

Discordant 16S and 23S rRNA Gene Phylogenies for the Genus *Helicobacter*: Implications for Phylogenetic Inference and Systematics

Floyd E. Dewhirst,^{1,2†} Zeli Shen,^{3†} Michael S. Scimeca,³ Lauren N. Stokes,¹
Tahani Boumenna,¹ Tsute Chen,¹ Bruce J. Paster,^{1,2}
and James G. Fox^{3*}

Department of Molecular Genetics, The Forsyth Institute, Boston, Massachusetts 02115¹; Department of Oral and Developmental Biology, Harvard School of Dental Medicine, Boston, Massachusetts 02115²; and Division of Comparative Medicine, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139³

Received 26 January 2005/Accepted 6 June 2005

Analysis of 16S rRNA gene sequences has become the primary method for determining prokaryotic phylogeny. Phylogeny is currently the basis for prokaryotic systematics. Therefore, the validity of 16S rRNA gene-based phylogenetic analyses is of fundamental importance for prokaryotic systematics. Discrepancies between 16S rRNA gene analyses and DNA-DNA hybridization and phenotypic analyses have been noted in the genus *Helicobacter*. To clarify these discrepancies, we sequenced the 23S rRNA genes for 55 *Helicobacter* strains representing 41 taxa (>2,700 bases per sequence). Phylogenetic-tree construction using neighbor-joining, parsimony, and maximum likelihood methods for 23S rRNA gene sequence data yielded stable trees which were consistent with other phenotypic and genotypic methods. The 16S rRNA gene sequence-derived trees were discordant with the 23S rRNA gene trees and other data. Discordant 16S rRNA gene sequence data for the *Helicobacter* are consistent with the horizontal transfer of 16S rRNA gene fragments and the creation of mosaic molecules with loss of phylogenetic information. These results suggest that taxonomic decisions must be supported by other phylogenetically informative macromolecules, such as the 23S rRNA gene, when 16S rRNA gene-derived phylogeny is discordant with other credible phenotypic and genotypic methods. This study found *Wolinella succinogenes* to branch with the unsheathed-flagellum cluster of *Helicobacter* by 23S rRNA gene analyses and whole-genome comparisons. This study also found intervening sequences (IVSs) in the 23S rRNA genes of strains of 12 *Helicobacter* species. IVSs were found in helices 10, 25, and 45, as well as between helices 31' and 27'. Simultaneous insertion of IVSs at three sites was found in *H. mesocricetorum*.

Helicobacter pylori was first identified in 1982 (39) as an infectious cause of chronic active gastritis and since then has been associated with peptic ulcers, gastric adenocarcinoma, and mucosa-associated lymphoid tissue lymphoma in humans (38). The genus *Helicobacter* now includes at least 26 formally named species (60). The tropism of the genus *Helicobacter* ranges from the stomach, cecum, and colon to the liver or genital tract of mammals and birds. *Helicobacter* spp. comprise a diverse group of potentially emerging pathogens in multiple species: humans, nonhuman primates, cats, dogs, ferrets, swine, sheep, cattle, wild birds, chickens, mice, rats, hamsters, gerbils, and a variety of wildlife, including dolphins, whales, and seals (60).

Over the past 15 years, microbiology has undergone a momentous shift from a determinative taxonomy to one based on phylogeny (61). Our current understanding of prokaryotic phylogeny rests largely on 16S rRNA gene sequence analysis, which has become the basis for reorganizing *Bergey's Manual of Systematic Bacteriology* and *The Prokaryotes* (3, 51). 16S rRNA

gene analysis has proved useful for defining prokaryotic relationships from the species to phylum levels. Cases of very closely related strains or species have highlighted the need to use other, more rapidly changing molecules for some genera (36). Of concern for the validity of prokaryotic systematics are cases where 16S rRNA gene-derived phylogeny seems to conflict with other credible molecular or phenotypic data. Within the genus *Helicobacter*, there have been several reported conflicts. For example, *Helicobacter nemestrinae* (ATCC 49396^T), was once believed to be a novel gastric *Helicobacter* isolated from a pigtailed macaque. However, it was later recognized to be a strain of *H. pylori* by correction of the 16S rRNA gene sequence and analysis of seven housekeeping and two flagellin genes, which clustered together with sequences from 20 or more *H. pylori* isolates from diverse sources (53). Vandamme et al. (58) have recently published data that indicated that "*Helicobacter westmeadii*" and *Helicobacter* sp. strain Mainz (27) are both strains of *Helicobacter cinaedi*. The 16S rRNA gene sequence for the Mainz strain differs by 4.3% from that of the type strain. The authors stated that similar within-species differences in 16S rRNA gene sequences have been reported in *Campylobacter hyointestinalis*, another species that belongs to the epsilon subdivision of the division *Proteobacteria* (25). The purpose of this study was to obtain extensive sequence data for a second phylogenetically informative molecule in order to

* Corresponding author. Mailing address: Division of Comparative Medicine, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Bldg. 16, Rm. 825C, Cambridge, MA 02139. Phone: (617) 253-1735. Fax: (617) 258-5708. E-mail: jgfox@mit.edu.

† These authors contributed equally to this work.

understand the source of the discrepancy between 16S rRNA gene analyses and those from other phenotypic and genotypic methods. The nearly complete 23S rRNA gene sequences (>94%) for 55 helicobacter strains representing 41 helicobacter taxa were determined.

MATERIALS AND METHODS

Bacterial strains. The strains sequenced and their sources of isolation are presented in Table 1. All strains were grown in an 80% N₂, 10% CO₂, 10% H₂ atmosphere at 37°C on Columbia blood agar plates.

DNA extraction. The High Pure PCR Template Preparation Kit (Roche Molecular Biochemicals, Indianapolis, IN) was used to extract DNA from bacterial pellets. Bacteria were scraped from plates and washed with 1 ml phosphate-buffered saline. After centrifugation, the cells were resuspended in 200 µl phosphate-buffered saline, and lysis was initiated by adding 15 µl of 10-mg/ml lysozyme. Following incubation at 37°C for 15 min, 40 µl of proteinase K (20 mg/ml) and 200 µl of binding buffer were added and incubated at 72°C for 10 min. The reaction mixture was loaded onto the filter tube and washed twice with wash buffer, and the DNA was eluted with 200 µl of Tris-EDTA.

PCR amplification of rRNA genes. The 23S rRNA genes were amplified using PCR primers O68 and M89, complementary to highly conserved regions on 23S rRNA genes (Table 2). The reaction mixture (50 µl) contained 1× polymerase buffer, a 0.4 µM concentration of each of the two primers, a 200 µM concentration of each deoxyribonucleotide, and 2.5 U Platinum or AccuPrime Taq DNA polymerase (Invitrogen, Carlsbad, California). The following stepdown conditions were used for amplification: activation of polymerase at 94°C for 2 min; 8 cycles of denaturation at 94°C for 45 s, annealing at 62°C to 55°C for 45 s (dropping 1°C per cycle), and elongation at 72°C for 3 min; and 22 cycles of 94°C for 45 s, 55°C for 45 s, and 72°C for 3 min, followed by a final elongation step of 7 min at 72°C. A 15-µl portion of the sample was then electrophoresed through a 1% agarose gel to check the size and concentration of the amplicon.

16S rRNA genes were amplified using primers F24 and F25 as described previously (13).

Sequencing. PCR-amplified fragments were purified with a QIAGEN PCR purification kit (QIAGEN Inc., Valencia, CA) for direct sequencing using a BigDye terminator cycle sequencing kit on an ABI 310 or ABI 3100 Genetic Analyser (Applied Biosystems, Foster City, CA). Quarter-dye chemistry was used following the manufacturer's instructions. 23S rRNA gene cycle sequencing was performed using 25 cycles as follows: denaturing at 94°C for 10 s and annealing and extension at 60°C for 4 min. Because intervening sequences (IVSs) of up to 415 bases occurred at four positions in helicobacter sequences, the full set of 20 primers was routinely used to ensure complete sequencing across any potential IVS. A map showing primers and potential IVS positions is given in Fig. 1.

Data analysis. RNA sequences were aligned manually based on secondary structure in our database (43). Base positions (columns) which did not contain specific base information (ambiguous bases or indels) for at least 52 of the 56 sequences were removed. The aligned sequences were exported in Phylip format for analysis using the Phylip set of programs, version 3.6b (19). Similarity matrices were constructed using the program Dnadist. Distances were corrected for multiple base changes using the method of Jukes and Cantor (29). Neighbor-joining (NJ) trees were constructed using the program Neighbor (47). Parsimony trees were constructed using the program Dnapars with the order of sequence additions jumbled 25 times. Maximum likelihood (ML) trees were constructed using the program Dnaml with Speedier off and Global on. Bootstrapping for neighbor joining and parsimony analyses was performed using the program Seqboot to generate 100 resampled sets of sequences. Consensus trees from bootstrapping analyses were generated using the program Consense. Trees for publication were prepared from the Phylip tree files using TreeView version 1.6.6 (42).

Whole-genome comparisons. The genome for *Helicobacter hepaticus* (NC_004917) was downloaded from GenBank. Each protein was subjected to a BLASTP search. The ranked order of similarity of *H. hepaticus* to *H. pylori* (either genome), *Wolinella succinogenes*, and *Campylobacter jejuni* was determined for those genes present in all four species. The mean similarity score (bits) for the set of shared genes was determined between *H. hepaticus* and the other three species.

Nucleotide sequence accession numbers. The GenBank accession numbers and culture collection accession numbers for the species and strains examined in this study are listed in Table 1.

RESULTS

23S rRNA gene sequences. Essentially complete 23S rRNA gene sequences (>2,700 bases) were determined for 55 *Helicobacter* strains representing 41 species or unnamed taxa. Each strain was sequenced on both strands using 16 to 20 primers listed in Table 2. For phylogenetic analysis, the sequences from the complete genome projects for *H. pylori* ATCC 26695 and J99 and *Campylobacter jejuni* were also included.

16S rRNA gene sequences. Some of the early helicobacter 16S rRNA sequences from our laboratory had been determined using direct reverse transcriptase sequencing of rRNA and contained numerous "Ns" or were missing intervening sequences. Fifteen previously deposited sequences were redone, and the GenBank entries were updated (noted in Table 1). Five helicobacter sequences were deposited in GenBank to give longer or more accurate 16S rRNA gene sequences for entries previously submitted by other laboratories (noted in Table 1). We did not have the 16S rRNA gene sequence for *H. canadensis* NLEP 16767 in time to include it in the 16S rRNA gene phylogenetic analyses; thus, the 16S rRNA gene analyses are based on 57 rather than 58 taxa.

NJ analyses. NJ trees based on 23S rRNA and 16S rRNA gene sequence analyses are shown in Fig. 2 and Fig. 3, respectively. The strain numbers, GenBank accession numbers, and species from which strains were isolated are included in the trees.

