Bacterial Diversity in Human Subgingival Plaque

BRUCE J. PASTER,^{1,2*} SUSAN K. BOCHES,¹ JAMIE L. GALVIN,¹ REBECCA E. ERICSON,¹ CAROL N. LAU,¹ VALERIE A. LEVANOS,¹ ASHISH SAHASRABUDHE,¹ AND FLOYD E. DEWHIRST^{1,2}

Department of Molecular Genetics, The Forsyth Institute, ¹ and Department of Oral Biology, Harvard School of Dental Medicine, ² Boston, Massachusetts

Received 8 November 2000/Accepted 28 March 2001

The purpose of this study was to determine the bacterial diversity in the human subgingival plaque by using culture-independent molecular methods as part of an ongoing effort to obtain full 16S rRNA sequences for all cultivable and not-yet-cultivated species of human oral bacteria. Subgingival plaque was analyzed from healthy subjects and subjects with refractory periodontitis, adult periodontitis, human immunodeficiency virus periodontitis, and acute necrotizing ulcerative gingivitis. 16S ribosomal DNA (rDNA) bacterial genes from DNA isolated from subgingival plaque samples were PCR amplified with all-bacterial or selective primers and cloned into Escherichia coli. The sequences of cloned 16S rDNA inserts were used to determine species identity or closest relatives by comparison with sequences of known species. A total of 2,522 clones were analyzed. Nearly complete sequences of approximately 1,500 bases were obtained for putative new species. About 60% of the clones fell into 132 known species, 70 of which were identified from multiple subjects. About 40% of the clones were novel phylotypes. Of the 215 novel phylotypes, 75 were identified from multiple subjects. Known putative periodontal pathogens such as Porphyromonas gingivalis, Bacteroides forsythus, and Treponema denticola were identified from multiple subjects, but typically as a minor component of the plaque as seen in cultivable studies. Several phylotypes fell into two recently described phyla previously associated with extreme natural environments, for which there are no cultivable species. A number of species or phylotypes were found only in subjects with disease, and a few were found only in healthy subjects. The organisms identified only from diseased sites deserve further study as potential pathogens. Based on the sequence data in this study, the predominant subgingival microbial community consisted of 347 species or phylotypes that fall into 9 bacterial phyla. Based on the 347 species seen in our sample of 2,522 clones, we estimate that there are 68 additional unseen species, for a total estimate of 415 species in the subgingival plaque. When organisms found on other oral surfaces such as the cheek, tongue, and teeth are added to this number, the best estimate of the total species diversity in the oral cavity is approximately 500 species, as previously proposed.

It has been previously estimated that about 500 species of bacteria inhabit the human oral cavity (32, 43, 52). While the majority of these organisms are commensals, a subset of them are likely to be opportunistic pathogens that can cause systemic disease. For example, oral bacteria have been implicated in bacterial endocarditis (3), aspiration pneumonia (42), osteomyelitis in children (10), preterm low birth weight (33), and coronary heart disease and cerebral infarction (or stroke) (2, 53). Consequently, it is important to know what microorganisms are present in the oral cavity for the diagnosis and rational treatment of systemic as well as oral diseases.

Studies with molecular techniques have shown that the bacterial diversity in most environments is severely underestimated in surveys with cultivation-based techniques (1, 19). In many natural environments, less than 1% of the organisms are cultivable (34). Because of the significant effort extended to cultivate oral bacteria, it is thought that about 50% of oral bacteria have been cultivated. However, any understanding of the oral environment requires knowledge of the entire bacterial community. There is no reason to expect that fewer pathogens exist among the uncultivated segment of the community than among the cultivated segment. Highly host-adapted organisms such as *Treponema pallidum* and *Mycoplasma pneu-*

moniae cannot yet be grown or are extremely difficult to grow in culture because they have lost the ability to synthesize many essential molecules that they normally obtain from their host. In previous work, we have found that 75% of oral species of Treponema have not been cultivated (6, 9). The current best model for exploring microbial diversity is based on isolating DNA from the target environment, PCR amplifying the ribosomal DNA (rDNA), cloning the amplicons into Escherichia coli, and sequencing the cloned 16S rDNA inserts (18, 34). These culture-independent molecular phylogenetic methods have been used to deduce the identity of novel phylotypes from periodontitis subjects (6, 45), from dentoalveolar abscesses (11, 50), and from a single subject with mild gingivitis (21). More recently, Sakamoto et al. (41) used these methods to compare the bacterial species and phylotypes in saliva from a healthy subject and two periodontitis subjects. In preliminary studies with similar methods, we identified known species and novel phylotypes in subgingival plaque from subjects with refractory periodontitis (6, 22), adult periodontitis (6, 24), and acute necrotizing ulcerative gingivitis (ANUG) (6, 8); in advanced lesions of noma or facial gangrene (B. J. Paster, W. A. Falkler, Jr., C. O. Enwonwu, E. O. Idigbe, K. O. Savage, V. A. Levanos, M. A. Tamer, R. L. Ericson, C. N. Lau, and F. E. Dewhirst, Abstr. 98th Gen. Meet. Am. Soc. Microbiol., p. 480, 1998); on or in epithelial cells of the tongue dorsum of healthy subjects and subjects with halitosis (4); on or in crevicular epithelial cells from healthy and diseased subjects (25); and in

^{*} Corresponding author. Mailing address: The Forsyth Institute, 140 Fenway, Boston, MA 02115. Phone: (617) 262-5200, ext. 288. Fax: (617) 262-4021. E-mail: bpaster@forsyth.org.

Primer	Function	Position	Orientation	Bacterial specificity	Sequence
D88	PCR	7–27	Forward	Most	GAGAGTTTGATYMTGGCTCAG
E94	PCR	1525-1541	Reverse	Most	GAAGGAGGTGWTCCARCCGCA
C90	PCR	1483-1503	Reverse	Spirochaetes	GTTACGACTTCACCCTCCT
F01	PCR	1487-1508	Reverse	Bacteroidetes	CCTTGTTACGACTTAGCCC
C75	Sequencing	7–27	Forward	Most	GAGAGTTTGATYCTGGCTCAG
B34	Sequencing	344-358	Forward	Most	ACGGGAGGCAGCAGY
F16	Sequencing	789-806	Forward	Most	TAGATACCCYGGTAGTCC
F18	Sequencing	1099-1113	Forward	Most	GCAACGAGCGCAACC
F19	Sequencing	1099-1114	Forward	Bacteroidetes	ATAACGAGCGCAACCC
E94	Sequencing	1522-1541	Reverse	Most	GAAGGAGGTGWTCCARCCGCA
F20	Sequencing	1226-1242	Reverse	Most	CCATTGTARCACGTGTG
F21	Sequencing	1226-1242	Reverse	β-Proteobacteria	CCATTGTATGACGTGTG
F17	Sequencing	907-926	Reverse	Most	CCGTCWATTCMTTTGAGTTT
F15	Sequencing	519-533	Reverse	Most	TTACCGCGGCTGCTG
F22	Sequencing	344-358	Reverse	Most	RCTGCTGCCTCCCGT

TABLE 1. PCR and sequencing primers used in this study

dental plaque from children with early childhood caries (M. R. Becker, A. L. Griffen, E. J. Leys, S. G. Kenyon, S. K. Boches, J. L. Galvin, F. E. Dewhirst, and B. J. Paster, Abstr. 100th Gen. Meet. Am. Soc. Microbiol., p. 244, 2000).

The primary purpose of this study was to determine the diversity of bacteria in human subgingival plaque by using culture-independent methods. The secondary purpose was to obtain qualitative data on the diversity of bacteria in different periodontal disease states and periodontal health. This study reports on the analysis of 2,522 16S rRNA sequences, thus making it an order of magnitude larger than previous 16S rRNA clonal analyses of oral bacteria (6, 11, 21, 41).

