Microbiology (2011), **157**, 3388–3397 DOI 10.1099/mic.0.052332-0

Identification and classification of the genus *Bacteroides* by multilocus sequence analysis

Mitsuo Sakamoto and Moriya Ohkuma

Microbe Division/Japan Collection of Microorganisms, RIKEN BioResource Center, Wako, Saitama 351-0198, Japan

Correspondence
Mitsuo Sakamoto
sakamoto@jcm.riken.jp

Received 21 June 2011
Revised 20 August 2011
Accepted 20 September 2011

Multilocus sequence analysis (MLSA) was performed on representative species of the genus *Bacteroides*. Internal fragments of the genes selected, *dnaJ*, *gyrB*, *hsp60*, *recA*, *rpoB* and 16S rRNA, were amplified by direct PCR and then sequenced from 38 *Bacteroides* strains representing 35 species. Neighbour-joining (NJ), maximum-likelihood (ML) and maximum-parsimony (MP) phylogenies of the individual genes were compared. The data confirm that the potential for discrimination of *Bacteroides* species is greater using MLSA of housekeeping genes than 16S rRNA genes. Among the housekeeping genes analysed, *gyrB* was the most informative, followed by *dnaJ*. Analyses of concatenated sequences (4816 bp) of all six genes revealed robust phylogenetic relationships among different *Bacteroides* species when compared with the single-gene trees. The NJ, ML and MP trees were very similar, and almost fully resolved relationships of *Bacteroides* species were obtained, to our knowledge for the first time. In addition, analysis of a concatenation (2457 bp) of the *dnaJ*, *gyrB* and *hsp60* genes produced essentially the same result. Ten distinct clades were recognized using the SplitsTree4 program. For the genus *Bacteroides*, we can define species as a group of strains that share at least 97.5 % gene sequence similarity based on the fragments of five protein-coding housekeeping genes and the 16S rRNA gene. This study demonstrates that MLSA of housekeeping genes is a valuable alternative technique for the identification and classification of species of the genus *Bacteroides*.

INTRODUCTION

The genus *Bacteroides* is part of the indigenous microbiota of human and animal gastrointestinal tracts, but various species in this group are commonly associated with a variety of human and animal infections. More than one-third of all anaerobes isolated from clinical specimens are members of the *Bacteroides fragilis* group. Bacteraemia due to various species in this group contributes considerably to morbidity and mortality (Redondo *et al.*, 1995). Correct identification at the species level is necessary because the resistance to various antibiotics may differ according to the species (Snydman *et al.*, 2007).

Several identification kits, such as API ZYM and Rapid ID 32A, are available commercially for the rapid identification of anaerobes. However, the databases accompanying the kits are often incomplete or inaccurate, especially with a number of novel species. A short biochemical scheme described in the Wadsworth–KTL Anaerobic Bacteriology Manual (Jousimes-Somer *et al.*, 2002) was developed specifically for the identification of *B. fragilis* group species. However, it has been reported that this biochemical scheme is not as accurate as a multiplex PCR identification scheme (Liu *et al.*, 2003).

MALDI-TOF MS has been used for the rapid identification of micro-organisms associated with infectious diseases (Grosse-Herrenthey *et al.*, 2008; Mellmann *et al.*, 2008). Nagy *et al.* (2009) set up a database for *Bacteroides* species and tested the applicability of MALDI-TOF MS profiling for the identification of *Bacteroides* species. The discriminatory power and identification accuracy of MALDI-TOF MS were superior to biochemical testing in the cases of *Bacteroides thetaiotaomicron, Bacteroides ovatus* and *Bacteroides uniformis*. Extension of the database to include other novel *Bacteroides* species is necessary for more accurate identification of the genus *Bacteroides*.

A number of novel species have been recently proposed in the genus *Bacteroides* (Bakir *et al.*, 2006a, b, c; Hayashi *et al.*, 2007; Kitahara *et al.*, 2005, 2011; Lan *et al.*, 2006). These taxonomic studies were mainly based on 16S rRNA gene sequence analysis (Shah *et al.*, 2009). It has been reported that 16S rRNA gene sequencing is useful for the identification

Abbreviations: AU, approximately unbiased; DDH, DNA–DNA hybridization; ML, maximum-likelihood; MLSA, multilocus sequence analysis; MLST, multilocus sequence typing; MP, maximum-parsimony; NJ, neighbour-joining; SH, Shimodaira–Hasegawa.

The GenBank/EMBL/DDBJ accession numbers of the sequences studied are listed in Table 1.

Two supplementary figures and three supplementary tables are available with the online version of this paper.

of isolates of clinically significant *Bacteroides* species (Song *et al.*, 2005). However, accurate identification of an isolate closely related to known species is often difficult when solely based on the 16S rRNA gene analysis and other DNA-based methods such as DNA–DNA hybridization (DDH) are required for differentiating closely related species, as was shown in the case of *Parabacteroides johnsonii* and *Parabacteroides merdae* (Sakamoto *et al.*, 2007). Because the DDH method is time-consuming and laborious, an alternative molecular marker for the identification of species is anticipated.