In the 23S rRNA gene tree, there were eight major clusters. Cluster 1 included the gastric helicobacters *H. pylori*, *H. acinonychis*, *H. ceterum*, *H. bizzozeronii*, and *H. felis*. A partial sequence for *H. salomonis* (not deposited in GenBank or included in Fig. 1) placed it adjacent to *H. bizzozeronii* and *H. felis*. Cluster 2 included all helicobacters without sheathed flagella: *H. rodentium*, *H. ganmani*, *H. winghamensis*, *H. canadensis*, *H. pullorum*, and *H. mesocricetorum*. Also included in cluster 2 was the rumen organism *Wolinella succinogenes*. Cluster 3 contained *H. pametensis*, other bird strains, *H. cholecystus*, and *H. mustelae*. Clusters 4 to 7 were comprised of enterohepatic helicobacters. Cluster 8 was comprised of taxa with flexispira morphology (fusiform-shaped organisms with spiral periplasmic fibers and bipolar tufts of sheathed flagella), with the exception of *H. cinaedi* strains.

Eight major clusters were also identified in the 16S rRNA gene NJ tree. As in the 23S rRNA gene tree, cluster 1 was comprised of gastric strains, cluster 2 was comprised of helicobacters with unsheathed flagella (plus flexispira taxon 1), and cluster 3 was comprised of *H. pametensis*, *H. cholecystus*, *H. mustelae*, and helicobacter bird isolates. In the tree shown in Fig. 3, cluster 2 is broken into two adjacent paraphyletic clusters. Cluster C contained the same strains as cluster 6 in the 23S tree with the addition of the seal strain MIT 01 5529A. The remaining four clusters (B through E) are a random assortment of strains from 23S rRNA gene clusters 4, 5, 7, and 8.

The consensus NJ tree produced by bootstrap analysis of 100 resamplings of the 23S rRNA gene sequence data had branching identical to that of the simple neighbor-joining tree (bootstrapping values are shown at the nodes in Fig. 2). By contrast, the consensus NJ tree produced by bootstrap analysis of 100 resamplings of the 16S rRNA gene sequences produced a tree with several rearrangements of deep branching (Fig. 3). Nodes

TABLE 1. Accession numbers for strains examined in this study

| Species | Strain sequenced ^a | Culture collection ^b | Host | GenBank accession no. | | 23S rRNA gene sequenced (bp) |
|---|--|--|--------------------|-----------------------|-----------------------|------------------------------|
| | | | | 16S rRNA gene | 23S rRNA gene | |
| <i>Campylobacter jejuni</i> | NCTC 11168 | | Human | AL139074 ^c | AL139074 ^c | Genome |
| <i>Helicobacter acinonychis</i> | Eaton 90-119 ^T | ATCC 51101 ^T | Cheetah | M88148 ^d | AY596214 | 2,715 |
| <i>Helicobacter aurati</i> | MIT 97-5075 ^T | ATCC BAA-1 ^T | Hamster | AF297868 | AY596241 | 2,711 |
| <i>Helicobacter bilis</i> | MIT Hb1 ^T | ATCC 51630 ^T | Mouse | U18766 ^d | AY596255 | 2,875 |
| <i>Helicobacter bizzozeronii</i> | CCUG 35045 ^T | ATCC 700030 ^T | Dog | Y09404 ^c | AY596215 | 2,715 |
| <i>Helicobacter canadensis</i> | NLEP 16143 ^T | ATCC 700968 ^T | Human | AF262037 | AY596220 | 2,713 |
| <i>Helicobacter canadensis</i> | NLEP 16767 | ATCC 700969 | Human | | AY596221 | 2,713 |
| <i>Helicobacter canis</i> | 23S, ATCC 51401 ^T 16S, NCTC 12739 ^T | | Dog | AY631945 ^c | AY596234 | 2,710 |
| <i>Helicobacter canis</i> | ATCC 51402 | NCTC 12741 | Dog | AY631946 | AY596235 | 2,710 |
| <i>Helicobacter cetorum</i> | MIT 99-5656 ^T | ATCC BAA-540 ^T | Dolphin | AF292378 | AY596262 | 2,717 |
| <i>Helicobacter cholecystus</i> | ATCC 700242 ^T | ATCC 700242 ^T | Hamster | AY686606 ^c | AY596229 | 2,807 |
| <i>Helicobacter cinaedi</i> | CCUG 18818 ^T | CCUG 18818 ^T | Human | M88150 ^d | AY596254 | 2,718 |
| <i>Helicobacter cinaedi</i> | MIT 01-5002 | | Monkey | AY631947 | AY596261 | 2,718 |
| <i>Helicobacter felis</i> | Lee CS1 ^T | ATCC 49179 | Cat | M57398 ^d | AY596216 | 2,715 |
| <i>Helicobacter felis</i> | Lee CS3 | | Cat | AY631948 | AY596217 | 2,714 |
| <i>Helicobacter felis</i> | Lee CS5 | | Cat | AY631949 | AY596219 | 2,716 |
| <i>Helicobacter felis</i> | Lee DS2 | | Dog | AY686607 ^c | AY596218 | 2,717 |
| <i>Helicobacter fennelliae</i> | CCUG 18820 ^T | ATCC 35683 ^T | Human | M88154 ^d | AY596237 | 3,075 |
| <i>Helicobacter flexispira</i> taxon 1 | ATCC 43968 | | Pig | U96300 | AY596249 | 2,712 |
| <i>Helicobacter flexispira</i> taxon 2 | ATCC49314 | | Sheep | AF225546 | AY596257 | 3,052 |
| <i>Helicobacter flexispira</i> taxon 3 | ATCC 49320 | | Pig | AF225547 | AY596256 | 2,876 |
| <i>Helicobacter flexispira</i> taxon 4 | ATCC 49310 | | Sheep | AF225548 | AY596248 | 2,711 |
| <i>Helicobacter flexispira</i> taxon 5 | ATCC 43966 | | Sheep | M88137 | AY596250 | 2,712 |
| <i>Helicobacter flexispira</i> taxon 7 | Eaton 1302 | | Dog | U51874 ^d | AY596247 | 2,712 |
| <i>Helicobacter flexispira</i> taxon 8 | Schauer DBS59 | | Mouse | AY686608 ^c | AY596253 | 2,808 |
| <i>Helicobacter flexispira</i> taxon 8 | ATCC 49309 | | Human | AY631950 | AY686610 | 2,712 |
| <i>Helicobacter flexispira</i> taxon 8 | ATCC 43879 | | Human | M88138 ^d | AY596252 | 2,713 |
| <i>Helicobacter ganmani</i> | MIT 99-5102 | | Mouse | AY631951 | AY596224 | 2,714 |
| <i>Helicobacter hepaticus</i> | MIT Hh2 ^T | ATCC 51448 ^T | Mouse | U07574 ^d | AY596243 | 2,887 |
| <i>Helicobacter hepaticus</i> | MIT 96-1809 | | Mouse | AY631952 | AY596244 | 2,807 |
| <i>Helicobacter hepaticus</i> | MIT 96-284 | | Mouse | AY631953 | AY596242 | 2,712 |
| <i>Helicobacter marmotae</i> | MIT 98-6070 ^T | ATCC BAA-546 ^T | Woodchuck | AF333341 | AY596245 | 2,716 |
| <i>Helicobacter mesocricetorum</i> | ATCC 700932 ^T | | Hamster | AF072471 | AY686611 | 3,641 |
| <i>Helicobacter muridarum</i> | Lee ST1 ^T | ATCC 49282 ^T | Rat | M80205 ^d | AY596238 | 3,084 |
| <i>Helicobacter mustelae</i> | MIT 89-977 | | Ferret | AY631954 | AY596233 | 2,795 |
| <i>Helicobacter mastomyrinus</i> | MIT 94-022 | | Mouse | AF225550 ^d | AY596240 | 2,810 |
| <i>Helicobacter mastomyrinus</i> | MIT 97-5577 ^T | | Mastomys | AY631955 | AY596265 | 2,683 |
| <i>Helicobacter pametensis</i> | Seymour B9 ^T | CCUG 29255 ^T ATCC 51478 ^T | Tern | M88147 ^d | AY596230 | 2,745 |
| <i>Helicobacter pullorum</i> | ATCC 51801 ^T | NCTC 12824 ^T | Chicken | AY631956 ^c | AY596222 | 2,713 |
| <i>Helicobacter pullorum</i> | ATCC 51864 | NCTC 12827 | Human | L36144 ^c | AY596223 | 2,713 |
| <i>Helicobacter pylori</i> | ATCC 26695 | | Human | AE000644 ^c | AE000644 ^c | 2,843 |
| <i>Helicobacter pylori</i> | J99 | | Human | AE001553 ^c | AE001553 ^c | 2,886 |
| <i>Helicobacter rodentium</i> | MIT 95-1707 ^T | ATCC 700285 ^T | Mouse | U96296 | AY596225 | 3,086 |
| <i>Helicobacter rodentium</i> | MIT 95-2160 | ATCC 700286 | Mouse | U96297 | AY596227 | 3,086 |
| <i>Helicobacter rodentium</i> | MIT 95-2178 | | Mouse | AY631957 | AY596226 | 3,085 |
| <i>Helicobacter</i> sp. strain Cotton-top tamarin | MIT 97-6194-5 | | Cotton-top tamarin | AF107494 | AY596246 | 2,970 |
| <i>Helicobacter</i> sp. strain Seal-1 | MIT 01-5529-A | | Seal | AY203898 | AY596258 | 2,713 |
| <i>Helicobacter</i> sp. strain Seal-2 | MIT 01-5529-B | | Seal | AY203899 | AY596259 | 3,103 |
| <i>Helicobacter</i> sp. strain Rhesus-1 | MIT 99-5504 | | Rhesus monkey | AF333339 | AY596263 | 2,714 |
| <i>Helicobacter</i> sp. strain Rhesus-2 | MIT 99-5507 | | Rhesus monkey | AF333340 | AY596264 | 2,708 |
| <i>Helicobacter</i> sp. strain Mainz | CCUG 33804 | | Human | AF207739 ^c | AY596260 | 2,559 |
| <i>Helicobacter</i> sp. strain CLO3 | CCUG 14564 | | Human | M88151 ^d | AY596236 | 2,722 |
| <i>Helicobacter</i> sp. strain Bird-B | Seymour B10 | ATCC 51480 | Tern | M88139 ^d | AY596232 | 2,707 |
| <i>Helicobacter</i> sp. strain Bird-C | Seymour B52 | ATCC 51482 | House sparrow | M88144 ^d | AY596231 | 2,707 |
| <i>Helicobacter trogonum</i> | MIT 95-5368 | ATCC 700117 | Rat | AY686609 | AY596251 | 2,712 |
| <i>Helicobacter typhlonius</i> | MIT 97-6810 | | Mouse | AF127912 | AY596239 | 2,712 |
| <i>Helicobacter winghamensis</i> | ATCC BAA-430 | | Human | AF246984 | AY686612 | 2,712 |
| <i>Wolinella succinogenes</i> | Tanner 602W ^T | ATCC 29543 ^T | Cow | M88159 ^d | AY596228 | 2,881 |

^a Strains are from the following investigators: Eaton, Kathryn Eaton, University of Michigan; Lee, Adrian Lee, University of New South Wales, Australia; Seymour, Charles Seymour, deceased; Schauer, David Schauer, Massachusetts Institute of Technology; and Tanner, Anne Tanner, The Forsyth Institute.