MATERIALS AND METHODS

Subject populations. (i) Refractory periodontitis. Refractory periodontitis is defined as failure to respond to at least three forms of therapy, including scaling and root planing, periodontal surgery, and systemically administered tetracycline antibiotic. A poor treatment response was defined as a mean periodontal attachment loss computed on a whole-mouth basis or attachment loss of $>2.5\,$ mm at $>3\,$ periodontal sites occurring within 1 year posttherapy. Attachment refers to the periodontal connective tissue surrounding the root of a tooth that separates it from and attaches it to the alveolar bone. All subjects had at least 20 teeth.

(ii) Periodontally healthy. Periodontally healthy subjects had no pockets >3 mm and no attachment loss >2 mm at any site in the mouth. The subjects had less than 15% of sites with bleeding on probing or with redness.

(iii) Periodontitis. All periodontitis subjects had at least 20 teeth, at least eight sites with pocket depths of >4 mm, and six sites with attachment level of >3 mm. Subjects had received no systemic antibiotics in the previous 6 months and no prior periodontal therapy. Sites may or may not have shown signs of gingival inflammation or suppuration.

(iv) HIV periodontitis. The human immunodeficiency virus (HIV) periodontitis patients were classified according to the criteria used above, except that subjects tested positive for HIV. Stanley C. Holt (presently at The Forsyth Institute) kindly provided these samples.

(v) ANUG. ANUG subjects had no known systemic abnormality. The sites sampled were necrotic and ulcerative at the tips of the interdental papillae.

Microbiological sampling. After removal of supragingival plaque with a sterile Gracey curette, a subgingival plaque sample was removed from the four deepest or most diseased sites with individual sterile Gracey curettes.

Sample lysis. Samples were directly suspended in 50 μ l of a mixture of 50 mM Tris buffer (pH 7.6), 1 mM EDTA (pH 8), 0.5% Tween 20, and 200 μ g of proteinase K per ml. The samples were heated at 55°C for 2 h. Proteinase K was then inactivated by heating at 95°C for 5 min.

Amplification of 16S rRNA cistrons by PCR and purification of PCR products. The 16S rRNA genes were amplified under standard conditions with three different primer sets—all-bacterial selective, *Spirochaetes* selective, and *Bacteroidetes* selective. The sequences of the primers are shown in Table 1. PCR was performed in thin-walled tubes with a Perkin-Elmer 9700 Thermocycler. One

microliter of the DNA template was added to a reaction mixture (50-µl final volume) containing 20 pmol of each primer, 40 nmol of deoxynucleoside triphosphates (dNTPs), and 1 U of *Taq* 2000 polymerase (Stratagene, La Jolla, Calif.) in buffer containing Tagstart antibody (Sigma Chemical Co.). In a hot start protocol, samples were preheated at 95°C for 8 min followed by amplification under the following conditions: denaturation at 95°C for 45 s, annealing at 60°C for 45 s, and elongation for 1.5 min with an additional 5 s for each cycle. A total of 30 cycles were performed, which was followed by a final elongation step at 72°C for 10 min. The results of PCR amplification were examined by electrophoresis in a 1% agarose gel. DNA was stained with ethidium bromide and visualized under short-wavelength UV light.

Cloning procedures. Cloning of PCR-amplified DNA was performed with a Zero Blunt Cloning kit, TOPO TA Cloning kit (Invitrogen, San Diego, Calif.), or Prime PCR Cloner cloning system (5'-3', Inc., Boulder, Colo.) according to the manufacturers' instructions. Transformation was done with competent *E. coli* TOP10 cells provided by the manufacturer. The transformed cells were then plated onto Luria-Bertani agar plates supplemented with kanamycin and incubated overnight at 37°C. Colonies were then placed into 40 μl of 10 mM Tris. One microliter was used as the template to determine the correct sizes of inserts in PCR with an M13 (–40) forward primer and an M13 reverse primer. The size of inserts (approximately 1,500 bp) was determined by PCR with flanking vector primers followed by electrophoresis on a 1% agarose gel.

16S rRNA sequencing. Purified DNA from PCR was sequenced with an ABI Prism cycle-sequencing kit (BigDye Terminator Cycle Sequencing kit with AmpliTaq DNA polymerase FS; Perkin-Elmer). The primers in Table 1 were used for sequencing. Quarter dye chemistry was performed with 80 μ M primers and 1.5 μ l of PCR product in a final volume of 20 μ l. Cycle sequencing was performed with an ABI 9700 sequencer with 25 cycles of denaturation at 96°C for 10 s and annealing and extension at 60°C for 4 min. Sequencing reactions were run on an ABI 377 DNA sequencer.

16S rRNA sequencing and data analysis of unrecognized inserts. A total of 2.522 clones with the correct size insert of approximately 1.500 bases were analyzed—approximately 50 to 100 per subject. In these studies, approximately 500 bases were obtained first to determine the identity or approximate phylogenetic position. Full sequences (about 1,500 bases with five to six additional sequencing primers) (Table 1) were obtained for most of the novel species. For identification of the closest relatives, the sequences of the unrecognized inserts were compared to the 16S rRNA gene sequences of over 4,000 microorganisms in our database and the 16,000 sequences in the Ribosomal Database Project (RDP) (28) and GenBank. Programs for data entry, editing, sequence alignment, secondary structure comparison, similarity matrix generation, and phylogenetic tree construction were written by F. E. Dewhirst (36). The similarity matrices were corrected for multiple base changes at single positions by the method of Jukes and Cantor (20). Similarity matrices were constructed from the aligned sequences by using only those sequence positions for which 90% of the strains had data. Phylogenetic trees were constructed by the neighbor-joining method of Saitou and Nei (40). TREECON, a software package for the Microsoft Windows environment, was used for the construction and drawing of evolutionary trees (47). Two hundred bootstrap trees were generated, and bootstrap confidence levels were determined with the TREECON program.

We are aware of the potential creation of 16S rDNA chimera molecules

3772 PASTER ET AL. J. BACTERIOL.

TABLE 2.	Diversity	of species	and	phylotypes	in	healthy a	and
diseased sites							

Health status (n)	No. of clones analyzed	No. (%) of known species	No. (%) of novel phylotypes	Total no. of species/ phylotypes
Healthy (5)	268	42 (58)	30 (42)	72
Periodontitis (9)	417	46 (51)	45 (49)	91
Refractory periodontitis (11)	1,386	112 (53)	101 (47)	213
ANUG (4)	417	36 (38)	59 (62)	96
HIV (2)	34	6 (55)	5 (45)	11
Total	2,522	132 (38)	215 (62)	347

assembled during the PCR (26). The percentage of chimeric inserts in 16S rRNA libraries ranged from 1 to 15%. Chimeric sequences were identified by using the Chimera Check program in RDP, by treeing analysis, or by base signature analysis. Species identification of chimeras was obtained, but the sequences were not examined for phylogenetic analysis.

Estimation of unseen species. In 1943, Fisher et al. (13) described a model for estimating the number of unseen species in a population. In addition to ecological studies of biological diversity, similar methods have been used to estimate the number of words known but not used by Shakespeare and other authors (12). In this study, the number of unseen species that were missed was calculated with an improved estimator as proposed by Boneh et al. (5). The estimator is based on a continuous time model of parallel independent Poisson processes. The estimator is:

$$\hat{\Psi}(t) = \sum_{k=1}^{k_{\text{max}}} N_k e^{-k} - \sum_{k=1}^{k_{\text{max}}} N_k e^{-k(1+t)}$$

where $\hat{\Psi}(t)$ is the number of new species expected to be observed over time zero to t when the original observations were over time -1 to 0. k is the number of times a species is seen. N_k is the number of species seen k times. Bias correction and other details of the estimator are beyond the scope of the current discussion, and the reader is referred to the publication of Boneh et al. (5).