Recently, the *hsp60* gene has been found to be an alternative phylogenetic marker for the identification and classification of a broad range of Gram-negative anaerobic rods, including the genera *Barnesiella*, *Butyricimonas*, *Odoribacter*, *Parabacteroides*, *Paraprevotella*, *Porphyromonas*, *Prevotella* and *Tannerella* as well as *Bacteroides* (Sakamoto & Ohkuma, 2010; Sakamoto *et al.*, 2010). Multilocus sequence typing (MLST) has also been used for discriminating bacterial strains (Enersen *et al.*, 2006; Tanigawa & Watanabe, 2011). MLST is a typing scheme based on the DNA sequence of typically four to 10 loci in a bacterial genome to identify and classify bacterial strains and to unravel population genetics, molecular evolution and epidemiology of the species the strains belong to. This method was first proposed by Maiden *et al.* (1998). In MLST, allele sequences at each locus are assigned numbers. Strains that have the same alleles at all loci are considered to belong to the same sequence type. Also every sequence type has a number. Most of the downstream analyses in MLST are based on allele numbers and sequence types. Multilocus sequence analysis (MLSA) has been developed by using the concatenated sequences of multiple core (housekeeping) genes to assess clustering patterns and explore the relationships among the strains of similar species. MLSA is an extension of the MLST approach (Gevers *et al.*, 2005; Hanage *et al.*, 2006). In MLSA, the actual DNA sequences are used in the downstream analyses. MLST is usually applied to strains that belong to a well-defined species while MLSA is more often used when species boundaries are not well known and MLSA data are used to improve species descriptions. To date, MLSA has been widely used to infer the phylogeny of a variety of bacteria (Ah-You *et al.*, 2009; Bui Thi Ngoc *et al.*, 2010; Leon *et al.*, 2010; Margos *et al.*, 2009; Martens *et al.*, 2007, 2008; Pascual *et al.*, 2010; Rivas *et al.*, 2009; Sawabe *et al.*, 2007; Thompson *et al.*, 2005). However, only limited information has been accumulated for *Bacteroides* species. Besides the above-mentioned *hsp60* gene, it has been reported that housekeeping gene sequences, e.g. *rpoB*, could be used to clarify interspecies phylogenetic relationships within the genus *Bacteroides* (Ko *et al.*, 2007). The ad hoc committee has recommended evaluation of protein-coding gene sequence analysis (a minimum of five genes) for its applicability to genomically circumscribe the taxon species and differentiate it from neighbouring species detected by 16S rRNA gene sequence (Stackebrandt *et al.*, 2002).

In this study, we evaluated the usefulness of the MLSA approach for the identification and classification of the genus *Bacteroides*, using the fragments of five protein-coding housekeeping genes (*dnaJ*, *gyrB*, *hsp60*, *recA* and *rpoB*) and 16S rRNA gene from 38 strains representing 35 species.

METHODS

Bacterial strains. A total of 38 *Bacteroides* strains (35 species) used in this study were obtained from the Japan Collection of Microorganisms (JCM), RIKEN BioResource Center, Wako, Japan, and are listed in Table 1. The strains were maintained according to the protocols of the JCM Online Catalogue (http://www.jcm.riken.jp/JCM/catalogue.shtml). Valid species not included in the MLSA study are listed in Supplementary Table S1, available with the online version of this paper.

DNA preparation, PCR amplification and sequencing. Bacterial genomic DNAs were extracted by using the High Pure PCR Template Preparation kit (Roche Applied Science). The 16S rRNA gene was analysed as described by Sakamoto *et al.* (2002). The partial *hsp60* gene (558 bp) was analysed as described previously (Sakamoto & Ohkuma, 2010; Sakamoto *et al.*, 2010).

The partial *gyrB* gene (approx. 1200 bp) was amplified by PCR using the previously described primers UP-1 (5′-GAA GTC ATC ATG ACC GTT CTG CAY GCN GGN GGN AAR TTY GA-3′) and UP-2r (5′-AGC AGG GTA CGG ATG TGC GAG CCR TCN ACR TCN GCR TCN GTC AT-3′) (Yamamoto & Harayama, 1995). Each sequencing primer (UP-1S and UP-2Sr) is underlined. Amplification reactions were performed in a total volume of 50 µl containing 5 µl DNA (50 ng), 1.25 U *TaKaRa Ex Taq* (Takara Bio), 5 µl 10× *Ex Taq* buffer, 4 µl dNTP mixture (2.5 mM each) and 50 pmol of each primer. The *gyrB* genes were amplified in a Biometra Thermocycler TGradient (Biometra) using the following programme: 95 °C for 5 min, followed by 35 cycles consisting of 95 °C for 30 s, 50 °C for 30 s and 72 °C for 1 min, with a final extension period at 72 °C for 7 min.

A partial sequence of the *rpoB* gene (358 bp) was amplified by PCR using the previously described primers BF (5′-CAC TTG AGC AAY CGT CGT RT-3′) and BR (5′-CCT TCA GGA GTY TCA ATN GG-3′) (Ko *et al.*, 2007). Amplification reactions were performed in the above-mentioned PCR mixture. The *rpoB* genes were amplified using the following programme: 95 °C for 5 min, followed by 30 cycles consisting of 95 °C for 30 s, 55 °C for 30 s and 72 °C for 30 s, with a final extension period at 72 °C for 5 min.

Novel *dnaJ*-specific primers were designed from the available sequence data of *Bacteroides caccae* ATCC 43185, *Bacteroides coprocola* DSM 17136, *Bacteroides dorei* DSM 17855, *Bacteroides eggerthii* DSM 20697, *Bacteroides finegoldii* DSM 17565, *B. fragilis* NCTC 9343, *B. fragilis* YCH46, *Bacteroides intestinalis* DSM 17393, *B. ovatus* ATCC 8483, *Bacteroides plebeius* DSM 17135, *Bacteroides stercoris* ATCC 43183, *B. thetaiotaomicron* VPI-5482, *B. uniformis* ATCC 8492, *Bacteroides vulgatus* ATCC 8482, *Parabacteroides distasonis* ATCC 8503, *Parabacteroides merdae* ATCC 43184, *Porphyromonas gingivalis* ATCC 33277 and *Po. gingivalis* W83. The partial *dnaJ* gene (927 bp) was amplified by PCR using the primers dnaJ3F (5′-GHD AAA AGR GAY TAY TAY GAA-3′) and dnaJ4R (5′-RTC DAT IGT BGG IAY YTC HAC-3′). Amplification reactions were performed in the above-mentioned PCR mixture. The *dnaJ* genes were amplified using the following programme: 94 °C for 3 min, followed by 40 cycles consisting of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 1 min, with a final extension period at 72 °C for 10 min.