^b ATCC, American Type Culture Collection; CCUG, Culture Collection University Göteborg; MIT, Massachusetts Institute of Technology; NLEP, National Laboratory for Enteric Pathogens, Winnipeg, Canada.

^c This sequence was obtained from GenBank and was determined by investigators listed in the GenBank entry.

^d Genbank entry from our laboratory has been revised, in some cases extensively.

^e New, more accurate, or longer sequence for a strain previously deposited in GenBank by other investigators.

TABLE 2. 23S rRNA gene PCR and sequencing primers used in this study

| Primer designation | Position | Sequence (5'-3') |
|------------------------------------|-------------------|----------------------------|
| Final set^a | | |
| O68 | 44–60 Forward | AGG CGA TGA AGG ACG TA |
| V67 | 46–63 Forward | GCG ATG AAG GAC GTA CTA |
| M90 | 241–255 Forward | AGT AGY GGC GAG CGA A |
| V68 | 465–483 Forward | GAG GGA AAG GTG AAA AGA A |
| O20 | 571–587 Forward | CAT AAT GAT CCT GCG AGT T |
| M92 | 975–990 Forward | AGG GRA ACA RCC CAG A |
| M93 | 1090–1104 Forward | WGC GTA AYA GCT CAC |
| M94 | 1608–1623 Forward | AAA CCG WCA CAG GTR G |
| V63 | 2012–2032 Forward | GAA ATT GTA GTG GAG GTG AA |
| V65 | 2375–2393 Forward | GAA AGT CGG TCA TAG TGA T |
| M96 | 2498–2512 Forward | CCT CGA TGT CGR CTC |
| V66 | 2580–2596 Forward | TGG GTT CAG AAC GTC GT |
| M83 | 241–256 Reverse | KTT CGC TCG CCR CAT C |
| O21 | 571–587 Reverse | AAC TCG CAG GAT CAT TAT G |
| O22 | 681–698 Reverse | GGC CAT GGA TAG ATC ACT |
| M85 | 1091–1105 Reverse | AGT RAG CTR TTA CGC |
| V62 | 1331–1348 Reverse | CCC GAC TAA CCC TAC GAT |
| M86 | 1608–1623 Reverse | CYA CCT GTG WCR GTT T |
| V64 | 2012–2032 Reverse | TTC ACC TCC ACT ACA ATT TC |
| M88 | 2498–2512 Reverse | GAG YCG ACA TCG AGG |
| P46 | 2652–2668 Reverse | CGG TCC TCT CGT ACT AG |
| M89 | 2744–2659 Reverse | CTT AGA TGC YTT CAG C |
| Preliminary set^b | | |
| ZSF4 | 940–958 Forward | GAG TCA GGC GGT GGG TGA T |
| ZSF3 | 1817–1836 Forward | GTA TAA GGT GTG ACG CCT GC |
| M95 | 2053–2069 Forward | GAC RGA AAG ACC CCR TG |
| ZSF5 | 2366–2384 Forward | AGC AGA GAC GAA AGT CGG T |
| ZSR7 | 454–473 Reverse | CTT TCC CTC ACG GTA CTT GT |
| ZSR6 | 799–818 Reverse | CGA AGA GAA CCA GCT ATC AC |
| M87 | 2053–2069 Reverse | CAY RGG GTC TTT CYG TC |

^a Final-set primers were designed based on alignment of approximately 20 nearly complete helicobacter 23S rRNA sequences. They contain some wobbles that are not necessary for helicobacters but which may make them useful for many epsilon *Proteobacteria*.

^b Initial-set primers were designed based on those described by Lane (31). These primers have mismatches with some or many helicobacter sequences and were not used once the final set of primers was designed.

marked with Xs in Fig. 3 indicate differences between the simple neighbor-joining tree and the consensus bootstrap neighbor-joining tree. Thus, the 16S rRNA gene neighbor-joining tree was not stable to bootstrap resampling. The mean bootstrap value for all nodes in the 23S rRNA gene tree was 82.7, whereas it was only 66.9 for the 16S rRNA gene tree.

Parsimony analyses. Consensus parsimony trees produced by bootstrap analysis of 100 resamplings of the 23S rRNA and 16S rRNA gene sequences are shown in Fig. 4. The 23S rRNA consensus parsimony tree retains the eight clusters of the neighbor-joining consensus tree with minor rearrangements within clusters and the backbone branching of the clusters: cluster 3 was split into two adjacent paraphyletic clusters, and cluster 4 became a monophyletic cluster. The 16S rRNA gene consensus parsimony tree had substantial rearrangements of the branching of the eight clusters and the fragmentation of cluster A with a two-sequence fragment inserted in the middle of cluster B relative to the 16S rRNA gene NJ tree. The mean bootstrap value for the 23S rRNA gene parsimony tree was 71.9, whereas it was only 52.3 for the 16S rRNA gene tree. The 23S rRNA gene tree had only 1 node bootstrap value under 30, whereas the 16S rRNA gene tree had 11. Several parsimony analyses were run using 25 jumbles of the input sequence order with different random-number seeds. The tree with the fewest steps for the 23S rRNA gene had 3,086 steps with seed 13. This tree had minor rearrangements of the eight clusters relative to the consensus parsimony tree. Three trees with 1,265 steps were the smallest found for the 16S rRNA gene, also with a seed of 13. These trees were substantially rearranged compared to the bootstrap consensus parsimony tree.

ML analyses. ML trees for 23S rRNA and 16S rRNA gene sequence analysis were constructed (data not shown). The 23S rRNA gene ML tree was similar to the NJ tree and contained each of the eight clusters. The tree shown had a ln of -22,152.8. The 16S rRNA gene ML tree did not retain the major branching pattern shown in the NJ or parsimony tree. The ln was -97,337 for the 16S rRNA gene tree.

Within- and between-species sequence differences. Several *Helicobacter* species had sequences determined for multiple strains. The within-species sequence variabilities for 23S rRNA gene sequences were as follows: *H. canadensis*, 0.1%, two strains; *H. canis*, 0.2%, two strains; *H. cinaedi*, 0.3%, three strains; *H. felis*, 2.0%, four strains; *H. hepaticus*, 1.8%, three strains; “*H. mastomyrinus*,” 1.2%, two strains; *H. pullorum*, 0.1%, two strains; *H. pylori*, 0.5%, two strains; *H. rodentium*, 0.1%, three strains; and *Helicobacter* sp. flexispira taxon 8, 0.2%, three strains. The interspecies variability was as low as 1.1% between the closely related species *H. canadensis* and *H. pullorum* and as high as 12% between *H. felis* and *H. aurati*. The helicobacter sequences differed by about 16% from that of *Campylobacter jejuni*. By 23S rRNA gene analysis, *Wolinella*

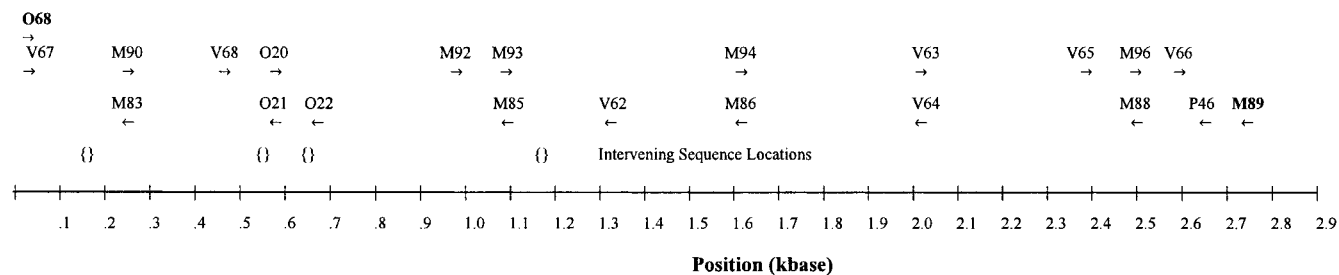


FIG. 1. Map of primer and IVS locations in helicobacter 23S rRNA gene. The positions of PCR and sequencing primers are shown by arrows with sequence designations above. The positions of four known intervening sequences are indicated by pairs of curly brackets.

