JOURNAL OF BACTERIOLOGY, Oct. 1985, p. 230–236
0021-9193/85/100230-07$02.00/0

Vol. 164, No. 1

Natural Relationship between Bacteroides and Flavobacteria

W. G. WEISBURG, Y. OYAIZU, H. OYAIZU, AND C. R. WOESE*

Department of Genetics and Development, University of Illinois, Urbana, Illinois 61801

Received 21 May 1985/Accepted 27 July 1985

Comparisons among 16S rRNA sequences from various eubacteria reveal a natural relationship between the bacteroides (represented by the *Bacteroides fragilis* sequence) and a phylogenetic unit that comprises the flavobacteria, cytophagae, flexibacteria, and others (represented by the *Flavobacterium heparinum* sequence). Although the relationship is not a close one, it is, nevertheless, specific. rRNAs from these two organisms are not only closer to one another in overall sequence than they are to outgroup species (such as *Bacillus subtilis*, *Escherichia coli*, *Desulfovibrio desulfuricans*, and *Agrobacterium tumefaciens*), but they show common idiosyncrasies (i.e., derived characteristics) in both rRNA sequences and higher-order structures.

Of all biological disciplines, microbiology seems the least affected by evolutionary considerations. A phylogenetic tree is an essential starting point for any serious consideration of microbiological evolution, and until quite recently, such a tree could not be determined for the bacteria. Bacterial morphologies and biochemistries are too simple to provide an adequate basis for the establishment of a phylogenetically valid system of bacterial classification.

All this is changing, however. The increasing ease with which nucleic acid sequences can be determined is automatically thrusting evolutionary considerations upon the field. In 1975, no microbial phylogeny existed to speak of. By 1980, partial sequencing methods had produced the outlines of such a phylogeny, and the monolithic world of "procaryotes" suddenly split into two distinct kingdoms, the eubacteria and the archaebacteria (5, 22), each no more related to the other than to the eucaryotes. With present technology, detailed phylogenies of the bacteria are readily determined, and in the near future we can expect a macromolecular sequence, such as that of an rRNA, to replace DNA base composition as a requisite for publication of a new genus or even species. All facets of microbiology may soon be reshaped by evolutionary considerations, as will our concepts of what a bacterium is and the role bacteria play in the history and physical state of this planet.

Partial sequencing of 16S rRNA, i.e., the oligonucleotide cataloging method (4), has brought our understanding of bacterial phylogeny to its present level (5, 26). Through 1984, over 400 bacterial species had been characterized by this method, and about 10 major phylogenetic groups of eubacteria, each at least the equivalent phylogenetically of a eucaryotic phylum or division, had been recognized, with the major subdivisions of a number of these taxa also being defined (26). These major groupings and their recognized subdivisions are as follows: (i) gram-positive eubacteria, comprising clostridial (or low G + C) and actinomycete (or high G + C) subdivisions; (ii) purple bacteria and relatives, comprising alpha, beta, gamma, and delta subdivisions, with the delta subdivision including sulfate and sulfur respirers, myxobacteria, and bdellovibrios; (iii) spirochetes and relatives, comprising three subdivisions, i.e., genera *Spirochaeta*, *Treponema*, and *Borrelia*, genus *Leptospira*, and obligately anaerobic halophiles; (iv) bacteroides-flavobacteria-cytophagae, with *Bacteroides* spp. forming one subdivision and flavobacteria-cytophagae forming another; (v) cyanobacteria and chloroplasts; (vi) green sulfur bacteria; (vii) green nonsulfur bacteria; (viii) radioresistant bacteria; and (ix) planctomyces and relatives. The definition of the high-level bacterial taxonomic units puts bacterial phylogeny on a par with eucaryotic phylogeny in the sense that the major or primary divisions have now been identified for both systems, but in both cases the relationships among the higher-level units, i.e., their branching orders relative to one another, remain uncertain.

The cataloging studies suggest a number of unexpected relationships among eubacteria and, in the process, vitiate much of the conventional wisdom concerning bacterial relationships (5, 16, 17, 24, 26–29). One of the more unusual associations suggested is the clustering of the bacteroides with the cytophagae, flavobacteria, and their relatives (16). We here firmly establish this relationship by using full sequencing of the 16S ribosomal RNA genes from *Bacteroides fragilis* and *Flavobacterium heparinum*, and we show that the unit so formed probably represents a deep branching in the eubacterial phylogenetic tree.

MATERIALS AND METHODS

Bacterial strains. Several grams of frozen *B. fragilis* ATCC 25285 cells were kindly provided by B. J. Paster, Department of Dairy Science, University of Illinois. *F. heparinum* IFO12017 (ATCC 13125) was grown in nutrient broth (Difco Laboratories) at 30°C and harvested by centrifugation.

Cloning. Nucleic acids RNA and DNA were isolated from the frozen cell pellets by standard procedures (12, 25). An rRNA gene for *F. heparinum*, contained in a DNA fragment produced by restriction endonuclease *Eco*RI cleavage, was cloned into a lambda gt variant (lambda gtWES.lambda B) (10). It was subsequently subcloned into single-stranded phage M13mp8 and M13mp9 (14) either as a 4-kilobase fragment produced by *Eco*RI or as a smaller fragment produced by double digestion with restriction endonucleases *Eco*RI and *Bam*HI. (Sites for restriction endonuclease *Eco*RI are located upstream of the 5′ terminus of this 16S rRNA gene and in the 23S rRNA gene, while a site for restriction endonuclease *Bam*HI occurs at position 545 in the 16S rRNA gene [*Escherichia coli* numbers].)

For *B. fragilis*, an original 7.6-kilobase fragment produced by restriction endonuclease *Bcl*I was cloned into the site for *Bam*HI in lambda L47 (11). The gene was subsequently removed as two pieces of 1.1 and 2.0 kilobases produced by endonuclease *Hin*dIII, which were then subcloned in both

* Corresponding author.

orientations in M13mp8 (a site for *Hin*dIII exists at position 82 in the 16S rRNA gene).

Sequencing methods. For sequencing, the dideoxynucleotide chain termination method (19) was used; templates were produced in the single-stranded phage M13 system (14). Synthesized strands were labeled by the inclusion of *d*(alpha-[^{35}S]thio)ATP (1). Two types of G sequencing reactions were routinely employed; one reaction was normal, and in the other, dGTP was replaced by dITP (2′,3′-dideoxy GTP being used to terminate chain growth; 15). The usual M13 priming site (14) as well as specific priming sites within the rRNA genes, for which primers were synthesized (most of them by the University of Illinois DNA Synthesis Facility), were used. The rRNA-specific primers were designed for regions of the 16S rRNA molecule whose sequence tends to be common to most if not all eubacteria and in some cases to archaebacteria as well (23). Oligonucleotides of 15 to 17 bases that prime in the forward (i.e., same sequence as the rRNA) and reverse directions were synthesized. Those oligonucleotides used in the present study cover positions 10F, 125R, 260F, 270R, 350F/R, 520F/R, 690R, 790F, 920R, 1100F/R, 1240F/R, 1400F/R and 1540R (*E. coli* numbers; F means the same sequence as rRNA, and R means its complement).

Approximately 80% of the *B. fragilis* gene sequence was determined in both the forward and reverse directions. For *F. heparinum*, 80% or so of the sequence was determined in the reverse direction only (20% in both directions), and 20 bases near the 3′ end were not determined. The rRNAs of both organisms have been cataloged (16), and the sequence of the gene agrees with the corresponding oligonucleotides in each case. (The 3′-terminal 10 or so nucleotides for *F. heparinum* were taken from the rRNA catalog [16].)

RESULTS AND DISCUSSION

Bacteroides spp. are widespread, major species found in the lower guts of animal hosts. Members of the genus account for approximately 30% of all fecal isolates (18). The phenotypic description of the genus is remarkably loose and nonspecific, i.e., anaerobic, chemo-organotrophic, gram-negative, non-sporeforming, pleiomorphic rods that are motile or nonmotile, many of which ferment a variety of compounds (6). It is surprising that such a description would define a phylogenetically coherent grouping, yet it does. With only a few exceptions, e.g., *Bacteroides succinogenes* and *B. amylophilus* (16, 28), the 13 species whose 16S rRNAs have been cataloged form a phylogenetically coherent unit (16).

