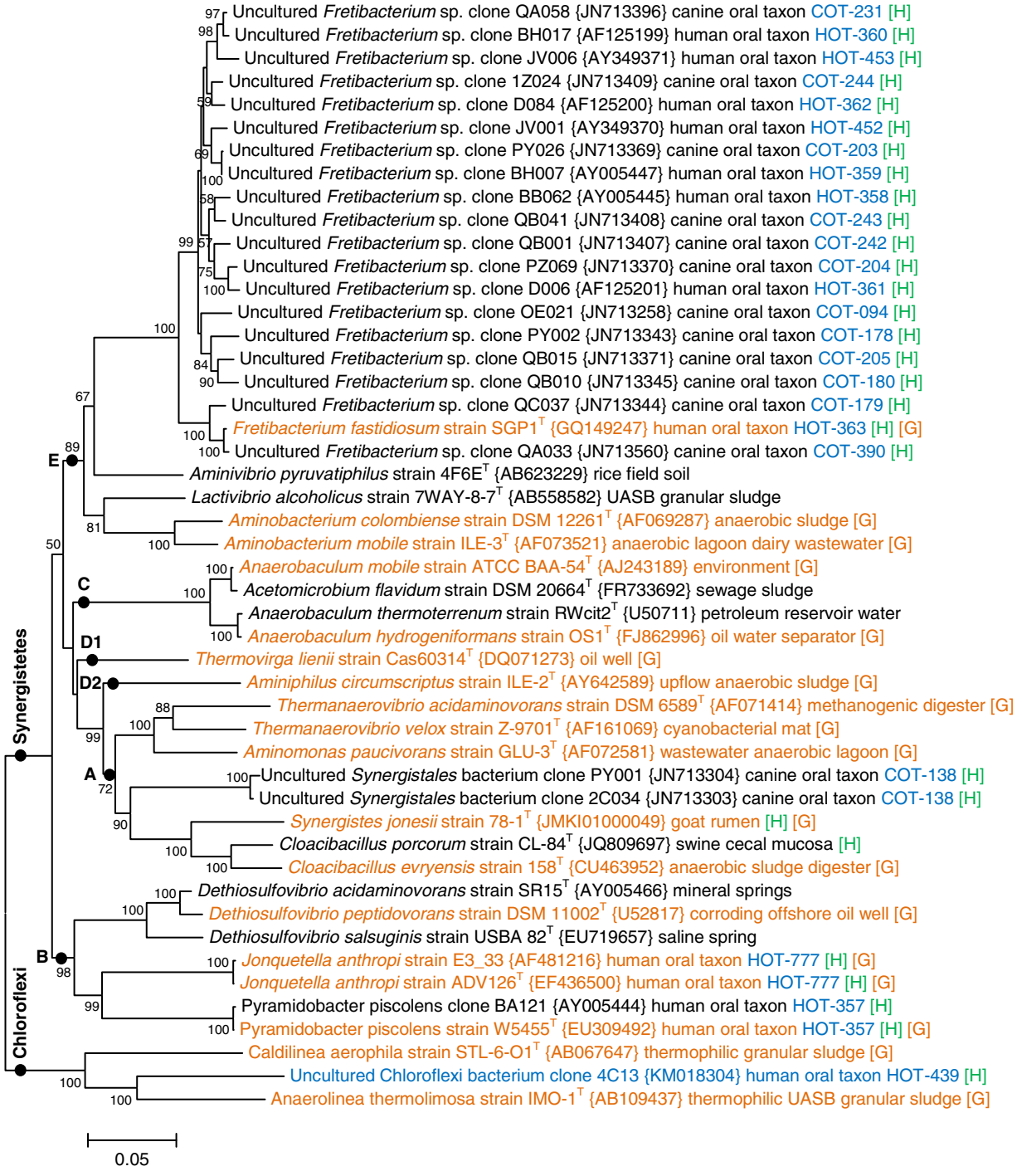

Fig. 7. Neighbor-joining tree for candidate division Synergistetes. Tree construction, text color, and symbols are the same as in Fig. 1. Class level lineages are lettered 'A' through 'E' as in Qiu et al. (62) and Jumas-Bilak et al. (24).

The bacterium *A. thermophila* is the closest sequenced organism to Chloroflexi oral taxon HOT-430. Its complete closed genome, AP012029, was released June 2011 and contains 3,532,378 bp with 3,167 protein sequences, but currently there is no paper describing this work. Sequences for uncultivated Chloroflexi bacterium Chl1-2 and *A. thermophila* are marked with a '[G]' for genome in Fig. 2, and their genomes have been added to Human Oral Microbiome Database. Information on these and three other Chloroflexi genomes are included in Supplementary Table 2. The three environmental genomes are not included in the Chloroflexi tree, Fig. 2, as 16S rRNA sequences were short or less than 80% similar to HOT-439. *A. thermophila* has been given a non-oral reference taxon

designation of HOT-995. The genome of *A. thermophila* should be useful for anchoring to the class *Anaerolineae*, for metagenomic, transcriptomic, and proteomic oral microbiome searches of genes not present in the partial genome of Chloroflexi human oral taxon HOT-439.