Novel *recA*-specific primers were designed from the available sequence data of *B. fragilis* NCTC 9343, *B. fragilis* YCH46, *B. thetaiotaomicron* VPI-5482, *B. vulgatus* ATCC 8482, *Pa. distasonis* ATCC 8503, *Po. gingivalis* ATCC 33277 and *Po. gingivalis* W83. A partial sequence of the *recA* gene (669 bp) was amplified by PCR using the primers recAF (5′-GAA TCI TCC GGT AAR ACI ACI-3′) and recAR (5′-CCA IGA ICC GCT YTT YTT GAT-3′). Amplification reactions were performed in the above-mentioned PCR mixture. The *recA* genes were amplified using the same programme as for *dnaJ*.

Amplified DNAs were visualized on 1.5 % agarose gels after electrophoresis at 100 V in TAE buffer for approximately 30 min, ethidium bromide staining and illumination with UV light. PCR products were purified using an UltraClean PCR Clean-up DNA purification kit (Mo Bio Laboratories). Sequencing of purified PCR products was performed with a BigDye Terminator v3.1 cycle sequencing ready reaction kit (Applied Biosystems) in an Applied Biosystems 3130*xl* Genetic Analyzer. Sequencing primers were the same ones as used in the amplification reaction (except for the *gyrB* and *hsp60* genes).

Phylogenetic data analysis. Sequences were aligned with the CLUSTAL_X 2.0.12 program (Larkin *et al.*, 2007) and corrected by manual inspection. Nucleotide substitution rates (K_{nuc} values) were calculated (Kimura, 1980) after gaps and unknown bases were eliminated. The phylogenetic tree was constructed by the neighbour-joining (NJ) method (Saitou & Nei, 1987). Bootstrap resampling analysis (Felsenstein, 1985) was performed to estimate the confidence of tree topologies. Maximum-likelihood (ML) analysis was performed using the program PhyML version 3.0 (Guindon & Gascuel, 2003). For ML analysis of each gene, the optimal models of nucleotide substitution were estimated through the program jModelTest 0.1.1 (Posada, 2008) using the Akaike information criterion. The models selected were TIM2 + G for *dnaJ*, TIM2 + I + G for *recA*, TIM3 + I + G

Table 1. Bacterial strains used in this study and their accession numbers in the databases

Accession numbers of new data are in bold type.

Species	Strain	*dnaJ*	*gyrB*	*hsp60*	*recA*	*rpoB*	16S rRNA
B. acidifaciens	JCM 10556^T	**AB619859**	**AB619897**	AB510667	**AB619935**	**AB619973**	AB510696
B. barnesiae	JCM 13652^T	**AB619860**	**AB619898**	AB510668	**AB619936**	**AB619974**	AB253726
B. caccae	JCM 9498^T	**AB619861**	**AB619899**	AB510669	**AB619937**	**AB619975**	AB510697
B. cellulosilyticus	JCM 15632^T	**AB619862**	**AB619900**	AB510670	**AB619938**	**AB619976**	AB510698
B. chinchillae	JCM 16497^T	**AB619863**	**AB619901**	AB547549	**AB619939**	**AB619977**	AB547636
B. chinchillae	JCM 16498	**AB619864**	**AB619902**	AB547550	**AB619940**	**AB619978**	AB547637
B. clarus	JCM 16067^T	**AB619865**	**AB619903**	AB547551	**AB619941**	**AB619979**	AB547638
B. coprocola	JCM 12979^T	**AB619866**	**AB619904**	AB510671	**AB619942**	**AB619980**	AB200224
B. coprophilus	JCM 13818^T	**AB619867**	**AB619905**	AB510672	**AB619943**	**AB619981**	AB260026
B. coprosuis	JCM 13475^T	**AB619868**	**AB619906**	AB510673	**AB619944**	**AB619982**	AB510699
B. dorei	JCM 13471^T	**AB619869**	**AB619907**	AB510674	**AB619945**	**AB619983**	AB242142
B. eggerthii	JCM 12986^T	**AB619870**	**AB619908**	AB510675	**AB619946**	**AB619984**	AB510700
B. faecis	JCM 16478^T	**AB619871**	**AB619909**	AB547553	**AB619947**	**AB619985**	AB547640
B. faecis	JCM 16477	**AB619872**	**AB619910**	AB547554	**AB619948**	**AB619986**	AB547641
B. finegoldii	JCM 13345^T	**AB619873**	**AB619911**	AB510676	**AB619949**	**AB619987**	AB222699
B. fluxus	JCM 16101^T	**AB619874**	**AB619912**	AB547555	**AB619950**	**AB619988**	AB547642
B. fragilis	JCM 11019^T	**AB619875**	**AB619913**	AB510677	**AB619951**	**AB619989**	AB510701
B. fragilis	JCM 11017	**AB619876**	**AB619914**	AB542762	**AB619952**	**AB619990**	AB542764
B. gallinarum	JCM 13658^T	**AB619877**	**AB619915**	AB510678	**AB619953**	**AB619991**	AB253732
B. graminisolvens	JCM 15093^T	**AB619878**	**AB619916**	AB547556	**AB619954**	**AB619992**	AB547643
B. helcogenes	JCM 6297^T	**AB619879**	**AB619917**	AB510679	**AB619955**	**AB619993**	AB510702
B. intestinalis	JCM 13265^T	**AB619880**	**AB619918**	AB510680	**AB619956**	**AB619994**	AB214328
B. massiliensis	JCM 13223^T	**AB619881**	**AB619919**	AB510681	**AB619957**	**AB619995**	AB510703
B. nordii	JCM 12987^T	**AB619882**	**AB619920**	AB510682	**AB619958**	**AB619996**	AB510704
B. oleiciplenus	JCM 16102^T	**AB619883**	**AB619921**	AB547557	**AB619959**	**AB619997**	AB547644
B. ovatus	JCM 5824^T	**AB619884**	**AB619922**	AB510683	**AB619960**	**AB619998**	AB510705
B. plebeius	JCM 12973^T	**AB619885**	**AB619923**	AB510684	**AB619961**	**AB619999**	AB200217
B. pyogenes	JCM 6294^T	**AB619886**	**AB619924**	AB510686	**AB619962**	**AB620000**	AB200229
B. rodentium	JCM 16496^T	**AB619887**	**AB619925**	AB547559	**AB619963**	**AB620001**	AB547646
B. salanitronis	JCM 13657^T	**AB619888**	**AB619926**	AB510687	**AB619964**	**AB620002**	AB253731
B. salyersiae	JCM 12988^T	**AB619889**	**AB619927**	AB510688	**AB619965**	**AB620003**	AB510707
B. stercoris	JCM 9496^T	**AB619890**	**AB619928**	AB510689	**AB619966**	**AB620004**	AB510708
B. suis (*B. pyogenes*)	JCM 6292^T	**AB619891**	**AB619929**	AB510690	**AB619967**	**AB620005**	AB510709
B. tectus (*B. pyogenes*)	JCM 10003^T	**AB619892**	**AB619930**	AB510691	**AB619968**	**AB620006**	AB200228
B. thetaiotaomicron	JCM 5827^T	**AB619893**	**AB619931**	AB510692	**AB619969**	**AB620007**	AB510710
B. uniformis	JCM 5828^T	**AB619894**	**AB619932**	AB510693	**AB619970**	**AB620008**	AB510711
B. vulgatus	JCM 5826^T	**AB619895**	**AB619933**	AB510694	**AB619971**	**AB620009**	AB510712
B. xylanisolvens	JCM 15633^T	**AB619896**	**AB619934**	AB510695	**AB619972**	**AB620010**	AB510713