A similarly loose description defines the genera *Flavobacterium* and *Cytophaga*. The primary characteristic defining the former genus is pigmentation; the latter is defined by gliding motility. It has been difficult to keep the two genera separate, for gliding motility is often seen among flavobacteria, at which point the tendency has been to reclassify them as cytophagae. (This situation is even the case for the type species *F. aquatile* [20].) The matter will not be resolved by making the existing definitions of the two groups more precise but, rather, by realizing that gliding motility (or the lack thereof) and most pigment characteristics are not the kinds of characters by which phylogenetically valid taxa can be defined. Indeed, rRNA oligonucleotide comparisons show the two taxa to be intermixed phylogenetically and the group so defined to contain other genera also, e.g., genera *Saprospira*, *Sporocytophaga*, and *Flexibacter* (16).

However, an even more unexpected relationship for these organisms is suggested by the rRNA cataloging studies. There seems to exist a phylogenetically distant but specific relationship between the flavobacterium-cytophaga cluster and *Bacteroides* spp. (16). This relationship is near the limit of detectability for the rRNA cataloging method, and given that a natural relationship between the two groups is not strongly suggested (if at all) by their phenotypes, stronger evidence is required for microbiologists to take the projected relationship seriously. Full sequencing of representative 16S rRNAs from this group, reported herein, provides such evidence.

The sequences of the 16S rRNA genes from *B. fragilis* and *F. heparinum* are shown aligned with those of *E. coli* (2), *Agrobacterium tumefaciens* (30), *Desulfovibrio desulfuricans* (H. Oyaizu, and C. R. Woese, Syst. Appl. Microbiol, in press), *Bacillus subtilis* (C. J. Green, G. C. Stewart, M. A. Hollis, B. S. Vold, and K. S. Bott, Gene, in press), "*Anacystis nidulans*" (21), and a representative archaebacterium, *Methanococcus vannielii* (9) (Fig. 1).

Table 1 gives the percent similarity for the various pairs of sequences. The lower-left values are calculated (from the Fig. 1 alignment) by using only those positions, a total of 1,461, represented in all eubacterial sequences. The upper-right values are calculated similarly, except that all positions of constant composition among the eubacterial sequences have been removed from consideration as well; these positions have no phylogenetic significance among the eubacteria in any case. This latter calculation helps to convince one that the relatively small differences in percent similarity seen in the lower-left values of Table 1 are indeed significant.

A natural relationship between *B. fragilis* and *F. heparinum* clearly emerges from these full-sequence data. With the data in Table 1, this relationship, for example, is more pronounced, than the relationships among three of the major branches of the purple bacteria (27–29; Oyaizu and Woese, in press), a group represented in Fig. 1 by the *A. tumefaciens*, *D. desulfuricans*, and *E. coli* sequences, despite the fact that the bacteroides, at least, are or were rapidly evolving lines of descent (see below).

Percent similarity is the result of three contributions, i.e., common ancestral composition at a given position (ancestral to the entire group, that is), common derived composition (i.e., ancestral only for a specific subline), and common composition due to a fortuitous evolutionary convergence at a given position. Only common derived composition bespeaks true specific relationship. For the more slowly evolving lines, the ancestral component can be disproportionately large and, therefore, sometimes gives the appearance of a specific relationship when none actually exists. Such seems to be the case for the apparent relationship between *B. subtilis* and "*A. nidulans*" suggested by the data in Table 1. These two sequences each have a higher percent similarity with the outgroup archaebacterial sequence (*M. vannielii*) than do any of the other eubacteria (in particular the bacteroides and flavobacteria); i.e., they seem to have retained more ancestral sequence pattern than have the other eubacterial rRNA sequences.