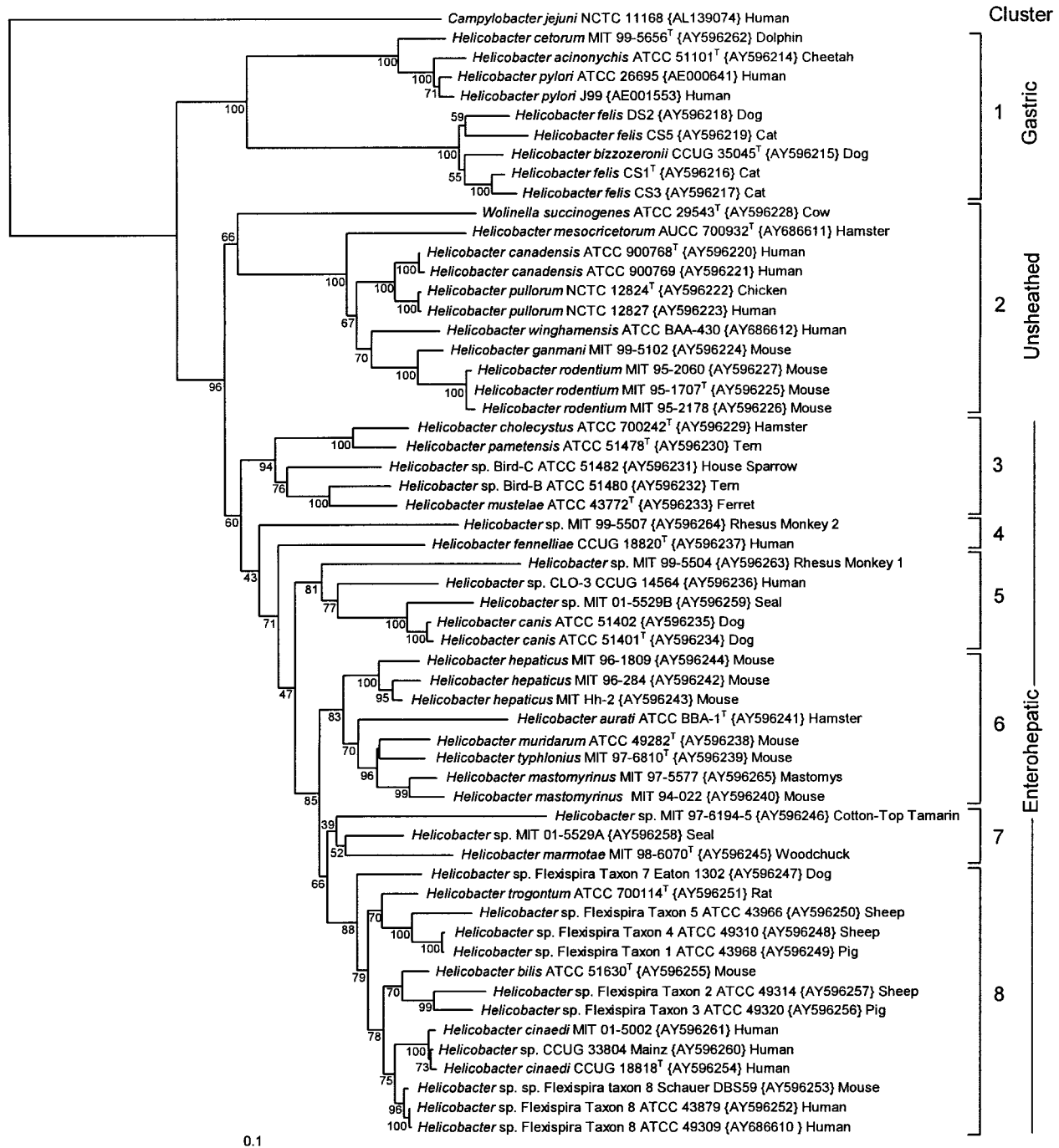


FIG. 2. Neighbor-joining tree based on comparison of 23S rRNA gene sequences. The tree shown is the simple neighbor-joining tree. All sequences are labeled by species, strain number, GenBank accession number in curly brackets, and the species from which it was isolated. The consensus bootstrap neighbor-joining tree had identical topology. The numbers immediately to the left of branches indicate the number of times out of 100 the clade was recovered by bootstrap resampling. The tree is divided into eight numbered clusters.

succinogenes does not represent a neighboring genus but rather branches within the helicobacters. Its sequence differs from other helicobacter sequences by 8 to 12%.

Intervening sequences in 23S rRNA genes. IVSs were found in 14 strains representing 12 species. The positions and lengths of these IVSs are shown in Table 3. IVSs were inserted into

four distinct locations in the 23S rRNA gene. Ten of the sequences demonstrated single IVS insertions, but *Helicobacter* sp. flexispira taxon 2 contained IVSs in two locations, and *H. mesocricetorum* contained IVSs in three locations. The IVSs for flexispira taxa 2 and 3 are identical and differ from that of *H. bilis* by only 6 bases (3.4%). The IVSs for three *H.*

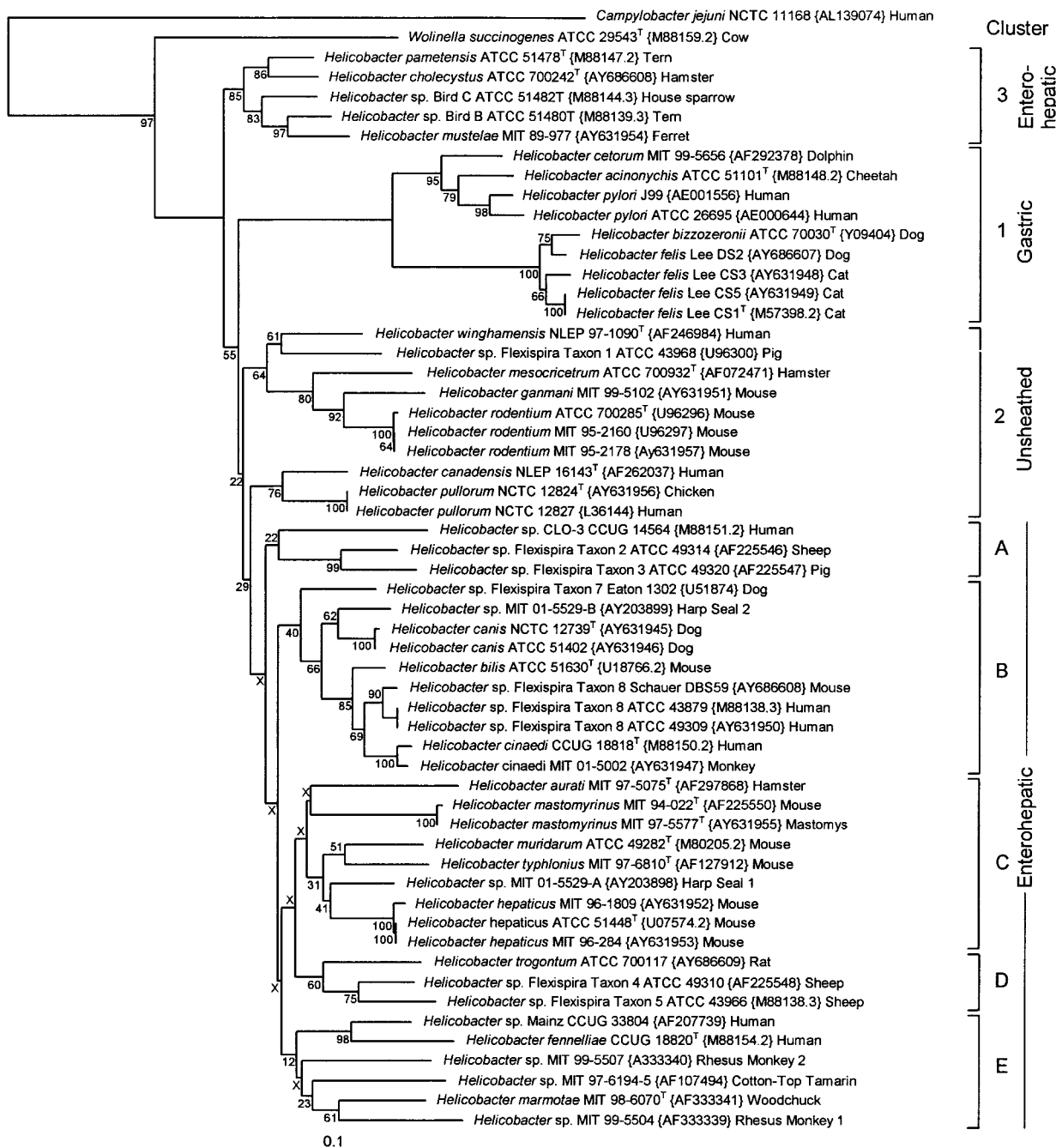


FIG. 3. Neighbor-joining tree based on comparison of 16S rRNA gene sequences. The tree shown is the simple neighbor-joining tree. All sequences are labeled as in Fig. 2. Nodes marked with Xs were not shared between the simple neighbor-joining tree and the consensus bootstrapping tree. The tree is divided into clusters. Clusters 1 to 3 correspond to those in the 23S rRNA gene tree, whereas clusters A to E differ from clusters 4 to 8 of the 23S rRNA gene tree.

rodentium strains differ from each other by 2 or fewer bases (or indels; 0.5%). While the IVS of *H. muridarum* is the same length as that from *H. rodentium*, it differs by approximately 240 bases and is only 37% similar (25% similarity is expected by chance).

Whole-genome comparisons. BLASTP (1) comparison of all *H. hepaticus* proteins with those in *H. pylori* (genomes for two

strains), *W. succinogenes*, and *C. jejuni* found that there were 870 proteins present in each of the four species with similarities for which *E* is less than e^{-5} . *E* is the probability due to chance that there is another alignment with a similarity greater than the given *S* score. The raw score *S* is based on the number of matches a sequence has to a sequence in the database, with a penalty subtracted for substitutions or gaps. For proteins,

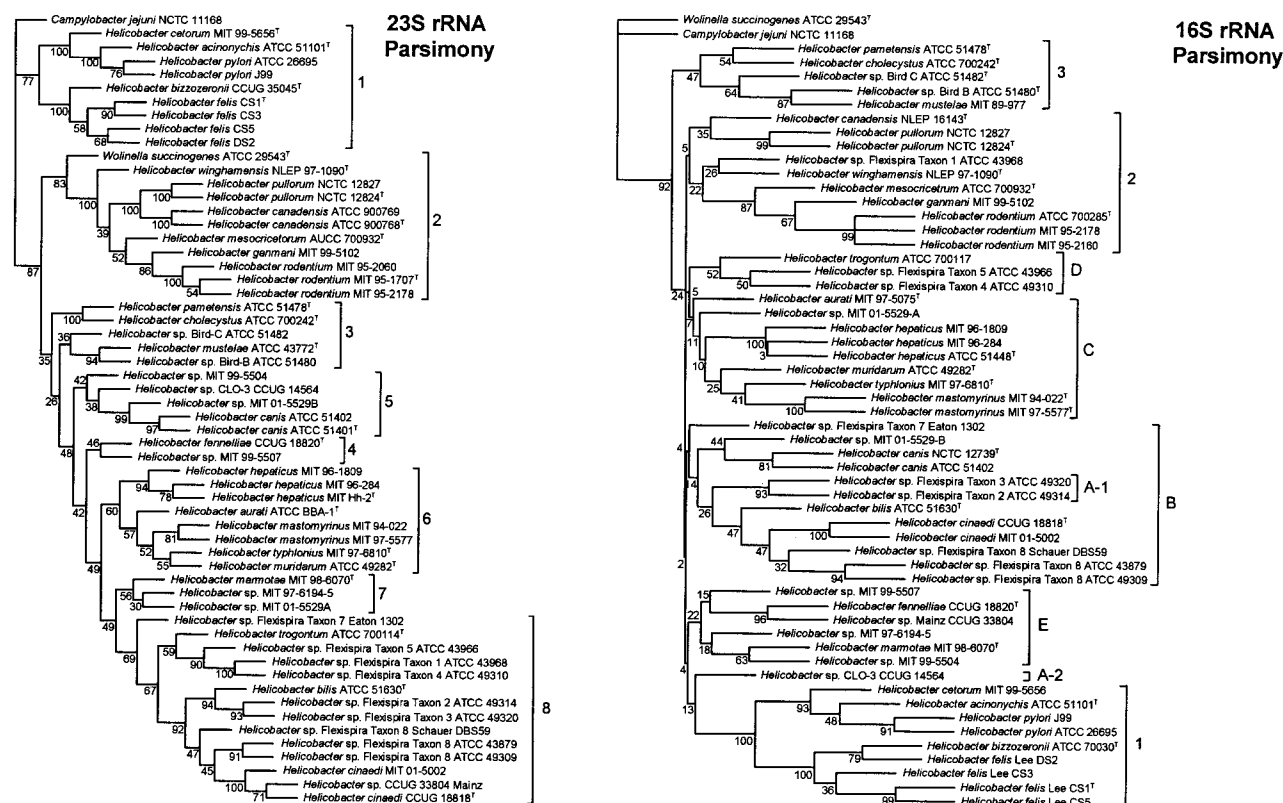


FIG. 4. Parsimony trees based on comparison of 23S rRNA gene and 16S rRNA gene sequences. The trees shown are consensus bootstrap trees based on 100 resampled parsimony trees. The numbers at the nodes are the number of times out of 100 resamplings that the node was present. Sequences are labeled as in Fig. 2 with GenBank numbers omitted. The clusters are labeled as in Fig. 2 and Fig. 3.