for 16S rRNA, TVM + I + G for *rpoB*, GTR + I + G for *gyrB*, *hsp60* and the concatenated gene datasets (three genes and six genes). Maximum-parsimony (MP) analysis was performed with MEGA version 4 (Tamura *et al.*, 2007) using a heuristic search option. Bootstrap analyses were performed using 1000 replications for NJ, ML and MP. Pairwise sequence similarities were calculated with GENETYX-MAC version 15.0.5 (GENETYX Corporation). DNA polymorphism data, the mean G + C content and Tajima's *D* value were analysed with DnaSP version 5.10.01 (Librado & Rozas, 2009). Phylogenetic calculations, including synonymous and non-synonymous substitutions, were performed using MEGA version 4. Phylogenetic congruence tests with a maximum-likelihood framework were performed using the PhyML topologies and the site likelihood values were computed by using Tree-Puzzle 5.2 (Schmidt *et al.*, 2002) and fed into CONSEL 0.1i (Shimodaira & Hasegawa, 2001) to perform the Shimodaira–Hasegawa (SH) test (Shimodaira & Hasegawa, 1999) and the approximately unbiased (AU) test (Shimodaira, 2002). A *P*-value of less than 0.05 was considered significant incongruence. Neighbour-Net analysis (Bryant & Moulton, 2004) was performed with the SplitsTree4 program (Huson & Bryant, 2006). Sequence similarities were corrected using Jukes–Cantor correction. Recombination in each gene was analysed by the Phi test (Bruen *et al.*, 2006) implemented in SplitsTree4.

Statistical analysis. Statistical analysis was performed using ANOVA with R version 2.10.1 (R Development Core Team, 2009). A *P*-value of less than 0.05 was considered significant.

RESULTS AND DISCUSSION

In this study, we analysed the partial sequences of six genes – five protein-coding housekeeping genes and the 16S rRNA gene – from 38 *Bacteroides* strains of 35 species (Table 1). The genes selected are widely distributed, unique within the genome, of adequate length to be phylogenetically informative, located separately on the main chromosome (as assessed from the *B. fragilis* NCTC 9343 complete genome) and have a relatively high degree of conservation. The length of the alignments used ranged from 315 bp (*rpoB*) to 1417 bp (16S rRNA) (Table 2). There were no gaps in the sequenced *hsp60*, *recA* and *rpoB* gene regions, whereas gaps resulting in variations in sequence length were observed in the 16S rRNA, *dnaJ* and *gyrB* gene sequences. The mean DNA G + C content for each gene ranged from 44.7 mol% for the *hsp60* gene to 51.7 mol% for the 16S rRNA gene, which is a little higher than the range of G + C content reported for the genus *Bacteroides* (40–49 mol%) (Sakamoto & Benno, 2006). The average d_N/d_S values are shown in Table 2. All values are <1, indicating that most of the sequence variability identified is purifying (stabilizing) selection. A negative Tajima's *D* value (*dnaJ*, *hsp60* and *rpoB*) signifies an excess of low-frequency polymorphisms, indicating population size expansion and/or probably purifying selection on those genes. A positive Tajima's *D* value (*gyrB* and *recA*) signifies low levels of both low- and high-frequency polymorphisms, indicating a decrease in population size and/or balancing selection on those genes (Table 2).