A true phylogenetic relationship should be demonstrable in terms of the derived characters unique to a particular subline. Scoring the Fig. 1 alignment for positions that are of the same composition in a given pair of sequences but of different composition in all remaining eubacterial sequences selectively enriches for the derived contribution. For the *B. fragilis-F. heparinum* couple, the number of such positions is 75, whereas for any other pair of eubacterial sequences in Fig. 1, the number is under 25. A further enrichment for

```
Bf UUACAACGAAGA GUUUGAUCCU  GGCUCAGGAU  GAACGCUAGC  UACAGGC..UUA ACACAUGCAA GUCGAGGGGC AUCAGGAAGA AAGCUUGCUUUC UUUGCUGGCG  ACCGGCGCAC  GGGUGAGUAA
Fh UUACAAUGGAGA GUUUGAUCCU  GGCUCAGGAU  GAACGCUAGC  GGCAGGC..CUA AUACAUGCAA GUCGAACGAG AU....UAAG GGGCUUGCUCCU UA....UGAA  AGUGGCGCAC  GGCUGCGUAA
Ec ..AAAUUGAAGA GUUUGAUCAU  GGCUCAGAUU  GAACGCUGGC  GGCAGGC..CUA ACACAUGCAA GUCGAACGGU AACAG.GAAG AAGCUUGCUUCU UU.GCUGACG  AGUGGCGGAC  GGGUGAGUAA  120
At CUCAACUUGAGA GUUUGAUCCU  GGCUCAGAAC  GAACGCUGGC  GGCAGGC..UUA ACACAUGCAA GUCGAACG.. .........C CCCGCAAGGGG. ..........  AGUGGCAGAC  GGGUGAGUAA
Dd .UGAACUGGAGA GUUUGAUCCU  GGCUCAGAUU  GAACGCUGGC  GGCGUGC..UUA ACACAUGCAA GUCGAACGCG AAA.....GG GACUUCGGUCCU ....GAGUAA  AGUGGCGCAC  GGGUGAGUAA
Bs .UUUAUCGGAGA GUUUGAUCCU  GGCUCAGGAC  GAACGCUGGC  GGCGUGC..CUA AUACAUGCAA GUCGAGCGGA CAGG...UGG GAGCUUGCUCCC G....AUGUU  AGCGGCGGAC  GGGUGAGUAA
An .CAAAAUGGAGA GUUUGAUCCU  GGCUCAGGAU  GAACGCUGGC  GGCGUGC..UUA ACACAUGCAA GUCGAACG.. ........GG GACUUCGGUCCU ..........  AGUGGCGGAC  GGGUGAGUAA
Mv .......AUUCC GGUUGAUCCC  GCCGGAGGCU  .ACUGCUAUU  GGGGUUCGACUA AGCCAUGCGA GUCUA..... .........U GGUUUCGGCCA. ..........  ..UGGCGGAC  GGCUCAUUAA

Bf CACGUAUCCAA CCUGCCCUUU ACUCGGGGAU AGCCUUUCGA AAGAAAGAUU AAUACCCGAU AGCAUAAUGAUUCC.GCAU.GGUUUCA UUAUUAAAGG .........AU U.......CCG GUAAAGGAUG GGGAUGCGUU
Fh CGCGUAUGCAA CCUACCUUAA UCAGGGGGAU AGCCCGAAGA AAUUCGGAUU AACACCGCAU AAAAACACAGGAUA.GCAU.UAUCCAA UGUUCAAAUA ........UUU A.......UAG GAUUAAGAUG GGCAUGCGUG
Ec UGUCUGGGAA. ACUGCCUGAU GGAGGGGGAU AACUACUGGA AACGGUAGCU AAUACCGCAU AACGUC.........GCAA......... GACCAAAGAG GGG.GACCUUC GGGC.CUCUUG CCAUCGGAUG UGCCCAGAUG
At CGCGUGGGAA. UCUACCGUGC CCUGCGGAAU AGCUCCGGGA AACUGGAAUU AAUACCGCAU ACGCCC.........UACG......... GGGGAAAGA. ........UUU A.......UCG GGGUAUGAUG AGCCCGCGUU
Dd CGCGUGGAUAA UCUGCCCUUA UGAUCGGGAU AACAGUUGGA AACGGCUGCU AAUACCGGAU ACGCUCAAAA....UGAACU...UUUU GAGGAAAGAU GGCCUCUGCUU GCAUGCUAUCA CGUAAGGAUG AGUCCCGCUC
Bs CACGUGGGUAA CCUGCCUCUA AGACUGGGAU AACUCCGGGA AACCGGGCGU AAUACCGGAU GGUUGUUUGAACC..GCAUGGUUCAAA CAUAAAAGGU GGC.....UUC G...GCUACCA CUUACAGAUG GACCCGCGGC
An CGCGUGAGAA. UCUGCCUACA GGACGGGGAC AACAGUUGGA AACGACUGCU AAUACCCGAU G.UGCC.........GACA........ GGUGAAACA. ........UUU A.......UGG CCUGUAGAUG AGCUCGCGUC
Mv CACGUGGUUAA CUUAACCUCA GGUGGAGCAU AACCUUGGGA AACUGAGGAU AAUUCUCCAU AAGAAAAGCAGUCUGGAACGAUUCUUU UCUGAAAGC. .........AUA U.......GCG CCCGAGGAUA GGACUGCGCU

Bf CCAUUAGGUU GUUGGUGAGG UAACGGCUCA CCAAGCCUUC GAUGGAUAGG GGUUCUGAGA GGAAGGUCCC CCACAUUGGA ACUGAGACAC GGUCCAAACU CCUACGGGAG GCAGCAGUGA
Fh UCAUUAGCUA GUUGGCGGGG UAACGGCCCA CCAAGGCGAC GAUGACUAGG GGAUCUGAGA GGAUGACCCC CCACACUGGU ACUGAGACAC GGACCAGACU CCUACGGGAG GCAGCAGUAA
Ec GGAUUAGCUA GUAGGUGGGG UAACGGCUCA CCUAGGCGAC GAUCCCUAGC UGGUCUGAGA GGAUGACCAG CCACACUGGA ACUGAGACAC GGUCCAGACU CCUACGGGAG GCAGCAGUGG  360
At GGAUUAGCUA GUUGGUGGGG UAAAGGCUCA CCAAGGCGAC GAUCCAUAGC UGGUCUGAGA GGAUGAUCAG CCACAUUGGG ACUGAGACAC GGCCCAAACU CCUACGGGAG GCAGCAGUGG
Dd CCAUUAGCUU GUUGGCGGGG UAACGGCCCA CCAAGGCAUC GAUGGGUAGC CGAUUUGAGA GGAUGAUCGG CCACACUGGA ACUGAAACAC GGUCCAGACU CCUACGGGAG GCAGCAGUGG
Bs GCAUUAGCUA GUUGGUAGGG UAACGGCUCA CCAAGGCAAC GAUGCGUAGC CGACCUGAGA GGGUGAUCGG CCACACUGGG ACUGAGACAC GGCCCAGACU CCUACGGGAG GCAGCAGUAG
An UGAUUAGCUA GUUGGUGGGG UAAGGGCCUA CCAAGGCGAC GAUCAGUAGC UGGUCUGAGA GGAUGAUCAG CCACACUGGG ACUGAGACAC GGCCCAGACU CCUACGGGAG GCAGCAGUGG
Mv CGAUUAGGUA GUUGGUGGGG UAAUGGCCCA CCAAGCCUAC GAUCGAUACG GGCCUUGAGA GAGGGAGCCC GGAGAUGGGG ACUGAGACAC GGCCCCAGGC CCUACGGGGC GCAGCAGGCG

Bf GGAAUAUUGG UCAAUGGGCG CUAGCCUGAA CCAGCCAAGU AGCGUGAAGG AUGAAGGCUCU AUGGGUCGUA AACUUCUUUU AUAUAAGAAU AAAG..UGCAG UA..UGUA.U ACUGU..UUU
Fh GGAAUAUUGG UCAAUGGAGG GAACUCUGAA CCAGCCAUGC CGCGUGCAGG AAGACAGCCCU CUGGGUCGUA AACUGCUUUU AUUCGGGAAU AAAC..CUACU UA..CGUG.U AAGUA..GCU
Ec GGAAUAUUGC ACAAUGGGCG CAAGCCUGAU GCAGCCAUGC CGCGUGUAUG AAGAAGGCCU. UCGGGUUGUA AAGUACUUUC AGCGGGGAGG AAGG.GAGUAA AGUUAAUACC UUUGCUCAUU  480
At GGAAUAUUGG ACAAUGGGCG CAAGCCUAGU CCAGCCAUGC CGCGUGAGUG AUGAAGGCCU. UAGGGUUGUA AAGCUCUUUC ACCGGAGAAG AUAA..... .......... .........U
Dd GGAAUAUUGC GCAAUGGGCG AAAGCCUGAC GCAGCGACGC CGCGUGAGGG AUGAAGGUUU. UCGGAUCGUA AACCUCUGUC AGAAGGGAAG AAACUACGUUG UGCUAAU.CA GCAGCGUACU
Bs GGAAUCUUCC GCAAUGGACG AAAGCUCGAC GGAGCAACGC CGCGUGAGUG AUGAAGGUUU. UCGGAUCGUA AAGCUCUGUU GUUAGGGAAG AACAAGUACCG UUCGAAUAGG GCGGUACCUU
An GGAAUUUUCC GCAAUGGGCG CAAGCCUGAC GGAGCAACGC CGCGUGGGGG AGGAAGGUUU. UUGGACUGUA AACCCCUUUU CUCAGGGAAG AAGA..... .......... ......AAGU