Examination of the Chloroflexi sequence alignment used for Fig. 2 (available from F.E. Dewhirst) allowed additional insight as to primer specificity and coverage. Forward primer AF22 is useful for PCR of many bacterial taxa in the class *Anaerolineae* of the phylum Chloroflexi but not for bacteria in other classes. Adding wobbles to the primer (see AF22-X1, Table 1) should slightly increase coverage of *Anaerolineae* bacteria. The reverse primer AF23 appears good for amplification of bacteria in clade 'A', but it is unclear of its utility for clades 'B' and 'C' as most members of these clades lack sequence information in the 1464–1509 region. The AF23 primer does not match sequences for Chloroflexi outside clades 'A', 'B', and 'C'. The reverse primer AF24 matches sequences from Actinobacteria and WPS-2 as well as Chloroflexi and must be paired with a selective forward primer. Thus, while we know of no primer pairs that will selectively amplify the 16S rDNA of bacteria in the entire Chloroflexi phylum, primer pair AF22-X1/AF24 should be useful for amplification of bacteria in the class *Anaerolineae*, the only class known to have host-associated species. It is also noteworthy that the 3′- complete 16S rRNA sequences for Chloroflexi human oral taxon HOT-439 and uncultivated bacterium clone OTU-X4-11 (JQ668612) both contain mismatches with the first two bases of the standard 'universal' 1525–1542 (*E. coli* numbering) reverse primer (Y36 in Table 1). At positions 1525–1526, nearly all bacteria have 'GG' whereas the two fully sequenced members of clades A and B in Fig. 2 contain 'CC'. Genome information for other Chloroflexi such as *A. thermophila* indicates that they have a single base mismatch with the 'universal' primer as they contain a 'GC' at position 1525–1526. The other commonly used 'universal' reverse primer 1492–1509 (AD43 in Table 1) also has a mismatch at the end of the primer at position 1509. The target sequence (not probe) contains a 'C' at position 1509 but the fully sequenced Chloroflexi oral taxon 439 and uncultivated bacterium clone OTU-X4-11 sequences contain 'G's. Thus, 16S rRNA PCR for making clone libraries or next generation sequencing that use either of the two commonly used 'universal' reverse primers may fail to amplify, or at least have a bias against, members of the Chloroflexi phylum.

The 10 currently cultivated species in the class *Anaerolineae* are all gram-negative, non-spore-forming, non-motile, non-photosynthetic, anaerobic, multicellular filamentous chemo-organotrophs (12, 44, 45, 47–50). The human and canine oral taxa are likely to be phenotypically similar to *A. thermophilia* and *Anaerolinea thermolimosa*, and may well be cultivable using conditions as described by Yamada et al. (12).

GN02

Because canine GN02 taxa were identified from two distinct lineages within the phylum, six primers (twice the number for other taxa) were designed in the hope of recovering taxa from both lineages. Using human samples, amplicons were produced for 10 primer pairs and the taxonomic specificity of the libraries is shown in Table 2. Seven libraries contained >96% clones falling in the GN02 division. Forward primers AF25 and AF29 and reverse primers AF26 and AF27 produced libraries with clones only from GN02 lineage 1. The primer pair AF29/AF33 produced clones only from GN02 lineage 2. The primer combination AF25/AF28 failed to produce GN02 clones indicating that the critical sequence at positions 1488–1491 (*E. coli* numbering) for human oral lineage 1 taxa is 'GCT*T*', and not 'GCT*G*' (as suggested by a sequence for one of the initially collected GN02 clones). Finally, use of 'universal' forward primer F24 in place of the GN02-selective primers causes partial to total loss of GN02 specificity in libraries.

Figure 3 shows a 16S rDNA neighbor-joining tree for GN02 with two major class-level clades marked lineage '1' and '2', and two minor lineages are marked '3' and '4'. The tree included the four previously described canine oral taxa COT-323, COT-364, COT-391, and COT-394 (2), human oral clone sequences from this study, and sequences from GenBank including clones from human oral and skin sites, oral sequences from other mammals, and sequences from selected uncultured bacterial GN02 sequences from environmental samples. Full sequences were obtained for one clone from each of our four lineage 1 libraries and from four clones from the lineage 2 library based on their clustering by preliminary 500 base sequences. All of the human oral clone sequences fell into three clusters, which we have designated uncultured GN02 bacterium human oral taxa HOT-871, HOT-872, and HOT-873. The three HOT-groups contain a number of human oral clone sequences from Relman's laboratory (51) and skin clone sequences from Segre's laboratory (52). Within each lineage is a genus-level clade containing only mammalian host-associated clones marked 'A' and 'B' in Fig. 3. Also within clade A are clones from sea lion stomach and wallaby saliva. One human skin clone sequence falls outside clade A in lineage 1. Within clade B are clones from the mouths and blowholes of dolphins (Relman, unpublished GenBank entries) and the canine duodenum. Thus, as in the phylum Chloroflexi, it appears that there are two genus-level mammalian host-associated clades in the division GN02 while the majority of division members have environmental habitats.

Recently, 201 single-cell genomes for more than 20 major uncultivated archaeal and bacterial lineages were

Citation: Journal of Oral Microbiology 2014, **6**: 25468 - http://dx.doi.org/10.3402/jom.v6.25468

described (18). Genomes for two GN02 bacteria were described (18), which fall in lineage 1, Fig. 3. The name 'Gracilibacteria' has been proposed for the GN02 candidate division (18). Based on alignment of 38 conserved marker genes, Rinke et al. proposed the superphylum 'Patescibacteria' containing the phyla 'Gracilibacteria' (GN02), 'Parcubacteria' (OD1), 'Microgenomates' (OP11), SR1, and PER. The 16S rRNA sequences for the single-cell genome taxa are marked with '[G]' and are highlighted in brown in Fig. 3. Genome analysis indicates that taxa within the 'Gracilibacteria' have reassigned the UGA stop codon to code for glycine (18, 23). Genetic code 25 (http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi) should be used for translation DNA to protein or many prematurely terminated open reading frames will result.