Individual gene sequences

The intra-species comparisons of the three species where more than one strain was represented demonstrated high sequence similarities for the *dnaJ*, *gyrB*, *hsp60*, *recA*, *rpoB* and 16S rRNA genes: *Bacteroides chinchillae* (100, 100, 100, 100, 100 and 99.7 %, respectively), *Bacteroides faecis* (100, 100, 100, 100, 100 and 100 %, respectively) and *B. fragilis* (99.9, 99.5, 99.8, 100, 99.7 and 99.7 %, respectively). The ranges of inter-species similarity were 67.9–100 % (*dnaJ*),

Table 2. Analysis of 35 *Bacteroides* sequences from the *dnaJ*, *gyrB*, *hsp60*, *recA*, *rpoB* and 16S rRNA genes

NA, Not applicable.

Sequence information	*dnaJ*	*gyrB*	*hsp60*	*recA*	*rpoB*	16S rRNA
No. of sites	816	1083	558	627	315	1417
Mean G + C content (mol%)	49.7	48.9	44.7	47.9	46.9	51.7
No. of polymorphic sites	435 (53.3 %)	597 (55.1 %)	234 (41.9 %)	271 (43.2 %)	121 (38.4 %)	358 (25.3 %)
No. of parsimony-informative sites	394 (48.3 %)	548 (50.6 %)	199 (35.7 %)	255 (40.7 %)	101 (32.1 %)	278 (19.6 %)
No. of nucleotide differences:						
Range	0–262	0–386	0–151	0–175	0–82	0–229
Mean ± SEM	187.1 ± 8.1	264.4 ± 8.9	83.8 ± 5.4	124.8 ± 6.1	46.1 ± 3.4	116.1 ± 5.9
Jukes–Cantor distance (*d*):						
Range	0–0.419	0–0.484	0–0.336	0–0.349	0–0.32	0–0.182
Overall mean ± SEM	0.276 ± 0.015	0.299 ± 0.012	0.169 ± 0.012	0.232 ± 0.014	0.164 ± 0.014	0.087 ± 0.005
Transition : transversion ratio (*R*)	1.5	1.2	1.5	1.6	1.7	1.1
d_S*	1.76 ± 0.073	1.66 ± 0.073	1.00 ± 0.074	1.85 ± 0.068	1.05 ± 0.099	NA
d_N*	0.099 ± 0.009	0.12 ± 0.010	0.031 ± 0.006	0.048 ± 0.008	0.015 ± 0.004	NA
d_N/d_S	0.056	0.072	0.031	0.026	0.014	NA
Tajima's *D* value	−0.0327	0.0385	−0.2444	0.0909	−0.3439	NA

*Synonymous substitutions per site (d_S) and non-synonymous substitutions per site (d_N) (means ± SEM) were determined by the Nei–Gojobori method using Jukes–Cantor distance.

64.4–100 % (*gyrB*), 72.9–100 % (*hsp60*), 72.1–100 % (*recA*) and 74.0–100 % (*rpoB*). These inter-species similarity values were clearly lower than that of the 16S rRNA gene (83.5–100 %) (Sakamoto & Ohkuma, 2010) ($P<0.001$). In all comparisons (six genes), *Bacteroides pyogenes* JCM 6294^T and *Bacteroides suis* JCM 6292^T shared 100 % sequence similarity with each other. Sakamoto *et al.* (2010) have already reported that *B. suis* is a later heterotypic synonym of *B. pyogenes*. The present study is in agreement with DDH data on the species *B. pyogenes* and *B. suis*. The above-mentioned finding shows the greater potential for species discrimination using the protein-coding housekeeping genes when compared with the 16S rRNA gene.

Among the housekeeping genes analysed, *gyrB* was the most informative (Table 2). The *gyrB* gene sequences exhibited 548 (50.6 %) parsimony-informative sites. Recently, a *gyrB*-based real-time PCR system has been developed for detecting *B. fragilis* as a human-specific marker of faecal contamination (Lee & Lee, 2010). This designing of *gyrB*-based primers for *B. fragilis* is based on a higher discrimination power than the 16S rRNA gene. The next most informative gene was *dnaJ*, with 48.3 % parsimony informative. On the other hand, the percentage of parsimony-informative sites for the 16S rRNA gene was the lowest (19.6 %).

In this study, the gene sequences were used for the phylogenetic analyses because the phylogenetic tree obtained from the amino acid sequences showed low resolution (data not shown). For each of the genes, the phylogenetic trees constructed by the NJ, ML or MP methods gave similar results, except for some minor differences at the deepest branching points (see Supplementary Fig. S1, available with the online version of this paper). In particular, the three algorithms produced very similar branching patterns for the *dnaJ* gene.

Concatenated gene sequences

Analysis of concatenated gene sequences improves the quality of the phylogenetic reconstruction and optimizes the taxonomic resolution. The concatenated alignment of all six genes comprised 4816 nt, consisting of 2799 invariable sites, 241 variable but parsimony-uninformative sites and 1775 (36.9 %) parsimony-informative sites. The NJ, ML and MP trees were very similar and almost fully resolved relationships of *Bacteroides* species were obtained, to our knowledge for the first time (Fig. 1).