substitution penalties are based on the PAM (percent of accepted mutations) matrix or BLOSUM (blocks substitution matrix). As shown in Fig. 5, *H. hepaticus* is more closely related to *W. succinogenes* than to *H. pylori* and most distantly related

to *C. jejuni* based on mean *S* scores. For a four-taxon comparison, there are six possible tree topologies. Support for each of the six possible trees is shown in Fig. 6. Analyses of 870 protein trees support the 23S rRNA gene phylogeny and the branching

TABLE 3. IVSs in 23S rRNA gene sequences

| Taxon | Strain ^a | Length (bases) | | | |
|---|----------------------------|---|-------------------------------|------------------------------------|-----------------------------------|
| | | Helix ^b 10-10' [150-176] ^c (9-11) ^d | Helix 25-25' [545-549] (8) | Helices 31'-27' [652-655] (4-5) | Helix 45-45' [1171-1178] (2-8) |
| <i>Helicobacter</i> sp. strain Cotton-top tamarin | MIT 97-6194-5 | 269 | | | |
| <i>Helicobacter bilis</i> | MIT Hb1 ^T | 173 ^e | | | |
| <i>Helicobacter</i> sp. flexispira taxon 2 | ATCC 49314 | 173 ^e | 185 | | |
| <i>Helicobacter</i> sp. flexispira taxon 3 | ATCC 49320 | 173 ^e | | | |
| <i>Helicobacter mesocricetorum</i> | ATCC 700932 ^T | 237 | 370 | 344 | |
| <i>Helicobacter</i> sp. strain Seal-2 | MIT 01-5529-B | 415 | | | |
| <i>Helicobacter rodentium</i> | MIT 95-1707 ^T | 381 ^f | | | |
| <i>Helicobacter rodentium</i> | MIT 95-2160 | 381 ^f | | | |
| <i>Helicobacter rodentium</i> | MIT 95-2178 | 380 ^f | | | |
| <i>Helicobacter muridarum</i> | Lee ST1 ^T | 380 ^f | | | |
| <i>Helicobacter mustelae</i> | MIT R85-13-6P ^T | 93 | | | |
| <i>Helicobacter fennelliae</i> | CCUG 18820 ^T | | | 271 | |
| <i>Helicobacter pametensis</i> | Seymour B9 ^T | | | | 159 |
| <i>Helicobacter cholocystus</i> | ATCC 700242 ^T | | | | 99 |

^a Strain sequenced. Abbreviations for sources are the same as in Table 1, notes *a* and *b*.

^b Helices are numbered as in reference 5.

^c In square brackets the position of the insertion taken from the alignment numbered relative to *Escherichia coli*.

^d In parentheses, the length of the region in *Helicobacter* species without an IVS.

^e The IVSs for *Helicobacter* spp. flexispira taxa 2 and 3 are identical. The IVS for *H. bilis* differs from that of flexispira taxa 2 and 3 by 6 bases.

^f The IVS for *H. rodentium* strains differ by 2 or fewer bases. The IVS for *H. muridarum* has only 37% similarity to that of *H. rodentium* strains.

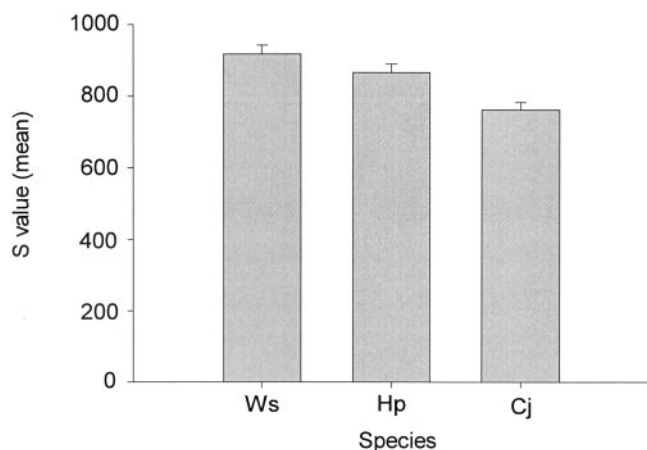


FIG. 5. Comparison of 870 *H. hepaticus* protein sequences to those from *H. pylori* (Hp), *W. succinogenes* (Ws), and *C. jejuni* (Cj). The sequences of 870 proteins shared by the four species were compared by BLASTP analysis. These 870 proteins were selected from all proteins found in these organisms based on reciprocal BLASTP matches with similarity values of *E* less than e^{-5} for all comparisons. The *S* value is a measure of similarity based on percent identity and length of comparison. The *E* value is the probability of obtaining an *S* value by chance based on comparison of the sequence to those in GenBank. The error bars represent standard deviations.

of *W. succinogenes* within the helicobacters. Also, analyses with more stringent cutoffs of *E* less than e^{-15} or e^{-25} give similar results with 769 or 657 proteins included in the four-way comparison.

DISCUSSION

In this study, we found that 16S rRNA and 23S rRNA gene phylogenetic trees are discordant. The 23S rRNA gene analyses are in agreement with other molecular and phenotypic methods as described below, suggesting that the 16S rRNA

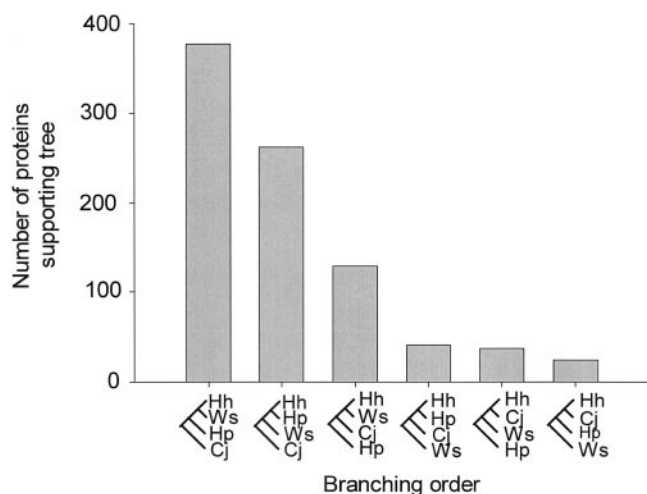


FIG. 6. Phylogeny of *H. hepaticus* (Hh), *H. pylori* (Hp), *W. succinogenes* (Ws), and *C. jejuni* (Cj) based on comparison of 870 protein trees. There are six possible trees for a four-taxon comparison. The number of proteins supporting each topology is shown above the respective tree.

gene sequence data analyses do not faithfully reflect phylogenetic relationships (22, 23, 40, 58). The findings support the majority of the criticisms by Vandamme et al. of the use of the 16S rRNA gene in helicobacter systematics (58). However, the implications for phylogenetic and taxonomic inference are not as straightforward as suggested in that article (58) or the study by Fox et al. (20). Discussion of our specific results will precede the discussion of the implications of our findings.

Helicobacter sp. strain Mainz (CCUG 33804) is the primary example cited for 16S rRNA gene sequence information giving discrepant phylogenetic information (58). The 16S rRNA gene sequence for the strain suggested that the Mainz strain is most closely related to *H. fennelliae*, whereas DNA-DNA hybridization, whole-cell protein profiles by polyacrylamide gel electrophoresis, and biochemical comparison suggested that it is a strain of *H. cinaedi*. As shown in Fig. 2, the 23S rRNA gene sequence for the Mainz strain is 99.8% similar to that of the type strain and to a monkey strain of *H. cinaedi*. Thus, in the Mainz strain example, the 23S rRNA gene sequence data, but not the 16S rRNA gene sequence data, are consistent with other phenotypic and genotypic analyses.

“*Flexispira rappini*” is a provisional name given by Bryner et al. (7, 8) to fusiform-shaped organisms with spiral periplasmic fibers and bipolar tufts of sheathed flagella that were first isolated by Kirkbride et al. (30). In a previous study, we demonstrated that strains with flexispira morphology fell into at least 10 *Helicobacter* taxa by 16S rRNA gene sequence comparison (14). Since that paper was published, *Helicobacter aurati* and strain MIT 01-5529A, isolated from a seal, representing two additional flexispira taxa, have been described (24, 44). As noted in Fig. 3, the 12 flexispira taxa are scattered throughout the 16S rRNA gene tree. Phylogenetic analysis of *Helicobacter* sp. flexispira taxa by 23S rRNA gene sequence data highlights a number of discrepancies with the previous 16S rRNA gene analyses. Most prominent is the position of *Helicobacter* sp. flexispira taxon 1, strain ATCC 43968, which falls in cluster 2 by 16S rRNA gene analysis (Fig. 3) but falls in cluster 8 and is essentially identical to *Helicobacter* sp. flexispira taxon 4, strain ATCC 49310, by 23S rRNA gene analysis (Fig. 2). Of note in the 23S rRNA gene tree is that strains with the flexispira morphology group together in clusters 7 and 8, with only a single strain, *H. aurati*, in cluster 6. Recently, Hänninen suggested that *Helicobacter* spp. flexispira taxa 1, 4, and 5 are members of a single species, *H. trogontum* (flexispira taxon 6) on the basis of sodium dodecyl sulfate-polyacrylamide gel electrophoresis protein pattern analysis, dot blot hybridization, PCR-restriction fragment length polymorphism patterns, and HSP60 gene analysis (23, 40). The 23S rRNA gene data support Hänninen’s contention that *Helicobacter* spp. flexispira taxa 1, 4, and 5 and *H. trogontum* are closely related, but we would argue against placing these taxa into a single species at this time. Hänninen has also suggested that flexispira taxa 2, 3, and 8 be included in *Helicobacter bilis* on the basis of phenotypic characteristics, dot blot DNA-DNA hybridization, and phylogeny of *ureB* and HSP60 genes (22). While the 23S rRNA data are in general agreement with these results, we believe placement in a single species requires further analysis. There appears to be wide variability in the genetic diversity of helicobacter strains included in a species (lumping versus splitting). Several named species differ by only approximately 2%

by both 16S and 23S rRNA gene sequence comparisons: *H. canadensis* and *H. pullorum*, *H. rodentium* and *H. ganmani*, *H. cholecystus* and *H. pametensis*, *H. muridarum* and *H. typhlonius*, and *H. bilis* and *H. cinaedi*. We believe that species with highly divergent 16S rRNA gene strain sequences, like *H. felis*, may be comprised of more than one species. Species boundaries need to be set by rigorous quantitative means, such as the renaturation rate DNA-DNA hybridization of De Ley et al. (12), or sequence comparisons examining within-species versus among-species comparisons (10, 37). Furthermore, we are not convinced that dot blot hybridization or other less strenuous DNA-DNA hybridization methods are either quantitative or reliable.