The Human Oral Microbiome Database (1) has been updated with the three 'Gracilibacteria' (GN02) taxa HOT-871, HOT-872, and HOT-873. As the single-cell genomes from Ranke et al. (53) are the only genomes available for the phylum, they have been entered into the Human Oral Microbiome Database as 'Gracilibacteria' non-oral reference taxa HOT-996 and HOT-997 to anchor metagenomic, transcriptomic, and proteomic oral microbiome searches.

Examination of the GN02 sequence alignment of taxa in Fig. 3 (available from F.E. Dewhirst) provides additional information on GN02 primer coverage and specificity (Table 1). Forward primer AF25 matches essentially all sequences in lineage 1. Forward primer AF29 matches all lineages 1, 2 and 4. The 16S rDNA of Lineage 3 bacteria should amplify with modified primer AF20-X1. The reverse primers AF26 and AF27 match all sequences in lineage A. The reverse primer AF28, which did not produce any GN02 clones in a human library, does match two sequences in linage 1, including the single-cell genome JGI_0000069-K10. The reverse primer AF33 matches all in lineage B as well as phyla SR1 and TM7 organisms. For PCR of host-associated lineages A and B we recommend primer pairs AF25/AF26 and AF29/AF33, respectively. We have not identified any good GN02 reverse primer inclusive of all lineages as there are at least six variants in the 1488–1491 region for sequences in the GN02 alignment. Design of inclusive primers pairs for PCR of 16S rDNA from GN02 bacteria may be possible and could build off the forward primers AF29 and AF29-X1.

SR1

Amplicons were successfully produced for human and canine oral samples with the four primer pairs presented in Table 1. Using forward primer AF31 and reverse primers AF32 or AF33 the resulting clone libraries produced 100% of clones with SR1 sequences (Table 2). Use of the 'universal' forward primer F24 with AF32 reverse reduced specificity to 90% SR1 clones and with the multi-phyla reverse primer AF33 to only 3% SR1.

A 16S rRNA based phylogenetic tree for SR1 and related taxa are shown in Fig. 4. Most of the sequences included in the initial tree of Harris et al. (19) proposing SR1 are included in Fig. 4. A more detailed description of the diversity of the SR1 division was published by Davis et al. (54) who described nine subgroups or lineages within the division. The Davis et al. 16S rRNA sequences were only partial sequences of approximately 900 bases and are not included in Fig. 4, which only includes sequences of greater than 1,300 bases. However, for Davis' groups III through VIII, sequence representatives with lengths of $>$1,300 bases were identified by BLASTN analysis at GenBank. Davis' groups III through VIII are marked with Roman numerals in Fig. 4. The oral SR1 clones fall into group III. Several clones from our human oral libraries fell within human oral taxon HOT-345; however, other clones clustered into two additional taxa, which we have designated uncultured SR1 bacterial human oral taxa HOT-874 and HOT-875. More than 100 clones from Relman's oral (51) and Segre's skin (52) studies fall within the three human oral taxa HOT-345, HOT-874 and HOT-875. Segre human skin clones also fall within canine oral taxon COT-369. As described above for other phyla, canine oral taxa have been found in human skin libraries from certain subjects. The canine oral taxon COT-369 group also contains a clone from an unpublished feline oral microbiome study, feline oral taxon FOT-313 (F.E. Dewhirst, manuscript in preparation). Two clones from a Relman study of the mouths of sea lions (GenBank, unpublished) form an adjacent clade. Canine oral taxon COT-380 also falls into this genus-level clade of oral taxa marked 'A' in Fig. 4. The remaining canine oral taxon COT-382 clusters with a clade of termite gut clones marked 'B'. A third clade marked 'C' contains cow and musk ox rumen clone sequences and a rhinoceros fecal clone sequence. It appears that lineage III is a class or family level clade comprising clones from host-associated SR1 bacteria, while the other lineages within SR1 appear to be comprised exclusively of environmental clones.

Since our study was initiated, three SR1 genomes have been completed. Two metagenomes from Banfield's group (ACD80 and RAAC1) (22, 55) and a single-cell genome supplemented with human microbiome metagenome scaffolds termed SR1-OR1 have been described (23). The 16S rRNA sequence from the SR1-OR1 assembly falls in human oral taxon HOT-874 though the metagenome sequence likely comes from all three closely related human oral SR1 taxa. The initial paper referring to ACD80 placed it phylogenetically as 'distantly related to group DB1-5′ (55) but can be seen to be a member of SR1 lineage IV in Fig. 4. The subsequent paper by Banfield's group (22) correctly placed the ACD80 and

RACC1 genomes in the SR1 candidate division. The RAAC1 genome 16S rRNA falls in a clade not included in the Davis scheme and we designate lineage 'X' in Fig. 4. The genome for RAAC1 is closed (See Supplementary Table 2) and is well described in the publication (22). Both groups recognized that the UGA stop codon now codes for glycine as in the closely related GN02 division of bacteria and genetic code 25 (http://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi) should be used for translation of DNA to protein. Recoding of stop codons has been examined and discussed by Ivanova et al. (56).

While we were successful in PCR amplification of human plaque DNA from a forward primer specific for Chloroflexi 16S rRNA to reverse primers in the 23S rRNA (described earlier), we were unsuccessful in a similar attempt to amplify from 16S to 23S rRNA using a SR1 oral taxon HOT-345 specific 16S rRNA forward primer and any of three 23S rRNA reverse primers (Supplementary Table 1). Ribosomal rRNA operons for the three recently released SR1 genomes were located in the sequences and analyzed to understand the SR1 16S to 23S rRNA PCR failure. Accession numbers and other information for the three SR1 genomes are presented in Supplementary Table 2. Surprisingly, the rRNA operons identified by BLASTN of the NCBI Whole Genome Shotgun (WGS) database using BioProjectIDs (Supplementary Table 2) were ordered 23S-5S-16S, an organization not previously seen in any organism. The annotated ribosomal RNA operon for the MEGHA/SR1-OR1 single-cell genome (23) is shown in Supplementary Fig. 1. The rRNA operons for RAAC 1 and ACD80 metagenome assemblies are in the identical order and the available by request from the author (F.E. Dewhirst).