Nucleotide sequence accession number. The complete 16S rRNA gene sequences of clones representing novel phylotypes defined in this study, sequences of known species not previously reported, and published sequences are available for electronic retrieval from the EMBL, GenBank, and DDBJ nucleotide sequence databases under the accession numbers shown in Fig. 1 through 7.

RESULTS

Partial sequences of about 500 bp were obtained for 2,522 16S rRNA clones in order to identify the predominant bacterial species present in the subgingival plaque of healthy and diseased subjects. Approximately 60% of the 2,522 clones had greater than 99% sequence similarity to one of 132 known species. The remaining 40% of the clones fell into 215 previously unrecognized clusters, termed phylotypes. Full 1,500base sequences were determined for representatives of each novel phylotype. The term "phylotype" is used for clusters of clone sequences that differed from known species by approximately 30 bases (or 2%) and were at least 99% similar to members of their cluster. The diversity of known species and novel phylotypes with respect to periodontal health status is shown in Table 2. Approximately 40 to 60% of clones analyzed in each category of healthy or diseased subjects were novel phylotypes. Overall, we detected 347 species or phylotypes in subgingival plaque.

The intent of using different PCR primer sets was to obtain the widest spectrum of phylogenetic groups. In previous studies, results obtained with the *Spirochaetes*-selective primers indicate that 85% of the clones have spirochetal inserts (6). Our results using the *Bacteroidetes*-selective primers were not as selective but yielded a wide range of bacterial types—including novel taxa at the phylum level. In designing the "universal" bacterial primer set, a "G" was added to the 5' end of the reverse primer (E94) to enhance cloning of amplicons with the TOPO-TA cloning kit. Despite presenting a mismatch at this position to most 16S rRNA sequences in the RDP, a continuous stretch of 20 matching bases were available for optimal PCR. Consequently, a wide range of phylogenetic types was obtained by using this "universal" primer set.

Estimation of unseen species. By using the estimator of Boneh et al. (5) and applying the suggested bias reduction procedure, we calculated that there are 68 unseen species in addition to the 347 found, for a total estimate of 415 subgingival species. Table 3 shows the number of additional species that one would expect to identify by examining various numbers of additional clones. As shown in Table 3, fewer new species were predicted to be found for each thousand additional clones analyzed. The model suggests that by examining 10,000 additional clones, we would find all but one of the estimated 415 species.

As shown in Fig. 1, the oral bacteria identified in this study fell into nine bacterial groups or phyla as follows: Obsidian Pool OP11 (19), TM7 (38); Deferribacteres, Spirochaetes, Fusobacteria, Actinobacteria, Firmicutes, Proteobacteria, and Bacteroidetes (the latter seven phyla are listed in the current taxonomic outline for Bergey's Manual of Systematic Bacteriology, 2nd ed., vol. 2, at http://www.cme.msu.edu/bergeys/). The distribution of known species and novel phylotypes within each of these phyla is shown in Table 4. In Fig. 1 through 7, the phylogenetic diversity within each phylum is shown and is discussed in detail below. The information presented includes bacterial species or phylotype; strain or clone identification, sequence accession number, total number of retrieved clones, and number of subjects in each health status category in which each species was identified (color- and shape-coordinated symbols). Prevalent species or phylotypes are defined as those identified from four or more subjects and are noted in bold and underlined in each dendrogram.

Obsidian Pool OB11. As shown in Fig. 1, oral clone X112 represents the sole oral member of this phylum. Although only one phylotype was identified, it was found in multiple subjects. Distant relatives include phylotypes identified from sludge and

TABLE 3. Number of unseen species

No. of additional clones analyzed	No. of expected new species	Increment in new species ^a	
1,000	26.86	26.86	
2,000	42.36	15.50	
3,000	51.72	9.36	
4,000	57.55	5.83	
5,000	61.27	3.72	
6,000	63.68	2.41	
7,000	65.26	1.58	
8,000	66.31	1.04	
9,000	67.00	0.69	
10,000	67.46	0.46	
	68.41	0.95	

^a Number of species expected minus that in the row above.

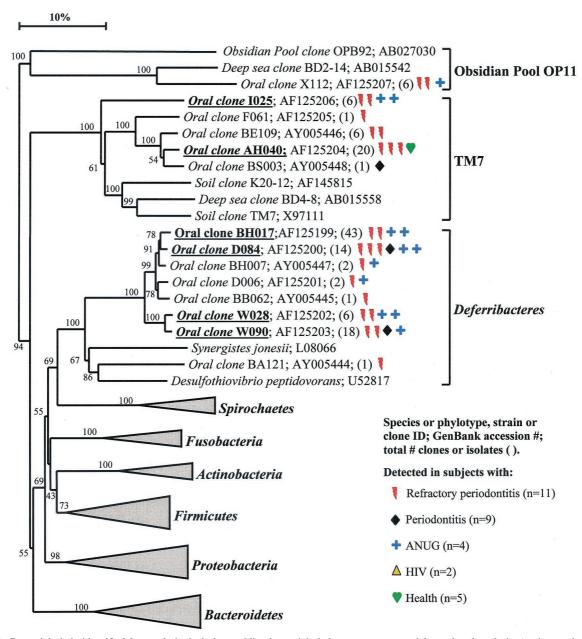


FIG. 1. Bacterial phyla identified from subgingival plaque. Nine bacterial phyla were represented from clonal analysis. As shown, the first two phyla, Obsidian Pool OP11 and TM7, have no cultivable representatives. Oral members of the phylum Deferribacteres form a cluster consisting of only not-yet-cultivated phylotypes. The information presented in Fig. 1 to 7 includes bacterial species or phylotype, strain or clone identification, sequence accession number, total number of retrieved clones, and number of subjects in each health status category in which each species was identified (as indicated by color-coordinated symbols). Novel phylotypes are defined as those taxa that are <98.5 to 99% similar in sequence comparisons to the phylotype's closest relative. Prevalent species are defined as those phylotypes or species identified in four or more subjects and are noted in boldface and underlined in each dendrogram. Two hundred bootstrap trees were generated, and bootstrap confidence levels as percentages (only values over 40%) are shown at tree nodes.

deep-sea sediments. There are no cultivable representatives of this phylum.

Vol. 183, 2001

TM7. Five oral phylotypes of the TM7 phylum were identified as shown in Fig. 1. Distant relatives of this phylum have been identified in 16S rRNA clones derived from DNA isolated from soil and deep-sea sediments. The oral phylotypes were initially identified with Bacteroidetes-selective primers, but later identified with all-bacterial-selective primers, which indicates that these phylotypes are relatively common in subgingival plaque. There are no cultivable representatives of this

Deferribacteres. In the Deferribacteres phylum, the oral species formed a coherent cluster of seven separate phylotypes, four of which were prevalent in diseased subjects, especially in the ANUG and refractory periodontitis subjects. Consequently, the latter phylotypes may be good candidates as putative pathogens. Members of this oral cluster are distantly related to Synergistes jonesii, a cultivable, bovine ruminal bac3774 PASTER ET AL. J. BACTERIOL.