We also concatenated the aligned sequences for the two most resolving genes, *dnaJ* and *gyrB* (Table 2), and the *hsp60* gene, which is a useful alternative phylogenetic marker (Sakamoto & Ohkuma, 2010), and obtained an alignment of 2457 nt comprising 1191 invariable sites, 125 variable but parsimony-uninformative sites and 1141 (46.4 %) parsimony-informative sites. NJ, ML and MP trees were also similar and bootstrap values were generally

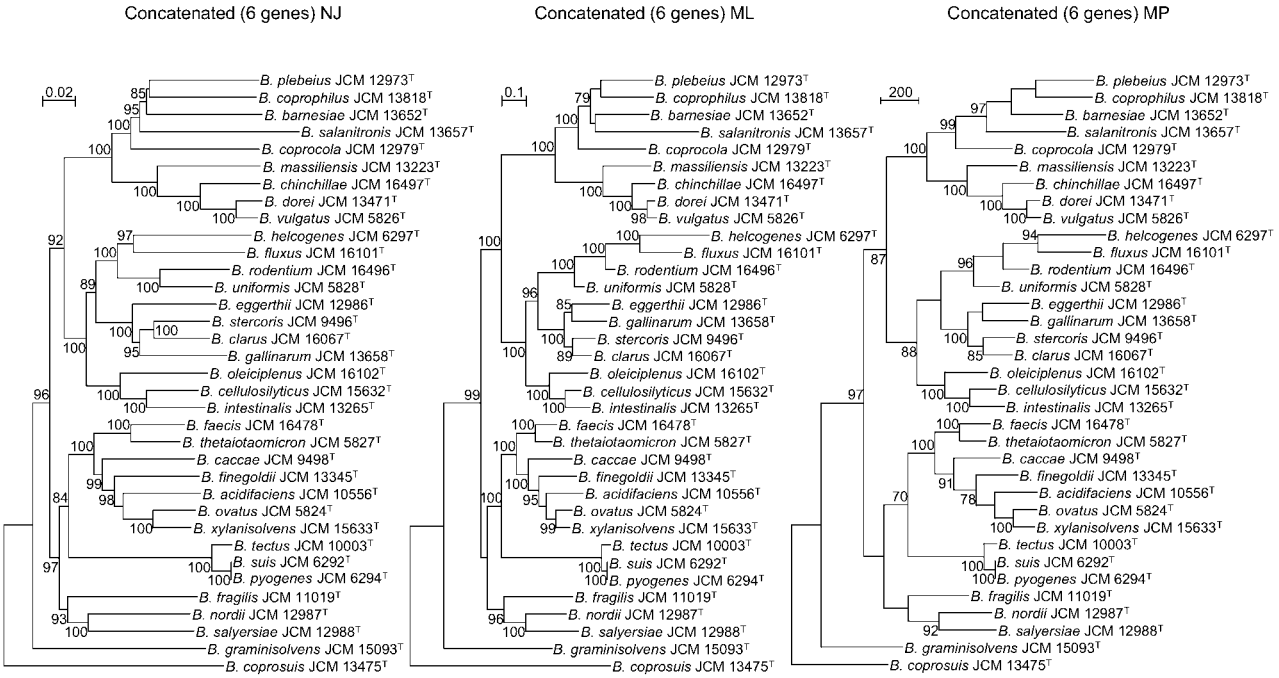

Fig. 1. Phylogenetic trees based on concatenated sequences of the *dnaJ*, *gyrB*, *hsp60*, *recA*, *rpoB* and 16S rRNA genes. The trees were constructed by the neighbour-joining (NJ), maximum-likelihood (ML) and maximum-parsimony (MP) methods. Numbers at nodes indicate the percentage bootstrap values of 1000 replicates (>70 %). Bars, 0.02 (NJ) or 0.1 (ML) expected nucleotide substitutions per site; 200 base changes between nodes (MP).

high (see Supplementary Fig. S2, available with the online version of this paper). In the NJ tree, the phylogeny inferred is essentially the same as that obtained with the six concatenated genes, except for the branching orders of the three clades Uniformis, Stercoris and Intestinalis (see below) and of *Bacteroides acidifaciens* and *B. finegoldii*. Nonetheless, the resolution, in terms of inter-species similarity, was superior for the three concatenated sequences (68.8–100 %; the pairwise differences ranged from 0 to 767 nt) to that obtained with the six concatenated sequences (74.9–100 %; the pairwise differences ranged from 0 to 1207 nt).

The SH and AU tests have already been performed to assess the congruence of the different gene phylogenies (Bilhère *et al.*, 2009; Keymer & Boehm, 2011). In this study, statistical comparisons of tree topologies performed by the SH and AU tests showed a significant incongruence ($P<0.05$) in many pairwise comparisons (see Supplementary Tables S2 and S3, available with the online version of this paper), indicating independent evolutionary mechanisms, and that the phylogeny of species of the genus *Bacteroides* cannot be inferred accurately from only one or a few genes. A likelihood score computed from the six concatenated genes was significantly congruent with that derived from the three concatenated sequences data ($P=0.984$ for the SH test and 0.679 for the AU test).