Additional insight on the specificity and coverage of the SR1 primers was obtained from examining the aligned sequences for generating the tree in Fig. 4 (available from F.E. Dewhirst). Forward primer AF31 matches sequences for lineage A, but requires two wobbles to cover the rest of lineage III and other lineages, see AF31-X1. Reverse primer AF32 matches all taxa in lineage A with sequences, and probably a majority of SR1 organisms, but many SR1 clones in GenBank were amplified with a 1,390 reverse primer and lack 3′ sequence information. Reverse primer AF33 appears universal for SR1 as well as for TM7 and GN02 lineage B. It appears that the primer pair AF31-X1/AF33 should selectively amplify more than 90% of SR1 bacterial 16S rDNA. Because of the unprecedented operon order in SR1, 16S rRNA operons should be selectively PCR amplified using a 23S rRNA forward primer such as M96 with AF33 reverse to capture the 3′ 500 bases of the 23S, the 5S, and the entire 16S rRNA up through position 1487 (see Supplementary Fig. 1). If any other rare candidate division organisms also have this operon structure and amplify with this primer pair, it would be serendipity rather than a problem.

The genome information obtained to date suggests that SR1 bacteria are fermentative anaerobes, lack tricarboxylic acid cycle and electron transport chain components, and lack flagella. Tripeptide synthase and murein biosynthesis genes suggest that SR1 is likely gram-positive. It is of interest to note that PCR of cells collected from deep-groundwater microorganisms that passed through 0.2-micron-pore-size filters included SR1 clones (57), so at least some SR1 bacteria are physically extremely small physically as well as having small genomes.

WPS-2

Amplicons were screened from human samples using two primer pairs and from canine samples using four primer pairs listed in Table 1. No amplicons were produced for human sample pairs that included the most selective oral WPS-2 reverse primer AF59. The specificity of the clone libraries is shown in Table 2. All of the clones from the human libraries using the multiple phyla-selective AF24 reverse primer (Actinobacteria/Chloroflexi/WPS-2 specificity) were Actinobacteria. Canine libraries using the oral WPS-2-selective AF59 reverse primer with either forward primer contained >96% clones in the WPS-2 division with identity to canine oral taxon COT-220. Thus, the WPS-2 primers apparently work and are selective for samples containing WPS-2 bacteria.

The Greengenes 2010 core set (http://greengenes.lbl.gov/cgi-bin/nph-index.cgi) included just 39 sequences for WPS-2/otu_4420. The updated May 2013 database (http://greengenes.secondgenome.com/downloads) contains 253 sequences assigned to WPS-2. A neighbor-joining phylogenetic tree for WPS-2 is shown in Fig. 5. Approximately, 60 representative WPS-2 sequences were selected for inclusion in the tree as well as Synergistetes sequences as the out-group. We believe Fig. 5 is the first published phylogeny for taxa within the WPS-2 candidate division. There are five major class-level lineages within WPS-2, labeled 1 through 5. The cluster marked 'A' in lineage 1 contains 66 rumen and feces sequences (12 representatives shown), many from the Ley et al. study of the evolution of mammals and their gut microbes (46). Lineages 2 and 3 contain only environmental isolates. Canine oral taxon 220 falls into lineage 4 with seven additional mammalian rumen and feces clone sequences in a cluster marked 'B'. The Writtenberg Polluted Soil clone WD272, for which the candidate division was named, falls in lineage 5. Scattered in the tree are 15 sequences from human skin microbiome studies from Segre's laboratory (52, 58). Thus, WPS-2 host-associated taxa are not confined to specific clades to the extent seen in Chlorobi, Chloroflexi, GN02, and SR1.

Using in-house BLASTN software with sequences from the May 2013 Greengenes dataset, we identified an interesting neighbor of clade 2 clone WD272: 'Carnoules arsenic-contaminated mine drainage metagenome,

Citation: Journal of Oral Microbiology 2014, 6: 25468 - http://dx.doi.org/10.3402/jom.v6.25468

Carnoules bin 1, Contig1, whole genome shotgun sequence' (GenBank entry CABL01000002.1). Unfortunately, the 16S rRNA sequence of this genome is invisible in the NCBI BLAST nr/nt database and can only be found using the NCBI BLAST WGS database with the specific BioProjectID PRJNA41535. The full 16S rRNA sequence was extracted from a contig and its phylogenetic position is shown in Fig. 5. Contig 1 contains a full 16S-23S-5S operon, but the true beginning and end of the rRNAs are poorly annotated in the GenBank entry. Supplementary Figure 2 provides full annotation. The 23S rRNA contains a 707 base intron with a homing endonuclease, which is a member of the LAGLIDADG family and related to I-CeuI, (59). The intron is inserted between positions 1931 and 1932 (*E. coli* numbering), which is adjacent to the location of the I-CeuI cleavage site. This type of intron is common in eukaryotic organelles such as chloroplasts and mitochondria, but relatively rare in bacterial genomes. The endonuclease in the WPS-2 organism genome is most similar to multiple endonucleases found in *Thermotoga* species. Whether similar introns with homing endonucleases are shared by other members of the WPS-2 division remains to be explored. The genome for the WPS-2 organisms is available under WGS project accession CABL00000000 and part of a metagenomic project on organisms from an arsenic-rich acid mine drainage in Carnoules, France (60). Seven genomes were reconstructed, CARN1–CARN7, and five represent uncultivated organisms. Two of the CARN genomes are WPS-2 taxa: CARN1 (the more complete genome containing a 16S rRNA sequence) and CARN4. The CARN1 genome comprises 21 contigs containing 2,539,671 bp with 2,616 proteins. The genome is described by Bertin et al. (60), but the paper does not give the GenBank accession numbers for the genomes and the genome entries in GenBank do not link to the paper. The authors indicated the CARN1 and CARN4 genomes were from organisms that could not be classified using the RDP classifier and suggested CARN1 and CARN4 (PRJNA41541: FN49743-FN49763) represent organisms in a new genus, which they named *Candidatus Fodinabacter communificans* (60). The CARN4 genome cannot be searched by BLAST, because it has not been assigned a WGS BioProjectID accession number. The genes in CARN1 highlighted in the paper include those for flagella, cofactor biosynthesis, and amino acid and purine utilization. This genomic information adds substantially to our understanding genomic capability bacteria in WPS-2 lineage 5. However, WPS-2 lineage 5 is distant from the bulk of the host-associated taxa in lineages 1 and 4 and may be of limited utility as an attractor sequence for proteome, transcriptome, metagenome studies of lineage 1 and 4 organisms.