TABLE	4. Known and novel species or phylotypes
	identified from clonal analysis

	Known spo	ecies	Novel phylotypes		
Phylum	No. of clones (% of total)	No. of species	No. of clones (% of total)	No. of phylotypes	
Obsidian Pool OB11	0	0	6 (100)	1	
TM7	0	0	34 (100)	5	
Deferribacteres	0	0	86 (100)	8	
Spirochaetes	199 (37)	9	338 (63)	49	
Fusobacteria	257 (73)	7	96 (27)	12	
Actinobacteria	229 (83)	14	46 (17)	18	
Firmicutes	454 (69)	53	205 (31)	62	
Proteobacteria	244 (72)	28	94 (28)	23	
Bacteroidetes	148 (63)	21	86 (37)	37	
Total	1,531 (60.7)	132	991 (39.3)	215	

terium able to degrade the pyridinediol toxin in the plant *Leucaena leucocephala* (27, 30). One oral clone, BA121, fell within the *S. jonesii* cluster rather than the oral cluster.

Spirochaetes. The human periodontal pocket harbors a highly diverse and numerous spirochetal community (6, 9). All species identified thus far fall in the genus *Treponema*, and presently there are about 60 oral treponemal species or phylotypes (6) (Fig. 2). As shown in Fig. 2, several prevalent species or phylotypes were identified. The predominant cultivable known *Treponema* species included *T. medium* (but not the closely related "*T. vincentii*"), *T. denticola* (which was one of the most commonly identified species in both disease and health), *T. maltophilum*, and two subspecies of *T. socranskii*. There were at least four predominant novel phylotypes, one of which, *Treponema* sp. 1:G:T21, was found only in diseased sites. The latter phylotype appears to be a good candidate as a putative pathogen.

It should be noted that the treponemal clones had been specifically sought by constructing 16S rRNA libraries with spirochete-selective primers in the initial PCR amplification. Microscopic examination of plaque samples from diseased sites, especially those from ANUG subjects, revealed that 20 to 50% of the members of the bacterial community were spirochetes (6). Subsequently, spirochetes were also detected in 16S rRNA libraries when all-bacterial primers were used, albeit at a lower frequency.

Fusobacteria. As shown in Fig. 3, the members of the oral Fusobacteria identified include species of the genera Fusobacterium and Leptotrichia. Most noteworthy was that Fusobacterium naviforme (earlier referred to as Fusobacterium nucleatum subsp. vincentii) was often identified, especially in periodontitis subjects (eight of nine subjects). In six of these eight periodontitis subjects, Fusobacterium naviforme represented from 50 to 80% of the clones. Other species and phylotypes of Fusobacterium and Leptotrichia buccalis from healthy and diseased subjects were also commonly identified. Fusobacterium animalis was detected only in diseased subjects but not in healthy subjects.

Actinobacteria. The Actinobacteria phylum was previously referred to as the high-G+C-content, gram-positive phylum. As shown in Fig. 3, oral genera of the phylum Actinobacteria that were detected include Actinomyces, Atopobium, Bifidobacterium, Corynebacterium, Propionibacterium, and Rothia. Within

these genera, Actinomyces naeslundii II, Corynebacterium matruchotii, and Rothia dentocariosa were commonly detected in health and disease. Corynebacterium matruchotii appeared to be more associated with "healthy" subgingival plaque. Two species of Atopobium were prevalent in diseased but not healthy subjects.

Firmicutes. Most of the members of the Firmicutes phylum were originally classified as belonging to the low-G+C-content, gram-positive phylum. The Firmicutes are presently divided into three classes—namely the proposed "Bacilli," the Mollicutes, and the proposed "Clostridia." The phylogenetic position of oral members of the "Bacilli" and Mollicutes is shown in Fig. 4. In the class "Bacilli," the predominant species included the streptococci Streptococcus oralis or S. mitis, S. mitis biovar 2, S. sanguis, S. intermedius, S. constellatus, and S. anginosus; Abiotrophia adiacens; and two species of Gemella. Since it was difficult to differentiate S. oralis and S. mitis by 16S rRNA comparisons, they were considered as one species in this study. Of the predominant species, S. constellatus and Gemella haemolysans were most associated with diseased sites, as they were not detected in subgingival plaque from any of the healthy subjects. Although most of the other predominant species were detected in both subjects with disease and those with health, many of the Streptococcus spp. were commonly identified from the healthy subjects.

Not many members of the class *Mollicutes* were detected: two species of *Mycoplasma*; *Solobacterium moorei*, a microorganism isolated from feces that is related to species of *Erysipelothrix*; and a misclassified *Lactobacillus* species, [*L.*] *catenaforme*. Infections by species of *Erysipelothrix* have been linked to endocarditis (16). It was encouraging to detect the purported hard-to-lyse mycoplasmas (23), since the results demonstrated that our procedures do indeed recover a broad spectrum of bacterial species.

Members of the class "Clostridia" represented one of the largest groups detected, as shown in Fig. 5. Oral members included species within the genera Catonella, Dialister, Eubacterium, Megasphaera, Peptostreptococcus, Selenomonas, and Veillonella. There were several good candidates as additional putative pathogens, i.e., those species or phylotypes that were commonly detected in disease, but not (or rarely) in health—namely Eubacterium saphenum, clone PUS9.170, Filifactor alocis (previously Fusobacterium alocis), Catonella morbi, Megasphaera sp. oral clone BB166, Dialister sp. strain GBA27, and Selenomonas sputigena.

Catonella morbi has been associated with periodontitis (31), clone PUS9.170 and Filifactor alocis have been associated with dentoalveolar abscesses (11, 50), and Selenomonas sputigena has been associated with periodontitis and, in rare cases, has been involved in fatal septicemias (51). Eubacterium saphenum and species of Megasphaera have not been considered as putative periodontal pathogens (44).

Oral isolate GBA27 is a slow-growing *Dialister* species that was commonly detected in clones from all disease categories and from one healthy subject (Fig. 6). Since strain GBA27 cells form only pinpoint colonies on agar medium (A. Tanner, The Forsyth Institute, personal communication), it is likely that this species has eluded detection in predominant cultivable studies over the years. The closely related *Dialister pneumosintes* has

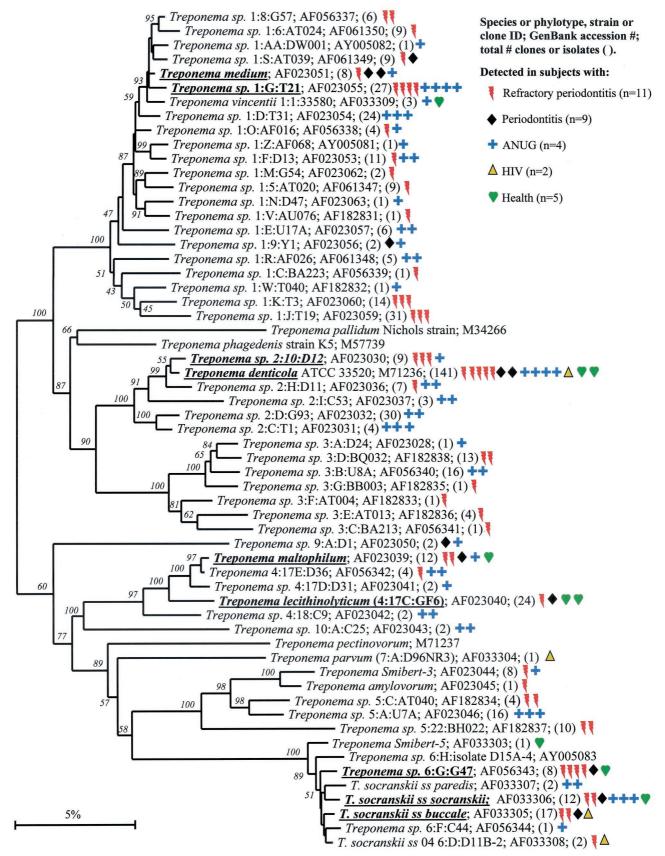


FIG. 2. Phylogenetic tree of oral treponemes identified from clone libraries. All species identified thus far fall in the genus *Treponema*, and presently there are about 60 oral treponemal species or phylotypes (6). The marker bar represents a 10% difference in nucleotide sequences.