Mv CGAAACCUCC GCAAUGCACG AAAGUGCGAC GGGGGGACCC CAAGUGCUCA UGC..... ACA.....GC AUGGGCUUUU AUCAAGU... ....GUAAACA .......... ..........

Bf GUAUGUAUUA UAUGAAUAAG GAUCGGCUAA. CUCCGUGCCA GCAGCGCGGU UAAUACGGAG GAUCCGAGCG UUAUCCGGAU UUAUUGGGUU UAAAGGGAGC GUAGGUGGAC UGGUAAGUCA
Fh GAAUGUACCG AAGGAAUAAG GAUCGGCUAA. CUCCGUGCCA GCAGCCGCGG UAAUACGGAG GAUCCAAGCG UUAUCCGGAU UUAUUGGGUU UAAAGGGUGC GUAGGCGGCC UGUUAAGUCA
Ec GACGUUACCC GCAGAAGAAG CACCGGCUAA. CUCCGUGCCA GCAGCCGCGG UAAUACGGAG GGUGCAAGCG UUAAUCGGAA UUACUGGGCG UAAAGCGCAC GCAGGCGGUU UGUUAAGUCA  600
At GACGGUAUCC GGAGAAGAAG CCCCGGCUAA. CUUCGUGCCA GCAGCCGCGG UAAUACGAAG GGGGCUAGCG UUGUUCGGAA UUACUGGGCG UAAAGCGCAC GUAGGCGGUU UGUCAAGUCA AUUUAAGUCA
Dd GACGGUACCU UCAAAGGAAG CACCGGCUAA. CUCCGUGCCA GCAGCCGCGG UAAUACGGAG GGUGCAAGCG UUAAUCGGAA UUACUGGGCG UAAAGCGCAC GUAGGCUGUA GUGUAAGUCA
Bs GACGGUACCU AACCAGAAAG CCACGGCUAA. CUACGUGCCA GCAGCCGCGG UAAUACGUAG GUGGCAAGCG UUGUCCGGAA UUAUUGGGCG UAAAGCGCUC GCAGGCGGUU UCUUAAGUCU
An GACGGUACCU GAGGAAUAAG CCUCGGCUAA. UUCCGUGCCA GCAGCCGCGG UAAUACGGGA GAGGCAAGCG UUAUCCGGAA UUAUUGGGCG UAAAGCGCCU GCAGGCGGUU AAUCAAGUCU
Mv ......GCUU GAGGAAUAAG GGCUGGGCAAG UUCGGUGCCA GCAGCCGCGG UAAUACCGAC GGCCCGAGUG GUAGCCACUC UUAUUGGGCC UAAAGCGUCC GUAGCCGGUC CAGUAAGUCC

Bf GUUGUGAAAG UUUGCGGCUC AACCGUAAAA UUGCAGCUGA UACUGUCAGU CUUGAGUACA GUAGAGGUGG GCGGAAUUCG UGGUGUAGCG GUGAAAUGCU UAGAUAUCAC GAAGAACUCC
Fh GGGGUGAAAG ACGGUAGCUC AACUAUCGCA GUGCCCUUGA UACUGACGGG CUUGAAUACA CUAGAGGUAG GCGGAAUGUG ACAAGUAGCG GUGAAAUGCA UAGAUAUGUC ACAGAACACC
Ec GAUGUGAAAU CCCCGGGCUC AACCUGGGAA CUGCAUCUGA UACUGGCAAG CUUGAGUCUC GUAGAGGGGG GUAGAAUUCC AGGUGUAGCG GUGAAAUGCG UAGAGAUCUG GAGGAAUACC  720
At GGGGUGAAAU CCCAGAGCUC AACUCUGGAA CUGCCUUUGA UACUGGCUAU CUUGAGUAUG GAAGAGGUAA GUGGAAUUCC GAGUGUAGAG GUGAAAUUCG UAGAUAUUCG GAGGAACGCC
Dd GGGGUGAAAU CCCACGGCUC AACCGUGGAA CUGCCUUUGA UACUGCACAA CUUGAAUCCG GGAGAGGGUG GCGGAAUUCC AGGUGUAGGA GUGAAAUCCG UAGAUAUCUG GAGGAACAUC
Bs GAUGUGAAAG CCCCCGGCUC AACCGGGAG GGUCAUUGGA AACUGGGGAA CUUGAGUGCA GAAGAGGAGA GUGGAAUUCC ACGUGUAGCG GUGAAAUGCG UAGAAGUCUG GAGGAACACC
An GUUGUCAAAG CGUGGGGCUC AACCUCAUAC AGGCAAUGGA AACUGAUUGA CUAGAGUAUG GUAGGGGUAG CGGGAAUUCC AGGUGUAGCG GUGAAAUGCG UAGAUAUCUG GAAGAACACC
Mv CUGUUUAAAU UCUCUGGCUU AACCAGAGGA CUGGCAGGGA UACUGCUGGA CUUGGGACCG GGAGAGGACA AGGGUACUCC UUGAUCCUGG CGUAAAUGUG UUGAUCCUUG GAGGACCACC

Bf GAUUGCGAAG GCAGCUCACU GGACUGCAAC UGACACUGAU GCUCGAAAGU GUGGGUAUCA AACAGGAUUA GAUACCCUGG UAGUCCACAC AGUAAACGAU GAAUACUCGC UGUUUGCGA.
Fh GAUUGCGAAG GCAGCUUACU AUGGUGUGAU UGACGCUGAG GCACGAAAGC GUGGGGAUCA AACAGGAUUA GAUACCCUGG UAGUCCACGC CCUAAACGAU GAAUacuggC UGUUAGCGA.
Ec GGUGGCGAAG GCGGCCCCCU GGACGAAGAC UGACGCUCAG GUGCGAAAGC GUGGGGAGCA AACAGGAUUA GAUACCCUGG UAGUCCACGC CGUAAACGAU GUCGACUUGG AGGUUGUGCC  840
At AGUGGCGAAG GCGGCUUACU GGUCCAUUAC UGACGCUGAG GUGCGAAAGC GUGGGGAGCA AACAGGAUUA GAUACCCUGG UAGUCCACGC CGUAAACGAU GAAUGUUAGC CGUCGGGCAG
Dd AGUGGCGAAG GCGGCCACCU GGACCGGUAU UGACGCUGAG GUGCGAAAGC GUGGGGAGCA AACAGGAUUA GAUACCCUGG UAGUCCACGC UGUAAACGAU GGAUGCUAGA GUCGGGGCA.
Bs AGUGGCGAAG GCGACUCUCU GGUCUGUAAC UGACGCUGAG GAGCGAAAGC GUGGGGAGCG AACAGGAUUA GAUACCCUGG UAGUCCACGC CGUAAACGAU GAGUGCUAAG UGUUAGGGGG
An AGCGGCGAAA GCGCCCUACU GGGCCAUAAC UGACGCUCAU GGACGAAAGC UAGGGGAGCG AAAGGGAUUA GAUACCCCUG UAGUCCUAGC CGUAAACGAU GAACACUAGG UGUUGCGUGA
Mv UAUGGCGAAG GCACUUGUCU GGAACGGGUC CGACGGUGAG GGACGAAAGC CAGGGGCGCG AACCGGAUUA GAUACCCGGG UAGUCCUGGC CGUAAACUCU GCGAACUAGG UGUCACCUGG

Bf ...UAUA...CAGU AAGCGGCCAA GCGAAAGCAU UAAGUAUUCC ACCUGGGGAG UACGCCGGCA ACGGUGAAAC UCAAAGGAAU UGACGGGGGC CCGC.ACAAGC GGAGGAACAU GUGGUUUAAU
Fh ...UAUA...CAGU UAGCGGCUAA GCGAAAGCGU UAAGUAUUCC ACCUGGGGAG UACGCCCGCA AGGGUGAAAC UCAAAGGAAU UGACGGGGGC CCGC.ACAAGC GGAGGAGCAU GUGGUUUAAU
Ec ..CUUGA..GGCGU GGCUUCCGGA GCUAACGCGU UAAGUCGACC GCCUGGGGAG UACGCCCGCA AGGUUAAAAC UCAAAUGAAU UGACGGGGGC CCGC.ACAAGC GGUGGAGCAU GUGGUUUAAU  960
At ...UAUA..CUGUU CGGUGGCGCA GCUAACGCAU UAAACAUUCC GCCUGGGGAG UACGGUCGCA AGAUUAAAAC UCAAAGGAAU UGACGGGGGC CCGC.ACAAGC GGUGGAGCAU GUGGUUUAAU
Dd ...GUAU...UCUU CGGUGUCGUA GUUAACGCGU UAAGCAUCCC GCCUGGGGAG UACGGUCGCA AGGCUGAAAC UCAAAGGAAU UGACGGGGGC CCGC.ACAAGC GGUGGAGUAU GUGGUUUAAU
Bs U..UUCC.GCCCCU UAGUGCUGCA GCUAACGCAU UAAGCACUCC GCCUGGGGAG UACGGUCGCA AGACUGAAAC UCAAAGGAAU UGACGGGGGC CCGC.ACAAGC GGUGGAGCAU GUGGUUUAAU
An A..UCGA.CCCGCG CAGUGCCGUA GCCAACGCGU UAAGUGUCCC GCCUGGGGAG UACGCACGCA AGUUGGAAAC UCAAAGGAAU UGACGGGGGC CCGC.ACAAGC GGUGGAGUAU GUGGUUUAAU
Mv GCCUCGAGCCCAGG UGGUGCCGAA GGGAAGCCGU UAAGUUCGCC GCCUGGGGAG UACGGUCGCA AGACUGAAAC UUAAAGGAAU UGGCGGGGGA GCACCACAACG GGUGGAGCCU GCGGUUUAAU

Bf UCGAUGAUAC GCGAGGAACC UUACCCGGGC UUAAAUUGCAG UGGAAUGAUG.. UCGAAACAUG UCAG.UGA.GC AA..UCACCGC UGUGAAGGUG CUGCAUGGUU GUCGUCAGCU CGUGCCGUGA
Fh UCGAUGAUAC GCGAGGAACC UUACCCGGGC UUGAAAGU.UA GUGAAUUAGA.. UAGAGAUAUU UAAG.UGA.GC AA..UCACACG AAACUAGGUG CUGCAUGGCU GUCGUCAGCU CGUGCCGUGA
Ec UCGAUGCAAC GCGAAGAACC UUACCUGGUC UUGACAUC.CA CGGAAGUUUU.. CAGAGAUGAG AAUGUGCC.UU CG.GGAACCGU GAGACAGGUG CUGCAUGGCU GUCGUCAGCU CGUGUCGUGA  1080