Examination of the aligned sequences for Fig. 5 (available from F.E. Dewhirst) provides additional information on the specificity and coverage of the WPS-2 primers. The forward primer AD81 matches about 2/3 of the aligned WPS-2 sequences, including most in clade A. The bottom three taxa in clade A have the sequence shown as AD81-X1, and thus an improved clade A forward primer would have the sequence AD81-X2. Most sequences in clade B match forward primer AD81, while others match AD81-X3, thus an inclusive primer for the clade would be AD81-X4. Some sequences in clade B are reported to have a 'G' at position 31, which is the common base in other phyla, but all other WPS-2 sequences have a 'T'. If the 'G' is not a sequencing error, then primer AD81-X5 would be a fully inclusive primer for clade B. A clade marked 'C' differs from other WPS-2 sequences and would require primer AD81-X6. The reverse primer AF59 is useful only for a portion of clade B, canine oral taxon COT-220 sequences and the two rumen clones above them in Fig. 5. The reverse primer AF24 matches all WPS-2 sequences as well as Chloroflexi and some Actinobacteria. Thus, for inclusive PCR of WPS-2 taxa, we suggest using equal parts AD81-X2, -X4, and -X6 forward primers with AF24 reverse primer.

TM7

Because 12 human and seven canine oral associated TM7 taxa had already been described, the research described here did not set out to design primers to seek novel TM7 taxa. However, 67 TM7 clones from human libraries were generated as off-target amplicons in the lower stringency GN02 and SR1 libraries. Approximately, half the clones (initial 500 base sequences) matched references sequences in the HOMD, but others did not and full 16S rRNA sequences were obtained. Two novel taxa were identified and added to HOMD as TM7 bacterium human oral taxa HOT-869 and HOT-870.

Figure 6 presents a phylogenetic tree for human and canine oral TM7 taxa and sequences from GenBank that match human oral taxa HOT-869 and HOT-870. GN02 sequences were used as out-group, but unlike other figures, environmental sequences are not included. For trees showing the broader TM7 diversity, see references by Hugenholtz et al. (32) or Dinis et al. (61). Clone sequences from oral studies by Relman (51) and skin studies by Segre (52) are essentially identical to those reported here. The TM7 taxa associated with canine and human hosts tend to segregate into separate clusters and no canine phylotype is identical to a human phylotype as is the case within the Chloroflexi phylum. The oral clones are scattered into six genus-level clusters labeled 1 to 6 in Fig. 6. One of the three single-cell genomes, TM7a, from Marcy et al. (34), contains a 16S rRNA sequence that places it in human oral taxon HOT-348. While this work was in progress, one complete and three near-complete TM7 genomes have been obtained from

environmental metagenomes (22, 62). The complete closed metagenome assembly by Albertsen et al. (62) of an environmental sample has been given the provisional name *Candidatus Saccharimonas aalborgensis*, and the authors have proposed the name 'Saccharibacteria' for the TM7 candidate division. *Candidatus Saccharimonas aalborgensis* is shown in Fig. 6 and given a non-oral taxon number HOT-988. A metagenome from Kantor et al. (22), assembly RAAC3, has been assigned the non-oral taxon number HOT-989. The TM7 genomes are very small, 845,109 to 974,669 bp. Genome reduction is associated with auxotrophy, which may help explain the reason why TM7 bacteria have not yet been cultivated axenically.

Examination of the aligned sequences for Fig. 6 and the cloning results in Table 2 indicate that reverse primer AF33 (Table 1) matches sequences for TM7 as well as for SR1 and GN02 lineage 2. Paired with a 'universal' forward F24, 70% of clone sequences in a library from human plaque DNA fell in the TM7 division. This specificity may be improved using forward primer TM7-X1, which matches aligned sequences for Fig. 6, in place of 'universal' primer F24.

Synergistetes

Primers for selective PCR of the Synergistetes phylum were not designed for this study as a highly selective reverse primer (M98, Table 1) for PCR and library construction of Synergistetes and Spirochaetes bacteria had been described previously (1, 2). More than a dozen clone libraries using the F24/M98 primer combination contained only taxa from Synergistetes and Spirochaetes phyla.