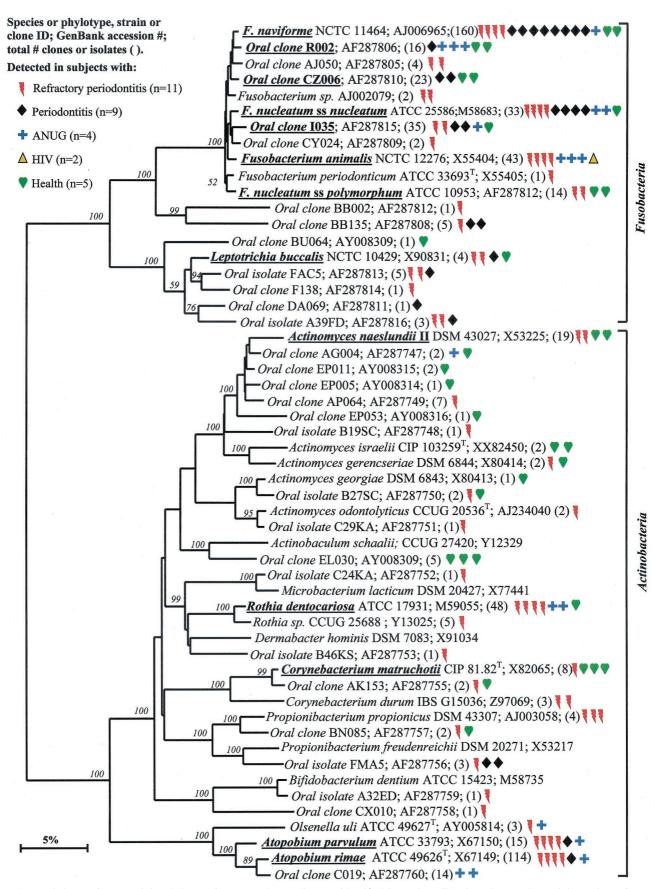


FIG. 3. Phylogenetic trees of the phyla Fusobacteria and Actinobacteria identified from clone libraries. The members of the oral Fusobacteria identified include species of the genera Fusobacterium and Leptotrichia. The marker bar represents a 5% difference in nucleotide sequences.

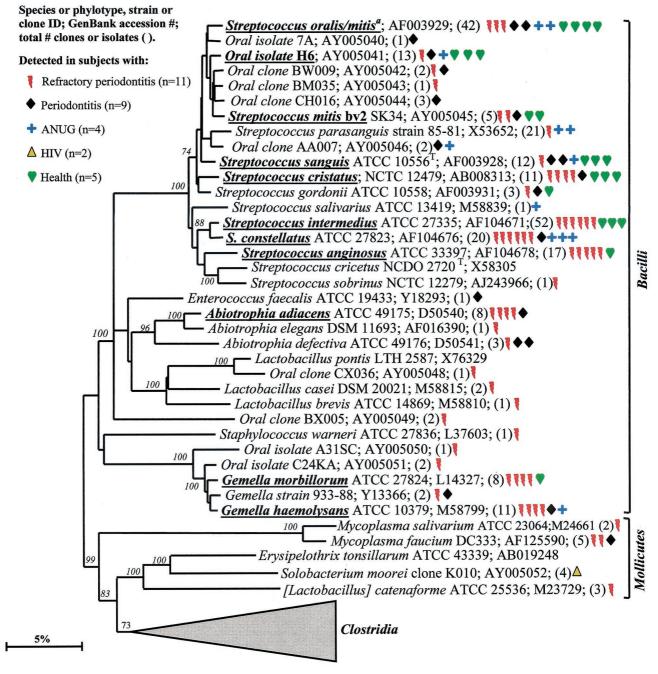


FIG. 4. Phylogenetic tree of the phylum *Firmicutes* identified from clone libraries. The phylum *Firmicutes* is presently divided into three classes: "Bacilli," Mollicutes, and "Clostridia." This phylum represented the largest group detected; 115 species or novel phylotypes were identified (Table 4). Streptococcus oralis and S. mitis could not be differentiated based on sequence comparisons. Consequently, the two species are grouped together. The marker bar represents a 5% difference in nucleotide sequences.

been associated with gingivitis (31) and human bite wounds (15).

Proteobacteria. Although there were 28 known species and 23 novel phylotypes of *Proteobacteria* detected (Table 4), most were only sporadically identified, as shown in Fig. 6. Those species or phylotypes detected that were most associated with disease include *Haemophilus parainfluenzae*, oral clone R004, and *Campylobacter rectus* (Fig. 6). *Haemophilus parainfluenzae*, mostly detected in refractory periodontitis sites, has been iso-

lated from the sputa of patients with chronic obstructive lung disease (29) and may be involved in endocarditis (3). Oral clone R004, detected in all disease categories (except HIV periodontitis), likely represents a novel sulfate-reducing bacterial species belonging to the genus *Desulfobulbus*. Sulfate-reducing bacteria, which have been reported in the periodontal pocket (48), are routinely isolated from the feces of patients with ulcerative colitis. Sulfide, an end product of sulfate reduction, may be involved in the pathogenesis of the disease (7).

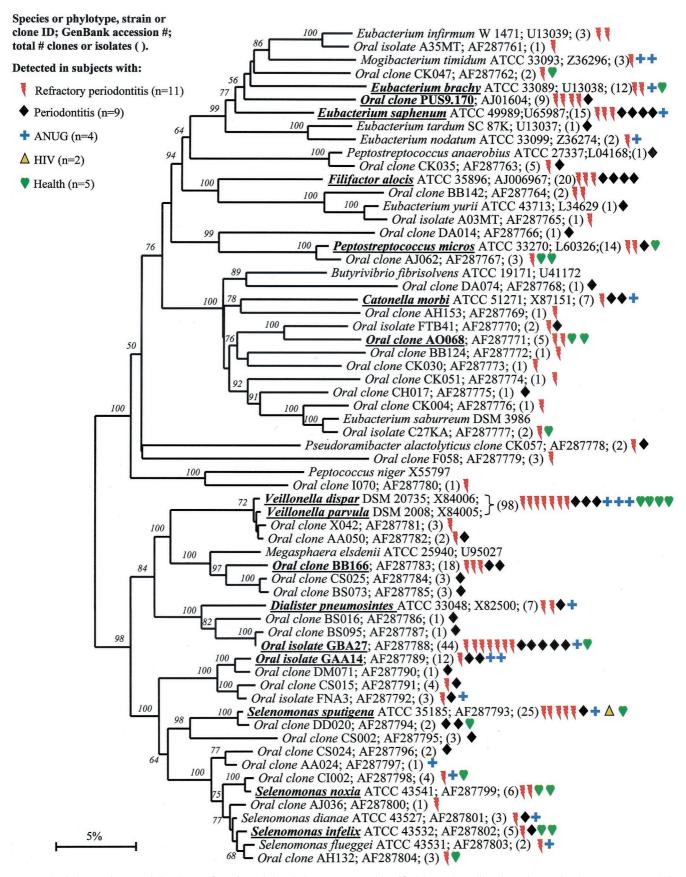


FIG. 5. Phylogenetic tree of the class "Clostridia" of the phylum Firmicutes identified from clone libraries. The marker bar represents a 5% difference in nucleotide sequences.