At UCGAAGCAAC GCGCAGAACC UUACCAGCUC UUGACAUU.CG GGGUUUGGGCAG UGGAGACAUU GUCCUUCAGUU AGGCUGGCCCC AGAACAGGUG CUGCAUGGCU GUCGUCAGCU CGUGUCGUGA
Dd UCGAUGCAAC GCGAAGAACC UUACCUAAGU UUGACAUC.CA CGGAACCUCC.. CCGAAAAGGA GGGGUGCCCUU CGGGGAGCCGU GAGACAGGUG CUGCAUGGCU GUCGUCAGCU CGUGUCGUGA
Bs UCGAAGCAAC GCGAAGAACC UUACCAGGUC UUGACAUC.CU CUGACAAUCC.. UAGAGAUAGG ACGUCCCC.UU CG.GGGGCAGA GUGACAGGUG GUGCAUGGUU GUCGUCAGCU CGUGUCGUGA
An UCGAUGCAAC GCGAAGAACC UUACCAGGGU UUGACAUC.CC CCGAAUCUCU.. UGGAAACGAG AGAGUGCC.UU CG.GGAGCGGG GAGACAGGUG CUGCAUGGCU GUCGUCAGCU CGUGUCGUGA
Mv UGGAUUCAAC GCCGGGCAUC UCACCAGGAG CGA...CA.GC AUGAUGACGGCC AGGUUGACGA CCU..... UGCCUGAAGCG CUGAGAGGUG GUGCAUGGCC AUCGUCAGCU CGUACCGCGA

Bf GGUGUCGGCU UAAGUGCCAU AACGAGCGCA ACCCUUAUCU UUAGUUACUA ACAG...GUUAUG...CUGAGGACUCUAGAGAGACUG CCGUCGUAAG AUGU.GAGGAA GGUGGGGAUG ACGUCAAAUG
Fh GGUGUUGGGU UAAGUCCCGC AACGAGCGCA ACCCCUAUGU UUAGUUGCCA GCAC...GUUAAG...GUGGGGACUCUAAACAGACUG CCUGUC.CAA ACAGAGAGGAA GGAGGGCACG ACGUCAAGUC
Ec AAUGUUGGGU UAAGUCCCGC AACGAGCGCA ACCCCUAUCC UUUGUUGCCA GCGG...UCCGG...CCGGGAACUCA.AAGGAGACUG CCAGUGAUAA ACUG.GAGGAA GGUGGGGAUG ACGUCAAGUC  1200
At GAUGUUGGGU UAAGUCCCGC AACGAGCGCA ACCCCUGCCC UUAGUUGCCA GCAU....UUAGU...UGGGCACUCU.AAGGGCAGUC CCGGUCAUAA GCCGAGAGGAA GGUGGGGAUG ACGUCAAGUC
Dd GAUGUUGGGU UAAGUCCCGC AACGAGCGCA ACCCCUAUGG AUAGUUGCCA GCAA....GUAAUG.UUGGGCACUCU.AUUCAGACUG CCCGGGUUAA CCGG.GAGGAA GGUGGGGACG ACGUCAAGUC
Bs GAUGUUGGGU UAAGUCCCGC AACGAGCGCA ACCCUUGAUC UUAGUUGCCA GCA....UUCAGU...UGGGCACUCU.AAGGUGACUG CCGGUGACAA ACCG.GAGGAA GGUGGGGAUG ACGUCAAAUC
An GAUGUUGGGU UAAGUCCCGC AACGAGCGCA ACCCACGUUU UUAGUUGCCA UCA....UUCAGU...UGGGCACUCU.AGAGAAACUG CCGGUGACAA ACCG.GAGGAA GGUGUGGACG ACGUCAAGUC
Mv GGCGUCCUGU UAAGUCAGGU AACGAGCGAG ACCCGUGCCC UAUGUUUGCGA CUACUUUCUCCGGAAGGUAAGCACUCAUAGGGGACCG CUAGCGCUAA GCUA.GAGGAA GGAGCGGGCA ACGAUAGGUC

Bf AGCACGGCCC UUACGUCCGG GGCUACACAC GUGUUACAAU GGGGGGUACA GAAGGCAGCU AGCGGGUGAC CGUAUGCUAA UCCCCAAAAU. CCUCUCUCAG UUCGGAUCGA AGUCUGCAAC
Fh AUCAUGGCCC UUACGUCCGG GGCUACACAC GUGCUACAAU GGAUGGUACA GAGGGCAGCA AGCUGGCAAC AGCAAGCGAA UCUCAAAAAG CCAUUCACAG UUCGGAUAGA GGUCUGCAAC
Ec AUCAUGGCCC UUACGACCAG GGCUACACAC GUGCUACAAU GGCGCAUACA AAGAGAAGCG ACCUCGCGAG AGCAAGCGGA CCUCAUAAAG UGCGUCGUAG UCCGGAUUGG AGUCUGCAAC  1320
At CUCAUGGCCC UUACGGGCUG GGCUACACAC GUGCUACAAU GGUGGUGACA GUGGGCAGCG AGACAGCGAU GUCGAGCUAA UCUCCAAAA. GCCAUCUCAG UUCGGAUUGC ACUCUGCAAC
Dd AUCAUGGCCC UUACGCCUAG GGCUACACAC GUACUACAAU GGCGCGCACA AAGGGGAGCG AGACCGCGAG GUGGAGCCAA UCCCAAAAAA CGCGUCCCAG UCCGGAUUGC AGUCUGCAAC
Bs AUCAUGCCCC UUAUGACCUG GGCUACACAC GUGCUACAAU GGACAGAACA AAGGGCAGCC AAACCCGCAG GUCUAAGCCAA UCCCACAAAU CUGUUCUCAG UUCGGAUCGC AGUCUGCAAC
An AUCAUGCCCC UUACAUCCUG GGCUACACAC GUACUACAAU GGCUCCGGACA GCGAGACGCC AAGCCCGCGAG GUGAAGCAAA UCUCCCAAAC CGGGGCUCAG UUCAGAUUGC AGGCUGCAAC
Mv CGCAUGCCCC GAAUCUCCUG GGCUACACGC GGGCUACAAU GGCUAGGACA AUGGGCUGCU ACCCUGAAAA GGGACGCGAA UCUCCGAAAC CUAGUCGUAG UUCGGAUCGU GGGCUGUAAC

Bf CCGACUUCGU GAAGCUGGAU UCGCUAGUAA UCGCGCAUCA GCCACGGCGCG GUGAAUACGU UCCCGGGCCU UGUACACACC GCCCGUCAAG CCAUGGGAGC CGGGGGUACC UGAAG.UA..
Fh UCGACCUCUU GAAGUUGGAU UCGCUAGUAA UCGCGUAUCA GCAAUGACGCG GUGAAUACGU UCCCGGGCCU UGUACACACC GCCCGUCAAG CCAUGGAAGU UGGGGGUACC UAAAGgUA..
Ec UCGACUCCAU GAAGUCGGAA UCGCUAGUAA UCGUGGAUCA G.AAUGCCACG GUGAAUACGU UCCCGGGCCU UGUACACACC GCCCGUCACA CCAUGGGAGU GGGUUGCAAA AGAAGUAGGU  1440
At UCGAGUGCAU GAAGUUGGAA UCGCUAGUAA UCGCAGAUCA G.CAUGCUGCC GUGAAUACGU UCCCGGGCCU UGUACACACC GCCCGUCACA CCAUGGGAGU UGGUUUUACC CGAAGGUAGU
Dd UCGACUGCAU GAAGUUGGAA UCGCUAGUAA UUCGAGAUCA G.CAUGCUCGG GUGAAUGCGU UCCCGGGCCU UGUACACACC GCCCGUCACA CCACGAAAGU CGGUUUUACC CGAAGCCGGU
Bs UCGACUGCGU GAAGCUGGAA UCGCUAGUAA UCGCGGAUCA G.CAUGCCGCG GUGAAUACGU UCCCGGGCCU UGUACACACC GCCCGUCACA CCACGAGAGU UUGUAACACC CGAAGUCGGU
An UCGCCUGCAU GAAGGCGGAA UCGCUAGUAA UCGCAGGUCA G.CAUACUGCG GUGAAUACGU UCCCGGGCCU UGUACACACC GCCCGUCACA CCAUGGAAGU UGGCCAUGCC CGAAGUCGGU
Mv UCGCCCACGU GAAGCUGGAA UCCGUAGUAA UCGCAGUUCA U.AAUACUGCG GUGAAUGUGU CCCUGCUCCU UGCACACACC GCCCGUCACA CCACCCGAGU UGGGUUCAGG UGAGGCCUUG

Bf ..CGUAACC.G CAA.GGAUCG. ...UCCUAGG GUAAAACUGG UGACUGGGGC UAAGUCGUAA CAAGGUAGCC GUACCGGAAG GUGCGGCUGG AACACCUCCUUUUCU
Fh ..CGUAACC.G CAA.GGAGCG. ..UCCUUAGG GUAAAACCGA UAACUGGGGC UAAGUCGUAA CAAGGUAgcc gUACCGGAAG GnnnnnnnnG AACACCUCCUUUUCU
Ec AGCUUAACC.U UCG.GGAGGGC GCUUACCACU UUGUGAUUCA UGACUGGGGU GAAGUCGUAA CAAGGUAACC CUAGCGGAUG CUGCGGUUGC AUCACCUCCUUA..  1542
At GCGCUAACC.G CAA.GGAGGCA GCUAACCACG GUAGGGUCAG CGACUGGGGU GAAGUCGUAA CAAGGUAGCC GUAGGGGAAC CUGCGGCUGG AUCACCUCCUUUCU
Dd CAGCCAACCAG CAAUGGAGCCA GCCGUCUACG GUAGGGCCGA UGAUUGGGGU GAAGUCGUAA CAAGGUAGCC GUAGGGGAAC CUGCGGCUGG AUCACCUCCUUU..
Bs GAGGUAACCUU UU.AGGAGCCA GCCGCCGAAG GUGGGACAGA UGAUUGGGGU GAAGUCGUAA CAAGGUAGCC GUAUCGGAAG CUGCGGCUGG AUCACCUCCUUUCU
An ACCCUAACCGU UCGCGGAGGGG GGCGCCGAAG GUAGGGCUGA UGACUGGGGU GAAGUCGUAA CAAGGUAGCC GUACCGGAAG GUGUGGCUGG AUCACCUCCUUU..
Mv GCC..... UUU....GGCU AGGGUCGAAC CUGCGGCUCAG CGAGGGGGGU GAAGUCGUAA CAAGGUAGCC GUAGGGGAAC CUGCGGCUGG AUCACCUCC.....
```