Significantly grouped taxa

As was expected, *B. pyogenes* JCM 6294^T robustly grouped with *B. suis* JCM 6292^T and *Bacteroides tectus* JCM 10003^T in all analyses (see Supplementary Fig. S1). The taxonomic

Table 3. Clades proposed by means of MLSA for *Bacteroides*

Clade	Described species included	No. of species	DDH value (%)	G+C content* (mol%)	MLSA concatenated similarity (%)	Phi test† (*P*-value)	Source
Coprosuis	*B. coprosuis* and *B. graminisolvens*	2	–	42–48	78.4	–	Rice-straw residue in a methanogenic reactor treating waste from cattle farms and swine-manure storage pit
Fragilis	*B. fragilis*, *B. nordii* and *B. salyersiae*	3	–	47–48	85.8–88.4	–	Human clinical specimens
Intestinalis	*B. intestinalis*, *B. cellulosilyticus* and *B. oleiciplenus*	3	<40‡	48	89.9–93.5	–	Human faeces
Ovatus	*B. ovatus*, *B. acidifaciens*, *B. caccae*, *B. finegoldii* and *B. xylanisolvens*	5	<42§	48–49	89.0–95.2	**3.9×10^{-5}**	Human faeces and mouse caecum
Plebeius	*B. plebeius*, *B. barnesiae*, *B. coprocola*, *B. coprophilus* and *B. salanitronis*	5	–	46–53	83.7–87.8	**2.7×10^{-2}**	Human faeces and chicken caecum
Pyogenes	*B. pyogenes*, *B. suis* and *B. tectus*	3	>70‖	51	97.5–100	–	Feline oral cavity and swine abscess and faeces
Stercoris	*B. stercoris*, *B. clarus*, *B. eggerthii* and *B. gallinarum*	4	–	51–53	90.1–93.5	6.5×10^{-2}	Human faeces and chicken caecum
Thetaiotaomicron	*B. thetaiotaomicron* and *B. faecis*	2	<22¶	48	93.9	–	Human faeces
Uniformis	*B. uniformis*, *B. fluxus*, *B. helcogenes* and *B. rodentium*	4	–	50–52	85.5–92.9	**5.9×10^{-14}**	Human faeces, chinchilla faeces, and swine abscess and faeces
Vulgatus	*B. vulgatus*, *B. chinchillae*, *B. dorei* and *B. massiliensis*	4	<53#	47–49	88.1–97.1	**5.6×10^{-17}**	Human faeces and blood, and chinchilla faeces

*Calculated based on the six genes.
†The Phi test was conducted for clades that included at least four species. Recombination ($P<0.05$) is indicated by bold type.
‡Data from Robert *et al.* (2007). DDH value between the pair *B. cellulosilyticus* and *B. intestinalis*.
§Data from Chassard *et al.* (2008). DDH value between the pair *B. ovatus* and *B. xylanisolvens*.
‖Data from Sakamoto *et al.* (2010).
¶Data from Kim *et al.* (2010).
#Data from Bakir *et al.* (2006c). DDH value between the pair *B. dorei* and *B. vulgatus*.

problem of these three species has already been resolved (Sakamoto *et al.*, 2010). In the NJ tree, *B. dorei* JCM 13471^T and *B. vulgatus* JCM 5826^T were grouped together with a high bootstrap value (91–100 %) in all analyses except the *hsp60* and *rpoB* gene phylogenies. In the cases of the *hsp60* and *rpoB* genes, *B. dorei* JCM 13471^T grouped with *B. chinchillae* JCM 16497^T. *Bacteroides rodentium* JCM 16496^T and *B. uniformis* JCM 5828^T were grouped together with high bootstrap value (86–100 %) in all analyses except the *gyrB* gene phylogeny. In the case of the *gyrB* gene, *B. rodentium* JCM 16496^T grouped with *Bacteroides fluxus* JCM 16101^T. *Bacteroides cellulosilyticus* JCM 15632^T and *B. intestinalis* JCM 13265^T were grouped together with high bootstrap value (88–100 %) in all analyses except the *rpoB* gene phylogeny. In the case of the *rpoB* gene, *B. cellulosilyticus* JCM 15632^T is closer to *B. intestinalis* JCM 13265^T and *Bacteroides oleiciplenus* JCM 16102^T. *B. faecis* JCM 16478^T and *B. thetaiotaomicron* JCM 5827^T were grouped together with high bootstrap value (84–100 %) in all analyses except the *hsp60* gene phylogeny. *B. ovatus* JCM 5824^T and *Bacteroides xylanisolvens* JCM 15633^T were grouped together with high bootstrap value (90–100 %) in all analyses except the *dnaJ* and *rpoB* gene phylogenies. In the case of the *dnaJ* gene, *B. ovatus* JCM 5824^T grouped with *B. finegoldii* JCM 13345^T; on the other hand, in the *rpoB* gene phylogeny, *B. ovatus* JCM 5824^T is closer to *B. acidifaciens* JCM 10556^T and *B. xylanisolvens* JCM 15633^T. Recombination between closely related species may occur for several loci.

The concatenated sequence (all six genes) similarity values for some pairs of above-mentioned closely related species within each clade (see below) were as follows: *B. cellulosilyticus* and *B. intestinalis*, 93.5 %; *B. ovatus* and *B. xylanisolvens* 95.2 %;, *B. pyogenes*/*B. suis* and *B. tectus*, 97.5 %; *B. faecis* and *B. thetaiotaomicron*, 93.9 %; and *B. dorei* and *B. vulgatus*, 97.1 %. DDH values of these pairs are shown in Table 3. The criterion of around 97.5 % concatenated