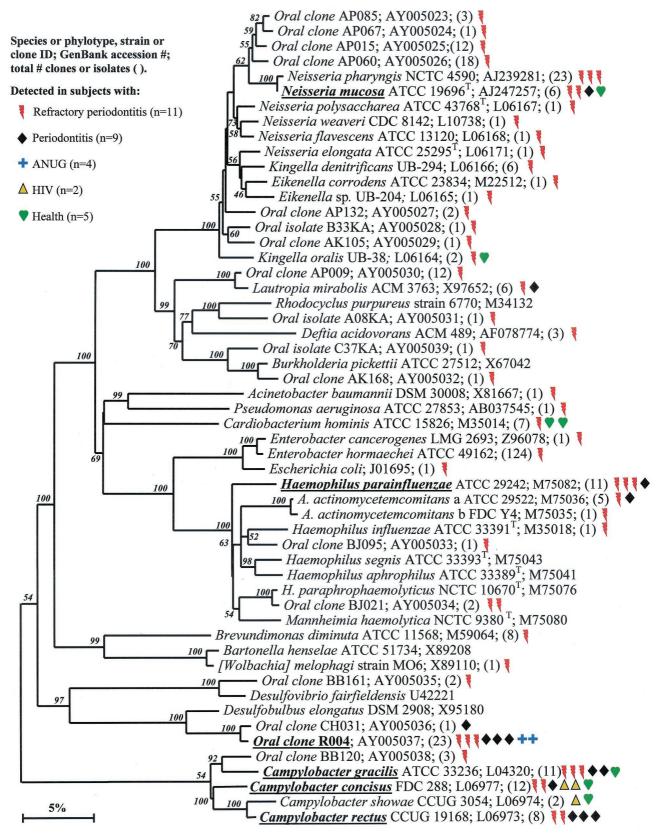


FIG. 6. Phylogenetic tree of the phylum *Proteobacteria* identified from clone libraries. The marker bar represents a 5% difference in nucleotide sequences.

3780 PASTER ET AL. J. BACTERIOL.

Campylobacter rectus has long been associated with periodontal disease (44). Species or phylotypes commonly detected in both diseased and healthy subjects include *Neisseria mucosa* and two additional species of *Campylobacter*, *C. gracilis* and *C. concisus*.

Bacteroidetes. There were 21 known species and 37 novel phylotypes of Bacteroidetes (previously referred to as the Capnocytophaga/Flavobacterium/Bacteroides phylum) detected in 16S rRNA libraries (Table 4). As anticipated, P. gingivalis and B. forsythus, two species typically associated with periodontitis, were detected, although not as frequently as some other species or phylotypes (Fig. 7). Other species or phylotypes that were commonly detected include Prevotella denticola, Prevotella oris, Prevotella tannerae, Porphyromonas endodontalis, oral clone AU126, Capnocytophaga ochracea, and Capnocytophaga gingivalis (Fig. 7). Porphyromonas endodontalis has been associated with endodontal infections (49), but it was often detected in diseased subjects and not in any of the healthy subjects (Fig. 7). Prevotella oris has been associated with dentoalveolar abscesses (11). The remaining cultivable species of this phylum have not been typically considered periodontal pathogens (44).

DISCUSSION

Human subgingival plaque harbors a highly diverse bacterial community, as demonstrated by our detection of 347 species or phylotypes by clonal analysis. Using the estimator of Boneh et al. (5), we predicted that approximately 68 additional unseen species were present and could be detected by examining 10,000 additional clones. This estimate is a rough approximation because the clones came from libraries constructed from subjects with different health and disease categories, violating the assumption of sampling with replacement from a homogeneous population. However, the model suggests that we have already identified 85% of the species and makes testable predictions as to how many unseen species should be found by examining additional clones (Table 3). Examination of 179 additional clones, not included in this data set, from subgingival plague from sites of necrotizing ulcerative periodontitis (29) yielded six novel species where the model would have predicted five. Thus, the model and predictions appear quite accurate. When organisms found in other studies (11, 21, 41, 45, 50) and on cheek, tongue, and other oral tissues are added to our estimate of 415 subgingival species, the best estimate of the total species diversity in the oral cavity is between 500 and 600 species, as has been previously proposed. A list of known human oral species and phylotypes can be obtained from the corresponding author.

Of the 215 phylotypes identified in this study, 33 were cultivable strains that have not yet been characterized. The remaining 182 were represented only by clones. It is likely that the vast majority of these phylotypes cannot be cultivated with standard anaerobic media and techniques. It is also likely that many of these phylotypes are "hidden" within culture collections, since the strains could not be differentiated from closely related known species. In this regard, we are currently examining isolates of over 100 unnamed taxa from the oral anaerobic collection of W. E. C. Moore and L. V. Moore (Virginia Polytechnic Institute, Blacksburg, Va.). It will be interesting to see how many of Moore's groups represent cultivable examples

of groups previously represented only by clone phylotypes. *Treponema amylovorum* and *Treponema lecithinolyticum* were identified first by clone data before the cultivable strains were characterized (6; this study). The phylogenetic position of notyet-cultivated novel phylotypes, especially if they are closely related to cultivable species, may provide insight into their cultivation. For example, physiological properties or antibiotic resistance can be inferred to some degree. DNA probes specific for the phylotype can be used to monitor enrichment efforts and to positively identify an isolate.

The definition of a species is controversial, particularly when only molecular sequence data exist (14, 46). Therefore, we have used the term "phylotype" in place of species for referring to novel clusters of clone sequences. In most cases, a 2% difference in 16S rRNA sequences does indicate separate species status, but there are exceptions. Formal naming of a species also requires a full description of the phenotypic characteristics of an organism. It is probable that the majority of the phylotypes identified in this report will eventually be validated as species. In the meantime, DNA probes can be designed to identify phylotypes and to assess their roles in disease or health. If a phylotype proves to be associated with disease, then efforts can be made to isolate and characterize the new species.

In general, the first 500 bases of the 16S rRNA gene are the most informative for identifying an organism, because there are several variable regions. However, for constructing phylogenetic trees, it is essential that nearly complete sequences of 1,500 bases be used. This is a deficiency of some previous studies. We have also obtained full sequences so that the data are most useful to databases such as GenBank or the RDP (28).

A total of 72 species or novel phylotypes were identified in the analysis of 268 clones from healthy subjects (Table 2). Based on the overall total of 347, there are 275 species or novel phylotypes that were identified in diseased subjects. Part of the explanation for the discrepancy between the number of species found in healthy subjects versus that found in diseased subjects is that only 10.6% of the clones analyzed were from healthy subjects. However, even with this bias, a number of species and phylotypes were found only in multiple (≥4) diseased subjects and not in any healthy subjects. These often-identified diseaseassociated species and phylotypes are shown in Table 5 and are clearly candidates for further study. Eighteen of the 29 species are named species, including Porphyromonas gingivalis and B. forsythus. One represents a cultivable, but unnamed species. Ten of the phylotypes are represented only by clones, and thus may represent currently not-yet-cultivated or unrecognized potential pathogens. Additional putative pathogens are likely, since there were several species or phylotypes that were often identified from diseased sites, but only rarely from health: e.g., Dialister sp. strain GBA27 (Fig. 5), other species or phylotypes of Fusobacterium (Fig. 3), or oral clone AU126 (Fig. 7). There were also several species that were more associated with health, namely Actinobaculum sp. clone EL030 and Corynebacterium matruchotii (Fig. 3). However, to rigorously assess the association of specific species or phylotypes with certain periodontal diseases or periodontal health, it will be necessary to analyze large numbers of clinical samples for the levels of all oral bacteria in well-controlled clinical studies. The definition of five subgingival plaque bacterial complexes by Socransky et