TABLE 1. Percent similarity for the sequences of Fig. 1[a]

Organism	% Similarity							
	B. fragilis	*F. heparinum*	*E. coli*	*A. tumefaciens*	*D. desulfuricans*	*B. subtilis*	*"A. nidulans"*	*M. vannielii*
B. fragilis		60.7	38.2	40.7	39.6	40.2	36.8	NA[b]
F. heparinum	82.8		42.9	46.5	42.9	44.0	42.0	NA
E. coli	72.9	74.9		54.4	56.9	52.7	50.7	NA
A. tumefaciens	74.0	76.5	80.0		58.2	53.7	51.2	NA
D. desulfuricans	73.5	74.9	81.1	81.7		56.8	53.8	NA
B. subtilis	73.8	75.4	79.3	79.7	81.0		57.1	NA
"A. nidulans"	72.3	74.5	78.4	78.6	79.7	81.2		NA
M. vannielii	60.0	60.8	61.3	62.1	62.1	63.9	63.4	

[a] Lower-left values are the percentage of positions in the Fig. 1 alignment in which each pair of sequences has the same composition. Those positions in which any one of the eubacterial sequences has no representation are not considered. The total number of positions considered in this case is 1,461. Upper-right values were determined similarly, except that all positions of constant composition among the eubacterial sequences have also been eliminated from consideration. The total number of positions considered in this case is 641.

[b] NA, Not applicable.

derived characters should result when the alignment is scored for positions not only uniquely common to a given pair of sequences, but of different yet common composition in the remaining eubacterial sequences. For the *B. fragilis-F. heparinum* couple, there are 41 such positions, but 10 or less for any other pair of eubacterial sequences in the alignment. When these 41 positions are screened against a wider set of sequences (mostly unpublished), their number shrinks to 21, or 15 if one does not count separately both members of a recognized base pair (23) in the 16S rRNA secondary structure (unpublished analysis). These stringently defined positions are shown in Table 2. They collectively form a sequence signature for the bacteroides-flavobacterium-cytophaga group.

On the table, bases at various positions in the 16S rRNA sequence have been listed either individually or, if they are involved in secondary structure, as a pair; thus, for example "290–310 G · C" in the table means a G · C pair involving sequence positions 290 to 310. Consider the following six entries on Table 2 in detail.

(i) The (paired) base at position 310 can be traced in the 16S rRNA catalogs of the bacteroides and flavobacteria by oligonucleotides containing the segment CCCCCACAY; all catalogs of *Bacteroides* species contain such an oligonucleotide, as do all but one of the catalogs from the flavobacteria and relatives, but none of the remaining eubacterial catalogs do. However, in over 95% of other eubacterial catalogs (and sequences), and oligonucleotide of the general form (G)YCACAYYG is found; i.e., the C residue at position 310 has been replaced by G (23, 26).

(ii) The U residue at position 570 is unique to the bacteroides-flavobacterium cluster among all published eubacterial 16S rRNA sequences (13, 23). Position 570 is covered by the oligonucleotide UUUAAAG, universal in the bacteroides-flavobacterium-cytophaga cluster, but found elsewhere among the eubacterial 16S rRNA catalogs only once in over 350 catalogs (16, 26). All other eubacterial species would seem to have a G residue at this position, evidenced by the universal occurrence (outside of the bacteroides-flavobacterium cluster) of (G)UAAAG in catalogs and at position 570 in the other eubacterial sequences (unpublished analysis).

(iii) The base at position 680 (paired to that at 710) (23) can be traced in the bacteroides by the oligonucleotide AAUUCG, which is found in 9 of the 11 bacteroides catalogs. The form of this oligonucleotide that is found in most catalogs from the other eubacterial "phyla" is AAUUC[AY] (unpublished analysis). (No member of this family of oligonucleotides is detectable in the flavobacteria and relatives, presumably because of the presence of a G residue located at position 678 [Fig. 1], which makes the oligonucleotide in this case too small to be specifically recognizable.)

(iv) The composition at position 724 can occasionally be measured by the oligonucleotides AAYACCAAUG and AAYACCAAUUG (26). (A much more common variant in this family is AAYACCRG (R is a purine; 26), which is too short to reach the position in question.) Ten examples of sequence AAYACCAAUG are spread among several eubacterial phyla, but none are found in the bacteroides-flavobacterium group, which, however, contains the only four examples among eubacterial catalogs of the alternate form, AAYACCAAUUG (unpublished analysis).

(v) Position 1340 (paired to position 943; 23) can be tracked by the pentamers AAUCG or AUUCG; AAUCG occurs in most eubacterial catalogs except for those of the bacteroides (unpublished analysis). On the other hand, the sequence AUUCG is found in all bacteroides catalogs, but occurs elsewhere among eubacterial catalogs only six times; however, in all six of these occurrences, the majority form, AAUCG, is also present (unpublished analysis), indicating that AUUCG in these cases may not actually occur in the vicinity of position 1340 in the sequence. (Pentamers were not routinely determined in catalogs of the flavobacteria and relatives; so the occurrence of AAUCG or AUUCG is these cases is not known.)

(vi) The final entry in Table 2, AAYACCUCCUU, represents the 3′ terminus of the 16S rRNA molecule. This particular version is confined solely to the bacteroides-

FIG. 1. Alignment of 16S rRNA sequences. Sequences are from *B. fragilis* (Bf), *F. heparinum* (Fh), *E. coli* (Ec) (2), *A. tumefaciens* (At), *D. desulfuricans* (Dd) (Oyaizu and Woese, in press), and *B. subtilis* (Bs) (Green et al., in press). *"A. nidulans"* (An) (21) and *M. vannielii* (Mv) (9) have been aligned by a standard procedure (23). Areas of uncertain sequence are designated by lowercase symbols corresponding to their uppercase counterparts or by N if the nucleotide is totally unknown. A dot signifies the lack of a base at that position in a given sequence.