Figure 7 is a phylogenetic tree for human and canine oral taxa as well as all currently named species in the phylum Synergistetes. Chloroflexi sequences serve as an out-group. Major lineages are lettered as in Qiu et al. (63) and Jumas-Bilak et al. (24). The large clade of oral species in lineage E had no cultivated representatives until Vartoukian et al. cultivated and named *Fretibacterium fastidiosum* (31). We have designated other members of this genus-level group *Fretibacterium* spp. Canine oral taxon COT-138 clusters with *Synergistes jonesii* and two *Cloacibacillus* spp. in lineage A. The two other named cultivated oral taxa, *Jonquetella anthropi* and *Pyramidobacter piscolens*, fall in lineage B with three named *Dethiosulfovibrio* spp. Several human and canine oral taxa have greater than 98.5% sequence similarity and may represent single species. Thus, Synergistetes species may be shared between mammalian hosts, but this will require comparison of characterized isolates.

The Synergistetes phylum currently has 14 genera with 23 named species. Quite remarkably, high coverage genomes are available for 14 of the named species, including the three human oral species (see Supplementary Table 2).

Examination of the aligned sequences for Fig. 7 allows analysis of previously used selective primers for Synergistetes (1, 2). The use of primer pair F24/M98 produced libraries with approximately 50%/50% Synergistetes and Spirochaetes (mostly the genus *Treponema*) from human and canine oral plaque. It should be possible to produce only Synergistetes amplicons by substituting primer SYN-X1 for the universal forward primer F24. Primer SYN-X1 matches most bacteria, but specifically mismatches with Spirochaetes and Proteobacteria at *E. coli* positions 28 and 29.

Conclusions

This study developed useful primers for making phylum-selective 16S rRNA clone libraries for the lesser-known phyla Chlorobi, Chloroflexi, 'Gracilibacteria' (GN02), SR1, Synergistetes, 'Saccharibacteria' (TM7), and WPS-2. Examination of these libraries has led to the identification of seven novel human and two novel canine species level oral taxa. In the majority of phyla examined, host-associated taxa are limited to specific clades. Taxa from mammalian oral samples often occurs in clades distinct from mammalian fecal/rumen samples or from insect samples demonstrating host site and host type specificity. These primers should prove useful for environmental studies of these phyla, and should be particularly useful for finding host-associated bacterial taxa in mammals, birds, fish, insects, and other animal and plant species. Phylogenetic trees for each phylum showing the placement of host-associated taxa in major lineages and in the context of environmental taxa have been presented. Reference genomes for human oral taxa and closely related non-oral taxa were identified in lesser-known phyla and added to the Human Oral Microbiome Database to anchor metagenomic, transcriptomic, and proteomic studies.

Acknowledgements

The authors thank David A. Relman, Julie A. Segre, Norman R. Pace, and other investigators who have contributed vast numbers of nearly full-length high quality 16S rRNA sequences to GenBank for mining by the research community. Research reported in this publication was supported by The National Institute of Dental and Craniofacial Research of the National Institutes of Health under award number R37DE016937. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest and funding

There is no conflict of interest in the present study for any of the authors.

References

1. Dewhirst FE, Chen T, Izard J, Paster BJ, Tanner AC, Yu WH, et al. The human oral microbiome. J Bacteriol 2010; 192: 5002–17.