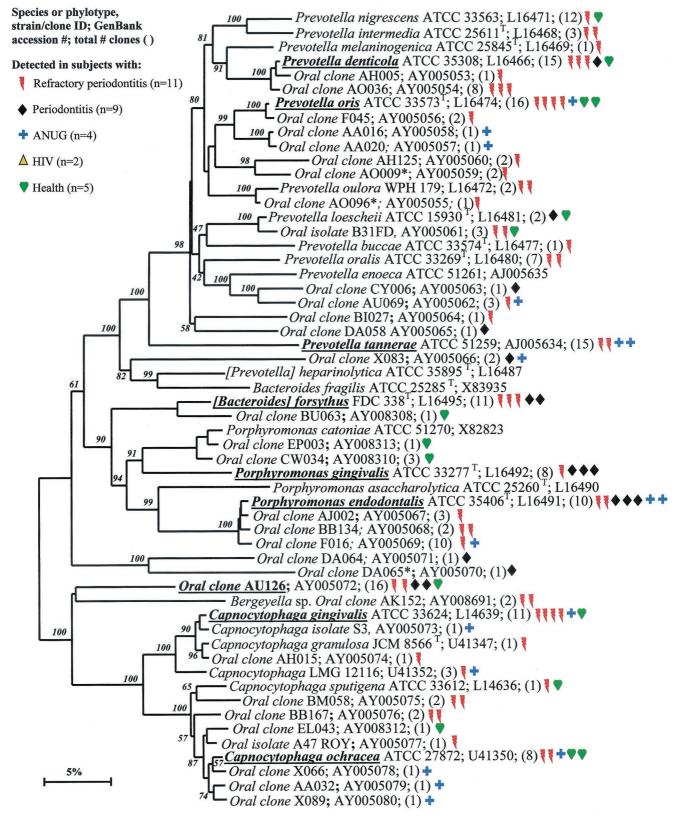


FIG. 7. Phylogenetic tree of the phylum *Bacteroidetes* identified from clone libraries. Oral species identified generally fall within the genera *Porphyromonas*, *Prevotella*, and *Capnocytophaga*. An asterisk denotes sequence of only about 1,200 bp. The marker bar represents a 5% difference in nucleotide sequences.

TABLE 5. Putative pathogens in subgingival plaque^a

	of clones dentified
TM7	
Clone I025	6
Deferribacteres	
Clone BH017	43
Clone D084	20
Clone W028	6
Clone W090	18
Spirochaetes	
Clone Treponema sp. 1:G:T21	27
Clone <i>Treponema</i> sp. 2:10:D12	9
Treponema socranskii sub sp. buccale	17
Fusobacteria	
Fusobacterium animalis	43
Actinobacteria	
Atopobium parvulum	15
Atopobium rimae	114
Firmicutes	
Clone Eubacterium sp. PUS9.170	9
Clone Megasphaera sp. BB166	18
Abiotrophia adiacens	8
Catonella morbi	
Dialister pneumonsintes	7
Eubacterium saphenum	15
Filifactor alocis	20
Gemella haemolysans	11
Selenomonas sp. strain GAA14	12
Streptococcus constellatus	
Proteobacteria	
Clone Desulfobulbus sp. R004	23
Campylobacter gracilis	
Campylobacter rectus	
Haemophilus parainfluenzae	
Bacteroidetes	
Bacteroides forsythus	11
Prevotella tannerae	15
Porphyromonas endodontalis	10
Porphyromonas gingivalis	

^a Detected in at least four diseased subjects, but not in healthy subjects. Data were derived from Fig. 1 through 7.

al. (44) was based on analysis of the microbial community of over 13,000 plaque samples from 185 subjects by using DNA probes in checkerboard hybridization assays. We are currently developing DNA probes for approximately all 500 known species and novel phylotypes for use in large clinical studies. These DNA probes are being developed for use in the checkerboard hybridization assay (35) and DNA microarray formats. Based upon knowledge of the full bacterial diversity, and using molecular techniques to identify these bacteria, future studies will be able to associate a number of additional oral bacteria with various oral and systemic diseases.

ACKNOWLEDGMENTS

This study was supported by NIH grants DE-11443 and DE-10374 from the National Institute of Dental and Craniofacial Research.

REFERENCES

- Amman, R., W. Ludwig, and K.-H. Schleifer. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. 59:143–169.
- Beck, J., R. Garcia, G. Heiss, P. S. Vokonas, and S. Offenbacher. 1996. Periodontal disease and cardiovascular disease. J. Periodontol. 67:1123–1137.
- 3. Berbari, E. F., F. R. Cockerill III, and J. M. Steckelberg. 1997. Infective

- endocarditis due to unusual or fastidious microorganisms. Mayo Clin. Proc. 72:532–542.
- Boches, S. K., P. M. Mitchell, J. L. Galvin, W. J. Loesche, C. E. Kazor, F. E. Dewhirst, and B. J. Paster. 2000. Cultivable and uncultivable bacteria on the healthy tongue dorsum. J. Dent. Res. 79:396.
- Boneh, S., A. Boneh, and R. J. Caron. 1998. Estimating the prediction function and the number of unseen species in sampling with replacement. J. Am. Statist. Assoc. 93:372–379.
- Choi, B. K., B. J. Paster, F. E. Dewhirst, and U. B. Göbel. 1994. Diversity of cultivable and uncultivable oral spirochetes from a patient with severe destructive periodontitis. Infect. Immun. 62:1889–1895.
- Christl, S. U., H. D. Eisner, G. Dusel, H. Kasper, and W. Scheppach. 1996. Antagonistic effects of sulfide and butyrate on proliferation of colonic mucosa: a potential role for these agents in the pathogenesis of ulcerative colitis. Dig. Dis. Sci. 41:2477–2481.
- Coleman, B. E., N. Tzellas, B. J. Paster, and F. E. Dewhirst. 1996. Identification of 16S rRNA clones from a subject with ANUG as *Atopobium* species.
 J. Dent. Res. 75:207.
- Dewhirst, F. E., M. A. Tamer, R. E. Ericson, C. N. Lau, V. A. Levanos, S. K. Boches, J. L. Galvin, and B. J. Paster. 2000. The diversity of periodontal spirochetes by 16S rRNA analysis. Oral Microbiol. Immunol. 15:196–202.
- Dodman, T., J. Robson, and D. Pincus. 2000. Kingella kingae infections in children. J. Paediatr. Child Health 36:87–90.
- Dymock, D., A. J. Weightman, C. Scully, and W. G. Wade. 1996. Molecular analysis of microflora associated with dentoalveolar abscesses. J. Clin. Microbiol. 34:537–542.
- 12. **Efron, B., and R. Thisted.** 1976. Estimating the number of unseen species. How many words did Shakespeare know? Biometrika **63**:435–447.
- Fisher, R. A., A. S. Corbet, and C. B. Williams. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. J. Anim. Ecol. 12:42–58.
- Fox, G. E., J. D. Wisotzkey, and P. Jurtshuk, Jr. 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. Syst. Bacteriol. 42:166–170.
- Goldstein, E. J., D. M. Citron, and S. M. Finegold. 1984. Role of anaerobic bacteria in bite-wound infections. Rev. Infect. Dis. 6:S177–S183.
- Gorby, G. L., and J. E. Peacock, Jr. 1988. Erysipelothrix rhusiopathiae endocarditis: microbiologic, epidemiologic, and clinical features of an occupational disease. Rev. Infect. Dis. 10:317–325.
- Harper-Owen, R., D. Dymock, B. Booth, A. J. Weightman, and W. G. Wade. 1999. Detection of unculturable bacteria in periodontal health and disease by PCR. J. Clin. Microbiol. 37:1469–1473.
- Hugenholtz, P., and N. R. Pace. 1996. Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. Trends Biotechnol. 14:190–197.
- Hugenholtz, P., C. Pitulle, K. L. Hershberger, and N. R. Pace. 1998. Novel division level bacterial diversity in a Yellowstone hot spring. J. Bacteriol. 180:366–376.
- Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules, p. 21–132. *In* H. N. Munro (ed.), Mammalian protein metabolism, vol. 3. Academic Press, Inc., New York, N.Y.
- Kroes, I., P. W. Lepp, and D. A. Relman. 1999. Bacterial diversity within the human subgingival crevice. Proc. Natl. Acad. Sci. USA 96:14547–14552.
- Lau, C. N., B. J. Paster, V. Levanos, R. L. Ericson, and F. E. Dewhirst. 1998. Cultivable and uncultivable predominant bacterial species in refractory periodontitis. J. Dent. Res. 77:139.
- Leng, Z., G. E. Kenny, and M. C. Roberts. 1994. Evaluation of the detection limits of PCR for identification of *Mycoplasma pneumoniae* in clinical samples. Mol. Cell. Probes 8:125–130.
- Levanos, V. A., J. L. Galvin, S. K. Boches, F. E. Dewhirst, and B. J. Paster. 2000. Predominant cultivable and uncultivable bacterial species in adult periodontitis. J. Dent. Res. 79:234.
- Levin, I. M., C. N. Lau, S. S. Socransky, A. D. Haffajee, L. Martin, J. L. Galvin, S. K. Boches, B. J. Paster, and F. E. Dewhirst. 1999. Cultivable and uncultivable species on or in gingival epithelial cells. J. Dent. Res. 78:453.
- Liesack, W., H. Weyland, and E. Stackebrandt. 1991. Potential risk of gene amplification by PCR as determined by 16S rDNA analysis of a mixedculture of strict barophilic bacteria. Microbiol. Ecol. 21:191–198.
- Ludwig, W., G. Wallner, A. Tesch, and F. Klink. 1991. A novel eubacterial phylum: comparative nucleotide sequence analysis of a tuf-gene of *Flexistipes sinusarabici*. FEMS Microbiol. Lett. 62:139–143.
- Maidak, B. L., J. R. Cole, T. G. Lilburn, C. T. Parker, Jr., P. R. Saxman, J. M. Stredwick, G. M. Garrity, B. Li, G. J. Olsen, S. Pramanik, T. M. Schmidt, and J. M. Tiedje. 2000. The RDP (Ribosomal Database Project) continues. Nucleic Acids Res. 28:173–174.
- Mitchell, J. L., and S. L. Hill. 2000. Immune response to *Haemophilus parainfluenzae* in patients with chronic obstructive lung disease. Clin. Diagn. Lab. Immunol. 7:25–30.
- McSweeny, C. S., M. J. Allison, and R. I. Mackie. 1993. Amino acid utilization by the ruminal bacterium *Synergistes jonesii* strain 78–1. Arch. Microbiol. 159:131–135.
- 31. Moore, L. V., and W. E. C. Moore. 1994. Oribaculum catoniae gen. nov., sp.