TABLE 2. Sequence signatures for the bacteroides and flavobacteria (and relatives)

Position no.(s)[a]	Composition[b]		Traced by oligonucleotide[c]	No. of occurrences of oligonucleotide in catalogs of (no. of catalogs)[d]:		
	Bf-Fh	Other		Bact (11)	Flav (12)	Other (>300)
38	A	G	None			
290, 310	G, $\overline{C}$	C, G	..CC$\overline{C}$CCACAY..	11	10	0
501, 544	G, C	C, G	None			
450, 483	U, A	G, C	None			
484	U	R	None			
569, 881	$\overline{U}$, A	G, C	$\overline{U}$UUAAAG	11	12	1
570	$\overline{U}$	G	U$\overline{U}$UAAAG	11	12	1
680, 710	$\overline{G}$, C	C, G	AAUUC$\overline{G}$	9		4
724	$\overline{U}$	G	AAYACCAAU$\overline{U}$G	0	4	0
866	A	C	None			
943, 1340	A, $\overline{U}$	U, A	A$\overline{U}$UCG	11	>3[e]	6[f]
975	G	A	None			
995	A	C	None			
1475	$\overline{A}$	G	UA$\overline{A}$AC[AY]..	8	8	0
1532	$\overline{A}$	U	A$\overline{A}$YACCUCCUU..	11	10	0

[a] Position(s) of base or base pair in the sequence (*E. coli* numbers [2].)
[b] Composition of position in *B. fragilis* and *F. heparinum* (Bf-Fh) sequences or in the remaining eubacterial sequences (other). Overbar indicates which base is being traced by the oligonucleotide.
[c] Oligonucleotide that covers a particular position. Y, Pyrimidine. Bases in brackets are alternatives to one another. When the full sequence is not given, the portion shown is preceded or followed by dots.
[d] Bact, Bacteroides (16); Flav, flavobacteria and their relatives (16); Other, the remaining eubacteria (26).
[e] Presence of oligonucleotide not checked in most cases; therefore, the number of occurrences is at least three.
[f] This number is falsely high; see the text for discussion.

flavobacterium group (16, 26). In all other known cases, i.e., catalogs and all 16S-type rRNA sequences, the base at position 1532 is U, not A (13, 16, 26). (Mitochondria, which show no sequence homology in this region, are not considered here.)

A distinct advantage of the full-sequencing over the cataloging approach is that full sequences permit the use of the molecular phenotype of the rRNA, i.e., various higher-order structural features of the molecule, in defining bacterial taxonomic categories. Certain of the detailed secondary structural features of the molecule seem to be of constant and unique composition in various major phylogenetic groupings. Three examples pertain as follows to the present discussion.

(i) In *B. fragilis* and *F. heparinum*, the helix 829–838/848–857 (23) has the same number of base pairs and an identical unique composition for the innermost three of these base pairs and for the loop spanning them (Fig. 1). Oligonucleotides of the form (G)AUAYAC (which cover the loop of this helix) account for 64 and 50% of the catalogs from bacteroides and flavobacteria, (and their relatives) (16), respectively, but this general composition is found in only 7 of 350 or so other eubacterial catalogs (unpublished analysis).

(ii) The helical structure 1025–1028/1033–1036 (23) has an identical and unique composition in the *B. fragilis* and *F. heparinum* sequences (Fig. 1). A form of the structure this short (1 base pair less than the *E. coli* version) is itself rare, but not unique (23; Woese, unpublished analysis). (Most of its counterparts are relatively large [Fig. 1].)

(iii) The helix 1409–1445/1457–1491 (23) is perhaps the most remarkable of the three helices (Table 3). The overall helix is notable for the number of noncanonical base pairs, bulges, etc., it contains (13, 23). Table 3 shows the portion of the helix whose detailed structure is phylogenetically interesting. The form of this structure that contains three contiguous G · A pairs is found only in *B. fragilis* and *F. heparinum* among the known eubacterial sequences. Oligonucleotides of the form UAAAACA or UAAAACY (see the A′ strand of the helix in Table 3) are found in 73% of the

TABLE 3. Structural feature of 16S rRNA characteristic of several eubacterial divisions

Organism	Sequence of strand[a]		Reference
	A	A′	
	1420 1430	1470 1480	
B. fragilis	CCGG[GGG]UACC	GGUA[AAA]CUGG	
F. heparinum	UUGG[GGG]UACC	GGUA[AAA]CCGA	
E. coli	UGGG·UUG·CAAA	UUUG·UGA·UUCA	2
A. tumefaciens	UUGG·UUU·UACC	GGUA·GGG·UCAG	30
D. desulfuricans	UCGG·UUU·UACC	GGUA·GGG·CCGA	Oyaizu and Woese, in press
B. subtilis	UUUGU[AA]CACC	GGUG[GG]ACAGA	Green et al., in press
H. chlorum	UCGGC[AA]CACC	GGUG[GG]GUCGA	21a
M. capricolum	UUGGU[AA]UACC	GGUA[GG]ACUAG	8
"*A. nidulans*"	UUGGCC[A]UGCC	GGUA[G]GGCUGA	21

[a] Strands A and A′ form double helical structures in each case. The bases involved in A·G and G·A pairs are enclosed in brackets. Sequence positions (*E. coli* numbers) are indicated below column headings.

bacteroides catalogs and 67% of those from the flavobacteria and their relatives (16), but not in any other eubacterial catalogs (26; unpublished analysis).

The helix of Table 3 also shows a version characteristic of the gram-positive bacteria (phylogenetically defined). The two contiguous A · G pairs seen in the *B. subtilis*, *Heliobacterium chlorum*, and *Mycoplasma capricolum* sequences (all of these organisms are gram-positive bacteria phylogenetically) have not been found elsewhere among the eubacteria. (An oligonucleotide of the form YAAYACCCR, which is characteristic of this particular form of the helix, is found in the majority of the 150 or so catalogs from gram-positive bacteria, but nowhere else [26].)

Analysis of the bacteroides-flavobacterium group in terms of oligonucleotide catalogs suggests that *Bacteroides* species are (or were) evolving at a more rapid rate than eubacteria in general and species of *Flavobacterium* and *Cytophaga* in particular (16). This point can be addressed here by determining the number of positions that are unique to a given organism (among the eubacteria) in the Fig. 1 alignment. (In tree construction algorithms, such positions contribute primarily to branch length beyond the terminal branching point for each organism.) Such a count reveals the *B. fragilis* sequence to have 104 unique positions. The next highest number, for *E. coli*, is 86; the lowest, for *B. subtilis*, is 58; while *F. heparinum* has 74. These counts are a strong indication that the bacteroides are, or were, evolving more rapidly than their specific relatives the flavobacteria.

The data presented here give no indication of any specific relationship between bacteroides-flavobacterium and any other major eubacterial group. In fact, the group would seem to represent a particularly ancient divergence in the eubacterial line of descent. Fig. 2 is a phylogenetic tree derived from the percentage similarities of Table 1 (13). The bacteroides-flavobacterium group is seen to diverge from the common line of eubacterial descent before such major eubacterial groups as the purple bacteria, gram-positive bacteria, and cyanobacteria do. (However, we will consider this branching order tentative until it is confirmed by a tree based upon a larger number of [distantly related] eubacterial sequences.)

Given the sequence of *B. fragilis* 16S rRNA and the set of oligonucleotide catalogs for various other *Bacteroides* species (16; unpublished data), it is possible to identify various short sequence stretches (in the range of 20 nucleotides) that might be constructed as probes for the clinical identification of *Bacteroides* species.