The discrepancy in the position of *Wolinella succinogenes* between 16S and 23S rRNA gene phylogenetic analyses is marked and has significant taxonomic implications. By 16S rRNA gene analysis, *W. succinogenes* branches outside the genus *Helicobacter* and appears to represent a genus (43). By 23S rRNA gene analysis, it branches within the genus *Helicobacter*, consistently branching with the unsheathed-flagellum helicobacters, cluster 2 (Fig. 2). Whole-genome comparison of 870 shared protein genes supports the 23S rRNA gene placement of *W. succinogenes* within the helicobacters. As shown in Fig. 5 and Fig. 6, *H. hepaticus* is more closely related to *W. succinogenes* than to *H. pylori*, supporting *W. succinogenes*' position within the nongastric subcluster of helicobacters. The original genome papers for *H. hepaticus* (52) and *W. succinogenes* (2) were published nearly simultaneously, so cross comparisons were included in neither paper. A recent paper by Eppinger et al. (17) comparing these four genomes also demonstrated that *H. hepaticus* shared more genes with *W. succinogenes* than with *H. pylori* or *C. jejuni* (see Fig. 1 in reference 17). Despite these findings, the authors concluded that the phylogenetic position of *W. succinogenes* was in agreement with published 16S rRNA gene analyses. We strongly disagree with their conclusion and believe their data support *W. succinogenes* branching within the helicobacters.

The 23S rRNA gene and whole-genome analyses indicate that *W. succinogenes* branches with, and is as much a helicobacter as, *H. canadensis*, *H. ganmani*, *H. mesocricetorum*, *H. pullorum*, *H. rodentium*, and *H. winghamensis*. *W. succinogenes* and these six *Helicobacter* spp. differ morphologically from other helicobacters in having unsheathed flagella. However, if *W. succinogenes* is recognized as a member of the genus *Helicobacter*, it presents an onerous issue for helicobacter nomenclature. *Wolinella* Tanner et al. 1981 (55) has priority over *Helicobacter* Goodwin et al. 1989 (21) as a genus name. Changing the name of the genus from *Helicobacter* to *Wolinella* would not be accepted by the medical or microbiological communities and is not suggested by the authors. Nevertheless, the 23S sequence and other phenotypic and genotypic analyses suggest that the genus *Helicobacter* has three major divisions, the gastric helicobacters (cluster 1), the unsheathed-flagellum helicobacters (cluster 2), and the enterohepatic helicobacters (clusters 3 to 8). The name *Helicobacter* could be preserved by restricting the name *Wolinella* to organisms in cluster 2. However, this would prompt the question of whether the enterohepatic helicobacters (clusters 3 to 8) should also be designated a novel genus, as the remaining helicobacters (minus cluster 2) would be members of two paraphyletic groups (clus-

ter 1 and clusters 3 to 8). We make no taxonomic recommendations at this time but raise these issues for discussion, as they are implicit in our findings. The Subcommittee for the Taxonomy of *Campylobacter*, *Helicobacter*, and Related Organisms may be an appropriate body to consider these issues.

As described in the results section, the phylogenetic trees derived by neighbor-joining, parsimony, and maximum likelihood analyses of 23S rRNA gene sequence data support eight major clusters of helicobacters (Fig. 2 and 4). The trees derived by three different methods are generally congruent and have high bootstrapping values. The trees derived from the 16S rRNA gene sequence data differ significantly from one another depending on the treeing methodology (Fig. 3 and 4). The 16S rRNA gene sequence-based trees have significantly lower bootstrap values than the 23S rRNA gene-derived trees and are sensitive to change in outgroup or strain inclusion (unpublished data). Thus, while the phylogeny inferred from 23S rRNA gene sequence data is quite stable and consistent by different treeing methods and is consistent with other phenotypic and genotypic data, phylogenetic inferences made from 16S rRNA gene sequence data are discordant.

While small-subunit rRNA gene-based sequence analysis has become the foundation for prokaryotic and eukaryotic phylogenetic analysis, a number of issues are now recognized which qualify the meaning and use of the rRNA gene-derived phylogenetic trees. Early doubts about the validity of a universal tree of life based on the 16S rRNA gene analysis were in part dispelled by finding that the 23S rRNA gene trees and trees for certain conserved proteins, such as elongation factors, ATPase subunits, and RNA polymerases, were highly congruent (34, 35, 46). With the completion of numerous prokaryotic genomes, it became evident that horizontal gene transfer was a widespread phenomenon, raising new questions regarding the validity of 16S rRNA gene-based phylogenetic inference (9, 16). The realization that there is a distinction between gene phylogenies and organismal phylogeny added to the questioning of 16S rRNA gene-based phylogeny (16). These doubts were eased in part by studies such as those by Brochier et al. (4), who, using whole-genome data, generated a tree by concatenating 57 proteins from the translational apparatus and found it highly congruent with that obtained by concatenating 16S and 23S rRNA gene sequences from 45 bacterial species for which complete genome sequences were available. Jain et al. (28) articulated the "complexity hypothesis," which states that genes whose proteins are components of complex interacting systems are less likely to be transferred than genes whose proteins are involved in few or no interactions. Several investigators have found that informational genes (transcription, translation, and related processes) are more phylogenetically informative than operational genes (metabolism, transporters, etc.). Lerat et al. (32) searched complete genomes of gamma *Proteobacteria* and identified genes for which all species had single orthologs. He found 203/205 genes gave the same tree, and gene concatenation gave a fully resolved phylogeny. Only two cases of lateral gene transfer were found. There appears to be a general (but not universal) consensus that an organismal phylogeny exists and that it can be deduced from a comparison of sets of conserved genes, but the best methods for such analyses are still being debated. However, the common theme emerging from this debate is that no individual

molecule, including 16S rRNA, is completely reliable for reconstructing organismal phylogeny.

The use of 16S rRNA gene sequences for determining phylogeny was initially based on the assumption that sequence diversity was purely due to evolutionary change and that the 16S rRNA gene was not influenced by horizontal gene transfer due to "complexity hypothesis" considerations (28). However, as early as 1993, Sneath provided evidence for horizontal exchange of 16S rRNA gene fragments in *Aeromonas* species (50). A careful analysis of 16S rRNA genes in *Streptomyces* strains by Ueda et al. (56) demonstrated that there were two classes of substitution patterns. The first class included individual base changes, which were mainly transitions, and the second class involved groups of five or more base changes and involved mainly transversions, which suggested horizontal gene transfer. These findings were extended by Wang and Zhang (59), who examined actinomycetes. They found evidence consistent with lateral gene transfer of short gene segments corresponding to hairpins in the 16S rRNA gene. They described the exchange of 16S rRNA gene fragments as the "simplified complexity hypothesis." They suggested that this mode of lateral gene transfer would create mosaic rRNA genes and gradually destroy the evolutionary history recorded in the sequence. The proposed transfer of 16S rRNA gene segments is supported by Schouls et al. (48), who examined the *Streptococcus anginosus* group, and van Berkum et al. (57), who examined rhizobia. In the case of helicobacters, neither the "complexity hypothesis" nor the "simplified complexity hypothesis" is supported by the sequence changes seen in the Mainz strain. Alignment of the Mainz strain with *H. fennelliae* and *H. cinaedi* suggests that the entire 5' 1,260 bases were obtained from a helicobacter closely related to *H. fennelliae*.

Natural transformation in helicobacters has been studied most intensively in *H. pylori*. It has been found that recombination in *H. pylori* occurs at such a high rate that the population structure is panmictic and polymorphisms within many gene loci are at linkage equilibrium (54). DNA is thought to be taken up through a transmembrane channel system related to type IV secretion systems (49). Components of the transformation apparatus that have been recently characterized include a ComE3 homologue and a nuclease component (41, 62). Thus, it is not in the least surprising that helicobacter 16S rRNA gene sequences have been modified by horizontal exchange and their phylogenetic signal degraded. But as noted in the discussion above, it is still possible in the presence of genetic exchange to construct an organismal phylogeny. What is required for reliable phylogenetic analysis is inclusion of a sufficiently large number of informative bases to recover the true phylogenetic signal from the noise created by horizontal transfer of genetic information (whole genes and gene fragments). In analysis of helicobacter phylogeny, it would appear that the approximately 220 informative bases in the 16S rRNA gene sequence are insufficient, whereas the approximately 690 informative bases in the 23S rRNA gene sequence appear sufficient to recover the phylogenetic signal. Interestingly, neighbor-joining analysis of concatenated 16S and 23S rRNA gene sequences produces a tree essentially identical to the 23S rRNA gene tree (tree not shown). However, when neighbor-joining trees are generated from halves of the 23S rRNA gene data set yielding about 345 informative bases per half, the eight

clusters start to be rearranged, and when data sets containing 220 bases are used, the trees are rearranged substantially. This suggests that phylogenetic robustness is dependent on the total number of informative bases rather than the particular gene, at least in the cases of 16S and 23S rRNA genes.

Despite the greater information content in the 23S rRNA molecule, there are sound reasons why 16S rRNA has become the primary molecule for phylogenetic inference. Obtaining a full 23S rRNA gene sequence requires substantially more than twice the effort to obtain twice the sequence length. Unlike the typical 16S rRNA operon, where highly conserved primer targets exist within 10 bases of the 3' and 5' ends, only semiconserved targets exist for 23S rRNA operons 250 and 150 bases from the 3' and 5' ends: Lane's primers 256f and 2447r, respectively (31). Lane's most 5' 23S rRNA gene primer (130f) is not adequate for helicobacters. Primers closer to the ends, such as helicobacter 44–60 forward (O68) (Table 2), were designed specifically to amplify helicobacters and related taxa but were not intended to be universal. Sites suitable for sequencing primer targeting are also less conserved for the 23S rRNA gene than for the 16S rRNA gene. The initial sequencing primers used in this project were based on those described by Lane (31) and modified by examining the aligned sequences of *H. pylori*, *E. coli*, and other divergent taxa. After about 20 full helicobacter 23S rRNA gene sequences were obtained, their alignment was used for designing additional and more appropriate helicobacter-specific primers. The final primer set (Table 2) worked well for sequencing helicobacter 23S rRNA genes. For completeness, Table 2 also includes primers that were subsequently discarded. The sequences for the final primers contain some wobbles that are probably not required for the genus *Helicobacter*, but these may make them more likely to work with other epsilon *Proteobacteria*.