- nov.; Catonella morbi gen. nov., sp. nov.; Hallella seregens gen. nov., sp. nov.; Johnsonella ignava gen. nov., sp. nov.; and Dialister pneumosintes gen. nov., comb. nov., nom. rev., anaerobic gram-negative bacilli from the human gingival crevice. Int. J. Syst. Bacteriol. 44:187–192.
- Moore, W. E. C., and L. V. H Moore. 1994. The bacteria of periodontal diseases. Periodontol. 2000 5:66–77.
- Offenbacher, S., H. L. Jared, P. G. O'Reilly, S. R. Wells, G. E. Salvi, H. P. Lawrence, S. S. Socransky, and J. D. Beck. 1998. Potential pathogenic mechanisms of periodontitis associated pregnancy complications. Ann. Periodontol. 3:233–250.
- Pace, N. R., D. A. Stahl, D. J. Lane, and G. J. Olsen. 1986. The analysis of natural microbial populations by ribosomal RNA sequences. Adv. Microb. Ecol. 9:1–55.
- Paster, B. J., I. M. Bartoszyk, and F. E. Dewhirst. 1998. Identification of oral streptococci using PCR-based, reverse-capture, checkerboard hybridization. Methods Cell Sci. 20:223–231.
- Paster, B. J., and F. E. Dewhirst. 1988. Phylogeny of campylobacters, wolinellas, *Bacteroides gracilis*, and *Bacteroides ureolyticus* by 16S rRNA sequencing. Int. J. Syst. Bacteriol. 38:56–62.
- Paster, B. J., F. E. Dewhirst, W. G. Weisburg, L. A. Tordoff, G. J. Fraser, R. B. Hespell, T. B. Stanton, L. Zablen, L. Mandelco, and C. R. Woese. 1991. Phylogenetic analysis of the spirochetes. J. Bacteriol. 173:6101–6109.
- Rheims, H., F. A. Rainey, and E. Stackebrandt. 1996. A molecular approach to search for diversity among bacteria in the environment. J. Ind. Microbiol. 17:159–169.
- Russell, M. K., T. Alpagot, S. K. Boches, J. L. Galvin, F. E. Dewhirst, and B. J. Paster. 2001. Bacterial species and phylotypes in necrotizing ulcerative periodontitis. J. Dent. Res. 80:167.
- 40. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.
- Sakamoto, M., M. Umeda, I. Ishikawa, and Y. Benno. 2000. Comparison of the oral bacterial flora in saliva from a healthy subject and two periodontitis patients by sequence analysis of 16S rDNA libraries. Microbiol. Immunol. 44:643–652.
- Scannapieco, F. A. 1999. Role of oral bacteria in respiratory infection. J. Periodontol. 70:793–802.

- Socransky, S. S., and A. D. Haffajee. 1994. Evidence of bacterial etiology: a historical perspective. Periodontol. 2000 5:7–25.
- Socransky, S. S., A. D. Haffajee, M. A. Cugini, C. Smith, and R. L. Kent, Jr. 1998. Microbial complexes in subgingival plaque. J. Clin. Periodontol. 25: 134–144.
- Spratt, D. A., A. J. Weightman, and W. G. Wade. 1999. Diversity of oral asaccharolytic *Eubacterium* species in periodontitis: identification of novel phylotypes representing uncultivated taxa. Oral Microbiol. Immunol. 14:56– 50
- Stackebrandt, E., and B. M. Goebel. 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44:846–849.
- Van de Peer, Y., and R. De Wachter. 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput. Appl. Biosci. 10:569–570.
- van der Hoeven, J. S., C. W. van den Kieboom, and M. J. Schaeken. 1995. Sulfate-reducing bacteria in the periodontal pocket. Oral Microbiol. Immunol. 10:288–290.
- van Winkelhoff, A. J., T. J. van Steenbergen, and J. de Graaff. 1992. Porphyromonas (Bacteroides) endodontalis: its role in endodontal infections. J. Endod. 18:431–434.
- Wade, W. G., D. A. Spratt, D. Dymock, and A. J. Weightman. 1997. Molecular detection of novel anaerobic species in dentoalveolar abscesses. Clin. Infect. Dis. 25:S235–S236.
- Westh, H., J. J. Christensen, J. Blom, and W. Frederiksen. 1991. Fatal septicaemia with Selenomonas sputigena and Acinetobacter calcoaceticus. A case report. APMIS 99:75–77.
- Wilson, M. J., A. J. Weightman, and W. G. Wade. 1997. Applications of molecular ecology in the characterisation of uncultured microorganisms associated with human disease. Rev. Med. Microbiol. 8:91–101.
- 53. Wu, T., M. Trevisan, R. J. Genco, J. P. Dorn, K. L. Falkner, and C. T. Sempos. 2000. Periodontal disease and risk of cerebrovascular disease: the first national health and nutrition examination survey and its follow-up study. Arch Intern Med 160:2749–2755