In conclusion, there can no longer be much doubt that the bacteroides are specifically related to the flavobacteria-cytophagae. The relationship can be seen in overall sequence similarity in 16S rRNA, in specific sequence similarity (i.e., derived characters), and in common unique secondary structural details of the molecule. Although a phenotypic clustering of bacteroides and flavobacteria is not now recognized, two phenotypic characteristics common to these genera are consistent with such a relationship. *Bacteroides* species are unique among bacteria in containing sphingolipids as a major membrane component (15, 18). However, at least one species of *Flavobacterium*, *F. multivorum* (related to the other members of the genus through DNA-DNA hybridization measurements), contains sphingophospholipids as the main cellular lipid (7, 20). The bacteroides and flavobacteria also share the property of resistance to several aminoglycoside antibiotics (7, 18). In the case of the bacteroides, this resistance is thought to reflect failure to transport these compounds (3). An unusual (if not unique) common feature of the two 16S rRNAs is that

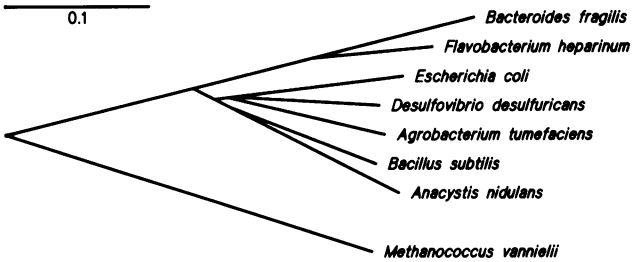

FIG. 2. Phylogenetic tree for the sequences of Fig. 1 derived from the percent similarities of Table 1 by the procedure of McCarroll et al. (13). The bar corresponds to 0.1 mutational events per sequence position. The root of the tree is arbitrary.

neither appears to possess a 5′-terminal phosphate residue. This lack can be deduced from the fact that there exists in the rRNA catalog in each case a large oligonucleotide (with no 5′ phosphate; ACUUUUACAAUG and UUUACAAUG in *B. fragilis* and *F. heparinum*, respectively [16]) that is a cleavage product of a predicted larger (T1 RNase) oligonucleotide in the gene sequence at the 5′ end of the molecule. These sequences are not shown in Fig. 1 (16; unpublished results).

An interesting type of flexibacterium, a strictly anaerobic flexible rod, which has a phenotype that in ways is intermediate between the phenotypes of the bacteroides and flavobacteria-cytophagae, has been isolated by K. O. Stetter (16). The organism holds an intermediate position between the two groups phylogenetically as well (16).

ACKNOWLEDGMENTS

This work has been supported by grants from the National Science Foundation and the National Aeronautics and Space Administration.

We thank G. J. Olsen for producing the phylogenetic tree of Fig. 2.

LITERATURE CITED

1. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
2. **Brosius, J., J. L. Palmer, J. P. Kennedy, and H. F. Noller.** 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. Proc. Natl. Acad. Sci. USA **75**:4801–4805.
3. **Bryan, L. E., S. K. Kowand, and H. M. Van Den Elzen.** 1979. Mechanisms of aminoglycoside antibiotic resistance in anaerobic bacteria: *Clostridium perfringens* and *Bacteroides fragilis*. Antimicrob. Agents Chemother. **15**:7–13.
4. **Fox, G. E., K. J. Pechman, and C. R. Woese.** 1977. Comparative cataloging of 16*S* ribosomal ribonucleic acid: molecular approach to procaryotic systematics. Int. J. Syst. Bacteriol. **27**:44–57.
5. **Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, and C. R. Woese.** 1980. The phylogeny of prokaryotes. Science **209**: 457–463.
6. **Holdeman, L. V., R. W. Kelley, and W. E. C. Moore.** 1984. Family I. *Bacteroidaceae* Pribram 1933, 10^{AL}, p. 602–631. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. The Williams & Wilkins Co., Baltimore.
7. **Holmes, B., R. J. Owen, and T. A. McKeekin.** 1984. Genus *Flavobacterium* Bergey, Harrison, Breed, Hammer and Huntoon 1923, 97^{AL}, p. 353–361. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. The Williams & Wilkins Co., Baltimore.

8. **Iwami, M., A. Muto, F. Yamao, and S. Osawa.** 1984. Nucleotide sequence of the *rrnB* 16S ribosomal RNA gene from *Mycoplasma capricolum*. Mol. Gen. Genet. **196**:317–322.
9. **Jarsch, M., and A. Bock.** 1985. Sequence of the 16S ribosomal RNA gene from *Methanococcus vannielii*. Syst. Appl. Microbiol. **6**:54–59.
10. **Leder, P., D. Tiemeier, and L. Enquist.** 1977. EK2 derivatives of bacteriophage lambda useful in the cloning of higher organisms: the lambda gtWES system. Science **196**:175–177.
11. **Loenen, W. A. M., and W. J. Brammar.** 1980. A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene **20**:249–259.
12. **Marmur, J.** 1961. A procedure for the isolation of deoxyribonucleic acid from micro-organisms. J. Mol. Biol. **3**:208–218.
13. **McCarroll, R., G. J. Olsen, Y. D. Stahl, C. R. Woese, and M. L. Sogin.** 1983. Nucleotide sequence of *Dictyostelium discoideum* small-subunit ribosomal ribonucleic acid inferred from the gene sequence: evolutionary implications. Biochemistry **22**:5858–5868.
14. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
15. **Mills, D. R., and F. R. Kramer.** 1979. Structure-independent nucleotide sequence analysis. Proc. Natl. Acad. Sci. USA **76**:2232–2235.
16. **Paster, B. J., W. Ludwig, W. G. Weisburg, E. Stackebrandt, R. B. Hespell, C. M. Hahn, H. Reichenbach, K. O. Stetter, and C. R. Woese.** 1985. A phylogenetic grouping of the bacteroides, cytophagas and certain flavobacteria. Syst. Appl. Microbiol. **6**:34–42.
17. **Paster, B. J., E. Stackebrandt, R. B. Hespell, C. M. Hahn, and C. R. Woese.** 1984. The phylogeny of the spirochetes. Syst. Appl. Microbiol. **5**:337–351.
18. **Salyers, A. A.** 1984. Bacteroides of the human lower intestinal tract. Annu. Rev. Microbiol. **38**:293–313.
19. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
20. **Shewan, J. M., and T. A. McMeekin.** 1983. Taxonomy (and ecology) of *Flavobacterium* and related genera. Annu. Rev. Microbiol. **37**:233–252.
21. **Tomioka, N., and M. Sugiura.** 1983. The complete nucleotide sequence of a 16S ribosomal RNA gene from a blue-green alga, *Anacystis nidulans*. Mol. Gen. Genet. **191**:46–50.
21a. **Woese, C. R., B. A. Debrunner-Vossbrinck, and H. Oyaizu.** 1985. Gram-positive bacteria: possible photosynthetic ancestry. Science **229**:762–765.
22. **Woese, C. R., and G. E. Fox.** 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA **74**:5088–5090.
23. **Woese, C. R., R. Gutell, R. Gupta, and H. F. Noller.** 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiol. Rev. **47**:621–669.
24. **Woese, C. R., J. Maniloff, and L. B. Zablen.** 1980. Phylogenetic analysis of the mycoplasmas. Proc. Natl. Acad. Sci. USA **77**:494–498.
25. **Woese, C. R., M. Sogin, D. A. Stahl, B. J. Lewis, and L. Bonen.** 1976. A comparison of the 16S ribosomal RNAs from mesophilic and thermophilic bacilli: some modifications in the Sanger method for RNA sequencing. J. Mol. Evol. **7**:197–213.
26. **Woese, C. R., E. Stackebrandt, T. J. Macke, and G. E. Fox.** 1985. A phylogenetic definition of the major eubacterial taxa. Syst. Appl. Microbiol. **6**:133–142.
27. **Woese, C. R., E. Stackebrandt, W. G. Weisburg, B. J. Paster, M. R. Madigan, V. J. Fowler, C. M. Hahn, P. Blanz, R. Gupta, K. H. Nealson, and G. E. Fox.** 1984. The phylogeny of purple bacteria: the alpha subdivision. Syst. Appl. Microbiol. **5**:315–326.
28. **Woese, C. R., W. G. Weisburg, C. M. Hahn, B. J. Paster, L. B. Zablen, B. J. Lewis, T. J. Macke, W. Ludwig, and E. Stackebrandt.** 1985. The phylogeny of purple bacteria: the gamma subdivision. Syst. Appl. Microbiol. **6**:25–33.
29. **Woese, C. R., W. G. Weisburg, B. J. Paster, C. M. Hahn, R. S. Tanner, N. R. Krieg, H.-P. Koops, H. Harms, and E. Stackebrandt.** 1984. The phylogeny of purple bacteria: the beta subdivision. Syst. Appl. Microbiol. **5**:327–336.
30. **Yang, D., Y. Oyaizu, H. Oyaizu, G. J. Olsen, and C. R. Woese.** 1985. Mitochondrial origins. Proc. Natl. Acad. Sci. USA **82**:4450–4455.