We conclude that the 23S rRNA gene sequence data are significantly more reliable than 16S rRNA gene sequence data for identification, classification, and phylogenetic analysis of helicobacters primarily because of the threefold-higher number of informative bases. This conclusion will best be supported or challenged by examining a third (or multiple) phylogenetically reliable molecule. This will not be a minor undertaking. The present study represented a major sequencing effort, as approximately 500,000 bp of raw sequence information were determined to generate the 150,000 bp of assembled 23S rRNA gene sequence information deposited in GenBank. To be useful for making a meaningful comparison with the 23S rRNA gene sequence data, the new molecule must have equal information content, about 700 informative bases for divergent *Helicobacter* species. Unfortunately, phylogenetic studies of helicobacters using the coding sequence for proteins, such as HSP60 (GroEL), examined a region of 590 bases with only approximately 120 informative bases (40). To obtain 700 informative bases will likely require sequencing 3 to 4 kb of finished sequence (a long single gene or multiple shorter genes) for each strain. Full double-strand coverage of a similar number of strains will require an effort comparable to the half million bases of raw sequence determined in the present study. Obtaining more reliable phylogenetic information than that provided by the 16S rRNA gene requires acquisition of greater sequencing information content, and unfortunately, there are no easy shortcuts.

The presence of IVSs in helicobacter 16S and 23S rRNA genes has been reported previously (26, 33), and they are known to occur in other prokaryotes, such as some *Enterobacteriaceae* (45). Where they have been examined, IVSs were cleaved without splicing by RNase III (18). Hurtado et al. (26) previously reported IVS insertions in helix 25. The present report documents IVS insertions in helices 10, 25, and 45 and between helices 31' and 27. This is also the first report of multiple IVS insertions into a helicobacter 23S rRNA gene, which we noted in *Helicobacter* sp. flexispira taxon 2 and *H. mesocricetorum*. The three IVSs in *H. mesocricetorum* increase the size of the 23S rRNA gene by 930 bases. Primers O68 and V62, which produce a 1.3-kb amplicon for most helicobacters, produce a 2.1-kb amplicon in *H. mesocricetorum*. The functions of IVSs in both 23S and 16S rRNA genes remain unknown. There are examples, such as *H. canis*, where only some strains of a species contain an IVS (26). IVSs may originate from the insertion of a mobile DNA element into the rRNA gene. It is not known whether IVSs carry either a positive or negative selective advantage, but a slight positive pressure would favor horizontal transfer of rRNA gene fragments carrying an IVS, and a slight negative pressure would favor horizontal transfer of rRNA gene fragments without an IVS.

Bacterial taxonomy or systematics deals with classification, nomenclature, and identification (11). Classification deals with the consistent arrangement of like taxa into groups. Nomenclature is the naming of taxa and is regulated by the International Committee of Systematic Bacteriology and the Bacteriological Code. Identification is the process of characterizing an unknown and determining if it matches a named taxon. Strains from a majority of helicobacter species can be reliably identified by a 16S rRNA gene sequence match of >99% to a species type strain. However, Fox et al. (20) provocatively raised the question of "how close is close," or how close 16S RNA sequences need to be to guarantee species identity. The answer is that a 99% 16S rRNA gene sequence similarity does not always guarantee species matching, as some closely related species have essentially identical 16S rRNA gene sequences. What was not addressed in the Fox paper is how often one would be led astray by assuming species identity for 99% 16S rRNA gene sequence similarity. Of the tens of thousands of bacterial strains that have 16S rRNA gene sequence information, less than a dozen cases have been cited where neighboring species have greater than 99% 16S rRNA gene sequence similarity. For routine identification, this level of error is substantially better than phenotypic analysis. A second important issue is how divergent 16S rRNA gene sequences can be and still represent a well-defined species (DNA-DNA hybridization of >70%). For the vast majority of species examined, <2% divergence, or >98% similarity, reliably defines a species. However, within the helicobacters, sequences for strains of *H. felis* differ from the type strain by up to 2% and could be confused with the closely related *H. bizzozeronii* and *H. salmonis*. The Mainz strain differs from other *H. cinaedi* strains by over 4% (58). Strains with sequences that are less than 99% similar to a type strain sequence fall into a poorly defined category where they cannot be reliably identified. Thus, while 16S rRNA gene analysis is generally more reliable than biochemical characterization for identification (because of the substantially greater information content), it is not perfect.

However, the assertion that investigators have routinely misidentified helicobacters based on 16S rRNA gene sequence analyses is false. What is true is that investigators have misclassified helicobacter strains based on the assumption that a 16S rRNA gene sequence difference of 2% or 3% from named species indicated a novel species. As reported by Vandamme et al. (58), at least one species of helicobacter, *H. cinaedi*, has strains with highly divergent 16S rRNA gene sequences (4.3%). The divergent sequence of the Mainz strain led Husmann et al. (27) to incorrectly classify the Mainz strain as a new species. The Mainz strain was misclassified, not misidentified. The proper standards for classification of novel helicobacters are fully described by Dewhirst et al. (15). While 16S rRNA gene divergence of greater than 2% or 3% (98% or 97% similarity) was used previously as one of the criteria to classify helicobacter strains as belonging to novel species, it is clearly invalid to use such an argument in light of current knowledge. A final point regarding the Mainz strain and the divergence of 16S rRNA gene within the species *H. cinaedi* needs to be made. Sequences for *H. cinaedi* strains do not differ randomly by 0 to 4% from the sequence of the type strain, but rather, the vast majority of strains are 99% similar to the type strain. Most likely, through genetic exchange of 16S rRNA genes or gene fragments with a species closely related to *H. fennelliae*, the Mainz strain's 16S rRNA gene sequence came to differ markedly (>4%) from other *H. cinaedi* strains. There are probably other *H. cinaedi* strains closely related to the Mainz strain (>99% similarity). Thus, *H. cinaedi* strains can be reliably identified if their 16S rRNA gene sequences are 99% similar to that of either the type strain or the Mainz strain. It is misleading for investigators to cite the *H. cinaedi* 4.3% within-species 16S rRNA gene sequence divergence without appropriate caveats.

The findings in this study have implications for prokaryotic phylogenetic inference and systematics. Where 16S rRNA gene-based phylogeny conflicts with other credible phenotypic and genotypic methods, as in the cases of helicobacters and the *Rhizobiaceae* (6), more robust molecular data should be obtained prior to making classification or taxonomic proposals. However, more robust data require determining more informative bases than are present in the 16S rRNA gene. Whether this is done by sequencing the 23S rRNA gene or sequencing other phylogenetically informative gene targets, it will likely require 3 to 4 kb of sequence information for each strain to be truly robust. Finally, identifying strains is always a trade-off between certainty and expense. Ideally, information on a plethora of phenotypic and genotypic traits would be obtained for this purpose. Despite the caveats presented in this paper, 16S rRNA gene sequence information is still among the most cost-effective and useful information to have for strain identification.

ACKNOWLEDGMENTS

This research was supported in part by grants from The National Institutes of Health: DE 10374 (F.E.D.) and DE 11443 (B.J.P.) from the National Institute for Dental and Craniofacial Research and R01 CA67529 (J.G.F.), R01 AI37750 (J.G.F.), and R01 AI50952 (J.G.F.).

REFERENCES

1. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
2. Baar, C., M. Eppinger, G. Raddatz, J. Simon, C. Lanz, O. Klimmek, R. Nandakumar, R. Gross, A. Rosinus, H. Keller, P. Jagtap, B. Linke, F. Meyer,