2. Dewhirst FE, Klein EA, Thompson EC, Blanton JM, Chen T, Milella L, et al. The canine oral microbiome. PLoS One 2012; 7: e36067.
3. Human Microbiome Project C. A framework for human microbiome research. Nature 2012; 486: 215–21.
4. Overmann J, Tuschak C. Phylogeny and molecular fingerprinting of green sulfur bacteria. Arch Microbiol 1997; 167: 302–9.
5. Iino T, Mori K, Uchino Y, Nakagawa T, Harayama S, Suzuki K. *Ignavibacterium album* gen. nov., sp. nov., a moderately thermophilic anaerobic bacterium isolated from microbial mats at a terrestrial hot spring and proposal of *Ignavibacteria* classis nov., for a novel lineage at the periphery of green sulfur bacteria. Int J Syst Evol Microbiol 2010; 60: 1376–82.
6. Podosokorskaya OA, Kadnikov VV, Gavrilov SN, Mardanov AV, Merkel AY, Karnachuk OV, et al. Characterization of *Melioribacter roseus* gen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class *Ignavibacteria*, and a proposal of a novel bacterial phylum Ignavibacteriae. Environ Microbiol 2013; 15: 1759–71.
7. Kadnikov VV, Mardanov AV, Podosokorskaya OA, Gavrilov SN, Kublanov IV, Beletsky AV, et al. Genomic analysis of *Melioribacter roseus*, facultatively anaerobic organotrophic bacterium representing a novel deep lineage within Bacteriodetes/Chlorobi group. PLoS One 2013; 8: e53047.
8. Liu Z, Frigaard NU, Vogl K, Iino T, Ohkuma M, Overmann J, et al. Complete genome of *Ignavibacterium album*, a metabolically versatile, flagellated, facultative anaerobe from the Phylum Chlorobi. Front Microbiol 2012; 3: 185.
9. Bjornsson L, Hugenholtz P, Tyson GW, Blackall LL. Filamentous Chloroflexi (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. Microbiology 2002; 148: 2309–18.
10. Garrity GM, Holt JG. Phylum BVI Chloroblexi phy. nov. In: Boon DR, Castenholtz RW, Garrity GM, editors. *Bergey's manual of systematic bacteriology, second ed.* New York: Springer; 2001. pp. 427–46.
11. Hugenholtz P, Stackebrandt E. Reclassification of *Sphaerobacter thermophilus* from the subclass *Sphaerobacteridae* in the phylum *Actinobacteria* to the class *Thermomicrobia* (emended description) in the phylum Chloroflexi (emended description). Int J Syst Evol Microbiol 2004; 54: 2049–51.
12. Yamada T, Sekiguchi Y, Hanada S, Imachi H, Ohashi A, Harada H, et al. *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes *Anaerolineae* classis nov. and *Caldilineae* classis nov. in the bacterial phylum Chloroflexi. Int J Syst Evol Microbiol 2006; 56: 1331–40.
13. Yabe S, Aiba Y, Sakai Y, Hazaka M, Yokota A. *Thermosporothrix hazakensis* gen. nov., sp. nov., isolated from compost, description of *Thermosporotrichaceae* fam. nov. within the class *Ktedonobacteria Cavaletti et al.* 2007 and emended description of the class *Ktedonobacteria*. Int J Syst Evol Microbiol 2010; 60: 1794–801.
14. Kawaichi S, Ito N, Kamikawa R, Sugawara T, Yoshida T, Sako Y. *Ardenticatena maritima* gen. nov., sp. nov., a ferric iron- and nitrate-reducing bacterium of the phylum 'Chloroflexi' isolated from an iron-rich coastal hydrothermal field, and description of *Ardenticatenia* classis nov. Int J Syst Evol Microbiol 2013; 63: 2992–3002.
15. Morris RM, Rappe MS, Urbach E, Connon SA, Giovannoni SJ. Prevalence of the Chloroflexi-related SAR202 bacterioplankton cluster throughout the mesopelagic zone and deep ocean. Appl Environ Microbiol 2004; 70: 2836–42.
16. Ley RE, Harris JK, Wilcox J, Spear JR, Miller SR, Bebout BM, et al. Unexpected diversity and complexity of the Guerrero Negro hypersaline microbial mat. Appl Environ Microbiol 2006; 72: 3685–95.
17. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006; 72: 5069–72.
18. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature 2013; 499: 431–7.
19. Harris JK, Kelley ST, Pace NR. New perspective on uncultured bacterial phylogenetic division OP11. Appl Environ Microbiol 2004; 70: 845–9.
20. Paster BJ, Boches SK, Galvin JL, Ericson RE, Lau CN, Levanos VA, et al. Bacterial diversity in human subgingival plaque. J Bacteriol 2001; 183: 3770–83.
21. Segata N, Haake SK, Mannon P, Lemon KP, Waldron L, Gevers D, et al. Composition of the adult digestive tract bacterial microbiome based on seven mouth surfaces, tonsils, throat and stool samples. Genome Biol 2012; 13: R42.
22. Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, et al. Small genomes and sparse metabolisms of sediment-associated bacteria from four candidate phyla. mBio 2013; 4: e00708–13.
23. Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, et al. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. Proc Natl Acad Sci USA 2013; 110: 5540–5.
24. Jumas-Bilak E, Roudiere L, Marchandin H. Description of 'Synergistetes' phyl. nov. and emended description of the phylum 'Deferribacteres' and of the family *Syntrophomonadaceae*, phylum 'Firmicutes'. Int J Syst Evol Microbiol 2009; 59: 1028–35.
25. Hugenholtz P, Hooper SD, Kyrpides NC. Focus: Synergistetes. Environment Microbiol 2009; 11: 1327–9.
26. Allison JJ, Mayberry WR, McSweeney CS, Stahl DA. *Synergistes jonesii*, gen. nov.: a rumen bacterium that degrades toxic pyridinediols. Syst Appl Microbiol 1992; 15: 522–9.
27. Marchandin H, Damay A, Roudiere L, Teyssier C, Zorgniotti I, Dechaud H, et al. Phylogeny, diversity and host specialization in the phylum Synergistetes with emphasis on strains and clones of human origin. Res Microbiol 2010; 161: 91–100.
28. Vartoukian SR, Palmer RM, Wade WG. Diversity and morphology of members of the phylum "Synergistetes" in periodontal health and disease. Appl Environ Microbiol 2009; 75: 3777–86.
29. Jumas-Bilak E, Carlier JP, Jean-Pierre H, Citron D, Bernard K, Damay A, et al. *Jonquetella anthropi* gen. nov., sp. nov., the first member of the candidate phylum 'Synergistetes' isolated from man. Int J Syst Evol Microbiol 2007; 57: 2743–8.
30. Downes J, Vartoukian SR, Dewhirst FE, Izard J, Chen T, Yu WH, et al. *Pyramidobacter piscolens* gen. nov., sp. nov., a member of the phylum '*Synergistetes*' isolated from the human oral cavity. Int J Syst Evol Microbiol 2009; 59: 972–80.
31. Vartoukian SR, Downes J, Palmer RM, Wade WG. *Fretibacterium fastidiosum* gen. nov., sp. nov., isolated from the human oral cavity. Int J Syst Evol Microbiol 2013; 63: 458–63.
32. Hugenholtz P, Tyson GW, Webb RI, Wagner AM, Blackall LL. Investigation of candidate division TM7, a recently recognized major lineage of the domain Bacteria with no known pure-culture representatives. Appl Environ Microbiol 2001; 67: 411–19.
33. Brinig MM, Lepp PW, Ouverney CC, Armitage GC, Relman DA. Prevalence of bacteria of division TM7 in human subgingival plaque and their association with disease. Appl Environ Microbiol 2003; 69: 1687–94.
34. Marcy Y, Ouverney C, Bik EM, Losekann T, Ivanova N, Martin HG, 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 USA 2007; 104: 11889–94.
35. Aas JA, Paster BJ, Stokes LN, Olsen I, Dewhirst FE. Defining the normal bacterial flora of the oral cavity. J Clin Microbiol 2005; 43: 5721–32.
36. Pei Z, Bini EJ, Yang L, Zhou M, Francois F, Blaser MJ. Bacterial biota in the human distal esophagus. Proc Natl Acad Sci USA 2004; 101: 4250–5.
37. Nogales B, Moore ER, Llobet-Brossa E, Rossello-Mora R, Amann R, Timmis KN. Combined use of 16S ribosomal DNA and 16S rRNA to study the bacterial community of polychlorinated biphenyl-polluted soil. Appl Environ Microbiol 2001; 67: 1874–84.
38. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987; 4: 406–25.
39. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 1985; 39: 783–91.
40. Jukes TH, Cantor CR. Evolution of protein molecules. New York: Academic Press; 1969.
41. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 2013; 30: 2725–9.
42. Dewhirst FE, Shen Z, Scimeca MS, Stokes LN, Boumenna T, Chen T, et al. Discordant 16S and 23S rRNA gene phylogenies for the genus *Helicobacter*: implications for phylogenetic inference and systematics. J Bacteriol 2005; 187: 6106–18.
43. Hugenholtz P, Pitulle C, Hershberger KL, Pace NR. Novel division level bacterial diversity in a Yellowstone hot spring. J Bacteriol 1998; 180: 366–76.
44. Yamada T, Imachi H, Ohashi A, Harada H, Hanada S, Kamagata Y, et al. *Bellilinea caldifistulae* gen. nov., sp. nov. and *Longilinea arvoryzae* gen. nov., sp. nov., strictly anaerobic, filamentous bacteria of the phylum Chloroflexi isolated from methanogenic propionate-degrading consortia. Int J Syst Evol Microbiol 2007; 57: 2299–306.
45. Imachi H, Sakai S, Lipp JS, Miyazaki M, Saito Y, Yamanaka Y, et al. *Pelolinea submarina* gen. nov., sp. nov., an anaerobic, filamentous bacterium of the phylum *Chloroflexi* isolated from subseafloor sediment offshore Shimokita, Japan. Int J Syst Evol Microbiol 2014; 64: 812–18. doi: 10.1099/ijs.0.057547-0.
46. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, et al. Evolution of mammals and their gut microbes. Science 2008; 320: 1647–51.
47. Sekiguchi Y, Yamada T, Hanada S, Ohashi A, Harada H, Kamagata Y. *Anaerolinea thermophila* gen. nov., sp. nov. and *Caldilinea aerophila* gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain Bacteria at the subphylum level. Int J Syst Evol Microbiol 2003; 53: 1843–51.
48. Gregoire P, Fardeau ML, Joseph M, Guasco S, Hamaide F, Biasutti S, et al. Isolation and characterization of *Thermanaerothrix daxensis* gen. nov., sp. nov., a thermophilic anaerobic bacterium pertaining to the phylum "Chloroflexi," isolated from a deep hot aquifer in the Aquitaine Basin. Syst Appl Microbiol 2011; 34: 494–7.
49. Nunoura T, Hirai M, Miyazaki M, Kazama H, Makita H, Hirayama H, et al. Isolation and characterization of a thermophilic, obligately anaerobic and heterotrophic marine Chloroflexi bacterium from a Chloroflexi-dominated microbial community associated with a Japanese shallow hydrothermal system, and proposal for *Thermomarinilinea lacunofontalis* gen. nov., sp. nov. Microbes Environ/JSME 2013; 28: 228–35.
50. Podosokorskaya OA, Bonch-Osmolovskaya EA, Novikov AA, Kolganova TV, Kublanov IV. *Ornatilinea apprima* gen. nov., sp. nov., a cellulolytic representative of the class *Anaerolineae*. Int J Syst Evol Microbiol 2013; 63: 86–92.
51. Bik EM, Long CD, Armitage GC, Loomer P, Emerson J, Mongodin EF, et al. Bacterial diversity in the oral cavity of 10 healthy individuals. Isme J. 2010; 4: 962–74.
52. Kong HH, Oh J, Deming C, Conlan S, Grice EA, Beatson MA, et al. Temporal shifts in the skin microbiome associated with disease flares and treatment in children with atopic dermatitis. Genome Res 2012; 22: 850–9.
53. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature 2013; 499: 431–7.
54. Davis JP, Youssef NH, Elshahed MS. Assessment of the diversity, abundance, and ecological distribution of members of candidate division SR1 reveals a high level of phylogenetic diversity but limited morphotypic diversity. Appl Environ Microbiol 2009; 75: 4139–48.
55. Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC, et al. Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. Science 2012; 337: 1661–5.
56. Ivanova NN, Schwientek P, Tripp HJ, Rinke C, Pati A, Huntemann M, et al. Stop codon reassignments in the wild. Science 2014; 344: 909–13.
57. Miyoshi T, Iwatsuki T, Naganuma T. Phylogenetic characterization of 16S rRNA gene clones from deep-groundwater microorganisms that pass through 0.2-micrometer-pore-size filters. Appl Environ Microbiol 2005; 71: 1084–8.
58. Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, et al. Topographical and temporal diversity of the human skin microbiome. Science 2009; 324: 1190–2.
59. Jurica MS, Monnat RJ, Jr, Stoddard BL. DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. Mol Cell 1998; 2: 469–76.
60. Bertin PN, Heinrich-Salmeron A, Pelletier E, Goulhen-Chollet F, Arsene-Ploetze F, Gallien S, et al. Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. Isme J. 2011; 5: 1735–47.
61. Dinis JM, Barton DE, Ghadiri J, Surendar D, Reddy K, Velasquez F, et al. In search of an uncultured human-associated TM7 bacterium in the environment. PLoS One 2011; 6: e21280.
62. Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol 2013; 31: 533–8.
63. Qiu YL, Hanada S, Kamagata Y, Guo RB, Sekiguchi Y. *Lactivibrio alcoholicus* gen. nov., sp. nov., an anaerobic, mesophilic, lactate-, alcohol-, carbohydrate- and amino-acid-degrading bacterium in the phylum Synergistetes. Int J Syst Evol Microbiol 2014; 64: 2137–45.