- H. Lederer, and S. C. Schuster. 2003. Complete genome sequence and analysis of *Wolinella succinogenes*. Proc. Natl. Acad. Sci. USA **100**:11690–11695.
3. Boone, D. R., R. W. Castenholz, and G. M. Garrity (ed.). 2001. Bergey's manual of systematic bacteriology: the Archaea and the deeply branching and phototrophic Bacteria, 2nd ed., vol. I. Springer, New York, N.Y.
 4. Brochier, C., E. Bapteste, D. Moreira, and H. Philippe. 2002. Eubacterial phylogeny based on translational apparatus proteins. Trends Genet. **18**:1–5.
 5. Brosius, J., M. L. Palmer, P. J. 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.
 6. Broughton, W. J. 2003. Roses by other names: taxonomy of the *Rhizobiaceae*. J. Bacteriol. **185**:2975–2979.
 7. Bryner, J. H. 1987. *Flexispira rappini*, gen. nov., sp. nov. A motile, urease-producing rod similar to *Campylobacter pyloridis*, p. 440–442. In B. Kaijer and E. Falsen (ed.), Proceedings of the Fourth International Workshop on Campylobacter Infections, Goteborg, Sweden. University of Göteborg, Göteborg, Sweden.
 8. Bryner, J. H., J. Littleton, C. Gates, C. A. Kirkbride, and A. E. Ritchie. 1986. *Flexispira rappini* gen. nov., sp. nov., a Gram-negative rod from mammalian fetus and feces, abstr. G11. XIV International Congress of Microbiology, Manchester, United Kingdom.
 9. Campbell, A. M. 2000. Lateral gene transfer in prokaryotes. Theor. Popul. Biol. **57**:71–77.
 10. Cladera, A. M., A. Bennisar, M. Barcelo, J. Lalucat, and E. Garcia-Valdes. 2004. Comparative genetic diversity of *Pseudomonas stutzeri* genomovars, clonal structure, and phylogeny of the species. J. Bacteriol. **186**:5239–5248.
 11. Cowen, S. T. 1968. A dictionary of microbial taxonomic usage. Oliver & Boyd, Edinburgh, United Kingdom.
 12. De Ley, J., H. Cattoir, and A. Reynaerts. 1970. The quantitative measurement of DNA hybridization from renaturation rates. Eur. J. Biochem. **12**:133–142.
 13. Dewhirst, F. E., C. C. Chien, B. J. Paster, R. L. Ericson, R. P. Orcutt, D. B. Schauer, and J. G. Fox. 1999. Phylogeny of the defined murine microbiota: altered Schaedler flora. Appl. Environ. Microbiol. **65**:3287–3292.
 14. Dewhirst, F. E., J. G. Fox, E. N. Mendes, B. J. Paster, C. E. Gates, C. A. Kirkbride, and K. A. Eaton. 2000. '*Flexispira rappini*' strains represent at least 10 *Helicobacter* taxa. Int. J. Syst. Evol. Microbiol. **50**:1781–1787.
 15. Dewhirst, F. E., J. G. Fox, and S. L. On. 2000. Recommended minimal standards for describing new species of the genus *Helicobacter*. Int. J. Syst. Evol. Microbiol. **50**:2231–2237.
 16. Doolittle, W. F. 1999. Lateral genomics. Trends Cell Biol. **9**:M5–M8.
 17. Eppinger, M., C. Baar, G. Raddatz, D. H. Huson, and S. C. Schuster. 2004. Comparative analysis of four *Campylobacteriales*. Nat. Rev. Microbiol. **2**:872–885.
 18. Evgenieva-Hackenberg, E., and G. Klug. 2000. RNase III processing of intervening sequences found in helix 9 of 23S rRNA in the alpha subclass of *Proteobacteria*. J. Bacteriol. **182**:4719–4729.
 19. Felsenstein, J. 2004. PHYLIP (Phylogeny Inference Package) version 3.6b. University of Washington, Seattle.
 20. 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.
 21. Goodwin, C. S., J. A. Armstrong, T. Chilvers, M. Peters, M. D. Collins, L. Sly, W. McConnell, and W. E. S. Harper. 1989. Transfer of *Campylobacter pylori* and *Campylobacter mustelae* to *Helicobacter* gen. nov. as *Helicobacter pylori* comb. nov. and *Helicobacter mustelae* comb. nov., respectively. Int. J. Syst. Bacteriol. **39**:397–405.
 22. Hanninen, M. L., R. I. Karenlampi, J. M. Koort, T. Mikkonen, and K. J. Bjorkroth. 2005. Extension of the species *Helicobacter bilis* to include the reference strains of *Helicobacter* sp. flexispira taxa 2, 3 and 8 and Finnish canine and feline flexispira strains. Int. J. Syst. Evol. Microbiol. **55**:891–898.
 23. Hanninen, M. L., M. Utriainen, I. Happonen, and F. E. Dewhirst. 2003. *Helicobacter* sp. flexispira 16S rDNA taxa 1, 4 and 5 and Finnish porcine *Helicobacter* isolates are members of the species *Helicobacter trogonum* (taxon 6). Int. J. Syst. Evol. Microbiol. **53**:425–433.
 24. Harper, C. G., S. Xu, A. B. Rogers, Y. Feng, Z. Shen, N. S. Taylor, F. E. Dewhirst, B. J. Paster, M. Miller, J. Hurlley, and J. G. Fox. 2003. Isolation and characterization of novel *Helicobacter* spp. from the gastric mucosa of harp seals *Phoca groenlandica*. Dis. Aquat. Organ. **57**:1–9.
 25. Harrington, C. S., and S. L. On. 1999. Extensive 16S rRNA gene sequence diversity in *Campylobacter hylotestinalis* strains: taxonomic and applied implications. Int. J. Syst. Bacteriol. **49**:1171–1175.
 26. Hurtado, A., J. P. Clewley, D. Linton, R. J. Owen, and J. Stanley. 1997. Sequence similarities between large subunit ribosomal RNA gene intervening sequences from different *Helicobacter* species. Gene **194**:69–75.
 27. Husmann, M., C. Gries, P. Jehnichen, T. Woelfel, G. Gerken, W. Ludwig, and S. Bhakdi. 1994. *Helicobacter* sp. strain Mainz isolated from an AIDS patient with septic arthritis: case report and nonradioactive analysis of 16S rRNA sequence. J. Clin. Microbiol. **32**:3037–3039.
 28. Jain, R., M. C. Rivera, and J. A. Lake. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. Proc. Natl. Acad. Sci. USA **96**:3801–3806.
 29. 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, New York, N.Y.
 30. Kirkbride, C. A., C. E. Gates, and J. E. Collins. 1986. Abortion in sheep caused by a nonclassified, anaerobic, flagellated bacterium. Am. J. Vet. Res. **47**:259–262.
 31. Lane, D. J. 1991. 16S/23S rRNA sequencing, p. 115–175. In E. Stackebrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics. John Wiley and Sons, New York, N.Y.
 32. Lerat, E., V. Daubin, and N. A. Moran. 2003. From gene trees to organismal phylogeny in prokaryotes: the case of the gamma-*Proteobacteria*. PLoS Biol. **1**:E19.
 33. Linton, D., J. P. Clewley, A. Burnens, R. J. Owen, and J. Stanley. 1994. An intervening sequence (IVS) in the 16S rRNA gene of the eubacterium *Helicobacter canis*. Nucleic Acids Res. **22**:1954–1958.
 34. Ludwig, W., J. Neumaier, N. Klugbauer, E. Brockmann, C. Roller, S. Jilg, K. Reetz, I. Schachtner, A. Ludvigsen, M. Bachleitner, et al. 1993. Phylogenetic relationships of Bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes. Antonie Leeuwenhoek **64**:285–305.
 35. Ludwig, W., and K. H. Schleifer. 1994. Bacterial phylogeny based on 16S and 23S rRNA sequence analysis. FEMS Microbiol. Rev. **15**:155–173.
 36. Ludwig, W., O. Strunk, S. Klugbauer, N. Klugbauer, M. Weizenegger, J. Neumaier, M. Bachleitner, and K. H. Schleifer. 1998. Bacterial phylogeny based on comparative sequence analysis. Electrophoresis **19**:554–568.
 37. Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95**:3140–3145.
 38. Marshall, B. J. 1995. *Helicobacter pylori* in peptic ulcer: have Koch's postulates been fulfilled? Ann. Med. **27**:565–568.
 39. Marshall, B. J., and J. R. Warren. 1983. Unidentified curved bacilli on gastric epithelium in active chronic gastritis. Lancet **i**:1273–1275.
 40. Mikkonen, T. P., R. I. Karenlampi, and M. L. Hanninen. 2004. Phylogenetic analysis of gastric and enterohepatic *Helicobacter* species based on partial HSP60 gene sequences. Int. J. Syst. Evol. Microbiol. **54**:753–758.
 41. O'Rourke, E. J., A. V. Pinto, E. A. Petroni, M. E. Tolmashy, and L. Ielpi. 2004. Evidence for the active role of a novel nucleic acid from *Helicobacter pylori* in the horizontal transfer of genetic information. J. Bacteriol. **186**:2586–2593.
 42. Page, R. D. 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12**:357–358.
 43. Paster, B. J., and F. E. Dewhirst. 1988. Phylogeny of campylobacters, wolinelas, *Bacteroides gracilis*, and *Bacteroides ureolyticus* by 16S ribosomal ribonucleic acid sequencing. Int. J. Syst. Bacteriol. **38**:56–62.
 44. Patterson, M. M., M. D. Schrenzel, Y. Feng, S. Xu, F. E. Dewhirst, B. J. Paster, S. A. Thibodeau, J. Versalovic, and J. G. Fox. 2000. *Helicobacter aurati* sp. nov., a urease-positive *Helicobacter* species cultured from gastrointestinal tissues of Syrian hamsters. J. Clin. Microbiol. **38**:3722–3728.
 45. Pronk, L. M., and K. E. Sanderson. 2001. Intervening sequences in *rrl* genes and fragmentation of 23S rRNA in genera of the family *Enterobacteriaceae*. J. Bacteriol. **183**:5782–5787.
 46. Puhler, G., H. Leffers, F. Gropp, P. Palm, H. P. Klenk, F. Lottspeich, R. A. Garrett, and W. Zillig. 1989. Archaeobacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proc. Natl. Acad. Sci. USA **86**:4569–4573.
 47. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.
 48. Schouls, L. M., C. S. Schot, and J. A. Jacobs. 2003. Horizontal transfer of segments of the 16S rRNA genes between species of the *Streptococcus anginosus* group. J. Bacteriol. **185**:7241–7246.
 49. Smeets, L. C., and J. G. Kusters. 2002. Natural transformation in *Helicobacter pylori*: DNA transport in an unexpected way. Trends Microbiol. **10**:159–162.
 50. Sneath, P. H. 1993. Evidence from *Aeromonas* for genetic crossing-over in ribosomal sequences. Int. J. Syst. Bacteriol. **43**:626–629.
 51. Stackebrandt, E. 1999. Unifying phylogeny and phenotypic diversity. In M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer, and E. Stackebrandt (ed.), The prokaryotes: an evolving electronic resource for the microbiological community, 3rd ed. Springer-Verlag, New York, N.Y.
 52. Suerbaum, S., C. Josenhans, T. Sterzenbach, B. Drescher, P. Brandt, M. Bell, M. Droge, B. Fartmann, H. P. Fischer, Z. Ge, A. Horster, R. Holland, K. Klein, J. Konig, L. Macko, G. L. Mendz, G. Nyakatura, D. B. Schauer, Z. Shen, J. Weber, M. Frosch, and J. G. Fox. 2003. The complete genome sequence of the carcinogenic bacterium *Helicobacter hepaticus*. Proc. Natl. Acad. Sci. USA **100**:7901–7906.
 53. Suerbaum, S., C. Kraft, F. E. Dewhirst, and J. G. Fox. 2002. *Helicobacter nemestrinae* ATCC 49396^T is a strain of *Helicobacter pylori* (Marshall et al. 1985) Goodwin et al. 1989, and *Helicobacter nemestrinae* Bronsdon et al. 1991 is therefore a junior heterotypic synonym of *Helicobacter pylori*. Int. J. Syst. Evol. Microbiol. **52**:437–439.
 54. Suerbaum, S., J. M. Smith, K. Bapumia, G. Morelli, N. H. Smith, E. Kunst-

- mann, I. Dyrek, and M. Achtman. 1998. Free recombination within *Helicobacter pylori*. Proc. Natl. Acad. Sci. USA **95**:12619–12624.
55. Tanner, A. C. R., S. Badger, C. H. Lai, M. A. Listgarten, R. A. Visconti, and S. S. Socransky. 1981. *Wolinella* gen. nov., *Wolinella succinogenes* (*Vibrio succinogenes* Wolin et al.) comb. nov., and description of *Bacteroides gracilis* sp. nov., *Wolinella recta* sp. nov., *Campylobacter concisus* sp. nov., and *Eikenella corrodens* from humans with periodontal disease. Int. J. Syst. Bacteriol. **31**:432–445.
56. Ueda, K., K. Matsuda, H. Takano, and T. Beppu. 1999. A putative regulatory element for carbon-source-dependent differentiation in *Streptomyces griseus*. Microbiology **145**:2265–2271.
57. van Berkum, P., Z. Terefework, L. Paulin, S. Suomalainen, K. Lindstrom, and B. D. Eardly. 2003. Discordant phylogenies within the loci of *Rhizobia*. J. Bacteriol. **185**:2988–2998.
58. Vandamme, P., C. S. Harrington, K. Jalava, and S. L. On. 2000. Misidentifying helicobacters: the *Helicobacter cinaedi* example. J. Clin. Microbiol. **38**:2261–2266.
59. Wang, Y., and Z. Zhang. 2000. Comparative sequence analyses reveal frequent occurrence of short segments containing an abnormally high number of non-random base variations in bacterial rRNA genes. Microbiology **146**:2845–2854.
60. Whary, M. T., and J. G. Fox. 2004. Natural and experimental *Helicobacter* infections. Comp. Med. **54**:128–158.
61. Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. USA **87**:4576–4579.
62. Yeh, Y. C., T. L. Lin, K. C. Chang, and J. T. Wang. 2003. Characterization of a ComE3 homologue essential for DNA transformation in *Helicobacter pylori*. Infect. Immun. **71**:5427–5431.