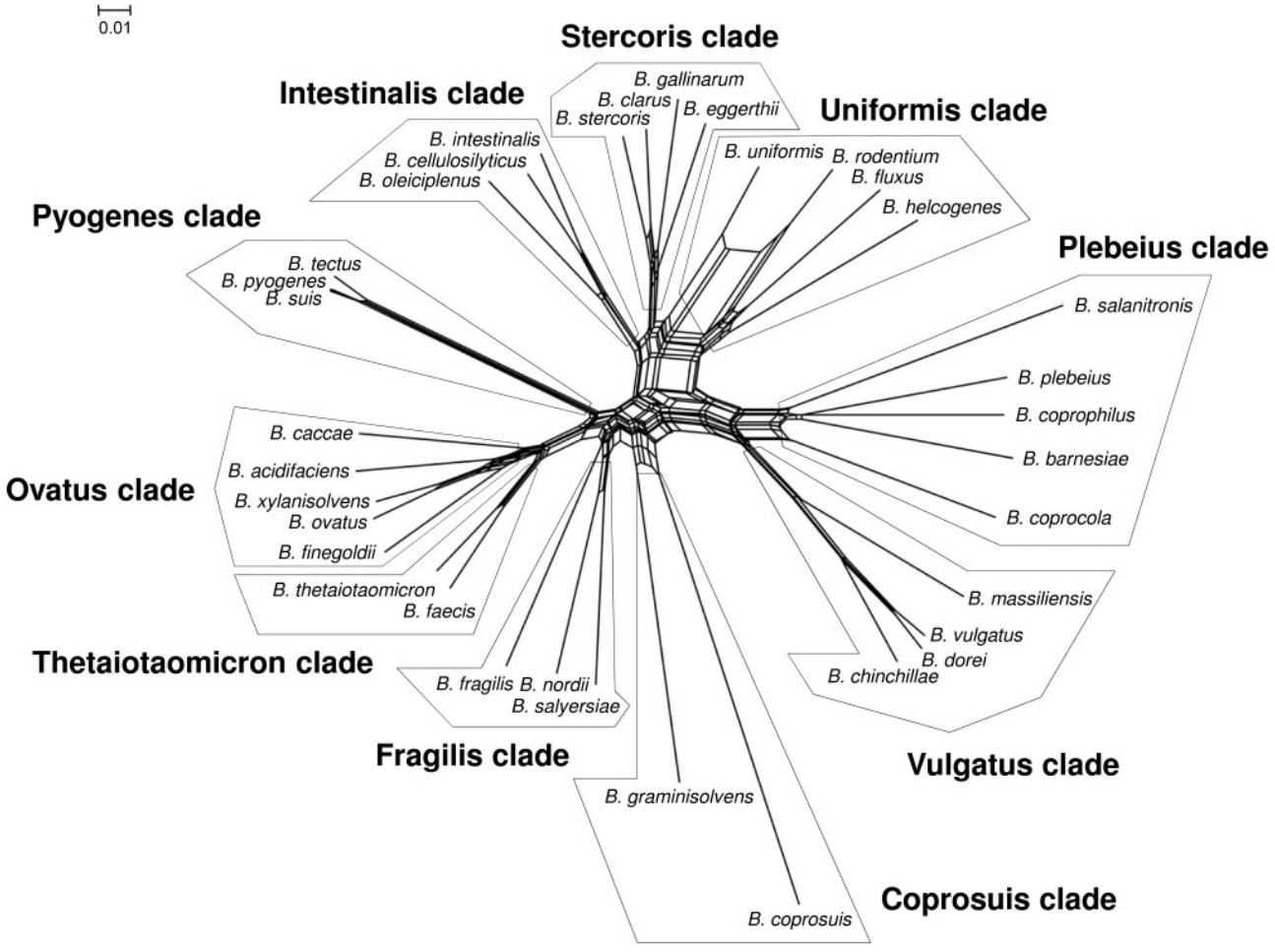

Fig. 2. Concatenated split network tree based on the six genes. The *dnaJ*, *gyrB*, *hsp60*, *recA*, *rpoB* and 16S rRNA gene sequences (4816 bp) from 35 *Bacteroides* species were concatenated and reconstructed using the SplitsTree4 program. Sequence similarities were corrected using the Jukes–Cantor correction. Bar, 0.01 expected nucleotide substitutions per site.

sequence similarity for defining species of the genus *Bacteroides* may be a useful idea because it has been reported that *B. pyogenes*, *B. suis* and *B. tectus* are a single species, on the basis of the *hsp60* gene sequence analysis and the levels of DNA–DNA relatedness (>70 %) (Sakamoto *et al.*, 2010).

Concatenated nucleotide sequences of six genes for 35 *Bacteroides* species were used to construct a Neighbour-Net split network with the EqualAngle algorithm. Ten distinct clades were recognized (Fig. 2). The Thetaiotaomicron clade was the sister group to the Ovatus clade. It has been reported that the previously described species closest to both of *Bacteroides nordii* and *Bacteroides salyersiae* was *B. thetaiotaomicron* (Song *et al.*, 2004). In this study, however, the clade of *B. nordii* and *B. salyersiae* was obviously sister to *B. fragilis* not *B. thetaiotaomicron* and these three species formed the Fragilis clade. Neighbour-Net analysis clearly separated the clades Intestinalis, Stercoris and Uniformis. The relationships of these clades have not been clear hitherto. These findings support the association between clustering and clade observed in the trees constructed with the concatenated alignment of six genes. The NJ, ML and MP trees (Fig. 1) were similar to that constructed using the SplitsTree4 program, except for the Stercoris (NJ tree), Plebeius and Uniformis (ML tree), and Uniformis (MP tree) clades. The Phi test has already been used to detect the presence of recombination (Sawabe *et al.*, 2007; Thompson *et al.*, 2007). In this study, the Phi test pointed to recombination within the three ($P=1.8\times10^{-7}$) and six ($P=8.9\times10^{-15}$) concatenated sequences. Recombination was detected in the *dnaJ* ($P=6.8\times10^{-3}$), *gyrB* ($P=1.6\times10^{-3}$) and 16S rRNA ($P=1.4\times10^{-5}$) genes. No recombination was detected in the *hsp60*, *recA* and *rpoB* genes. Other events might have occurred in the *hsp60* and *rpoB* genes because the positions of various pairs of species, e.g. *B. dorei* and *B. vulgatus*, *B. cellulosilyticus* and *B. intestinalis*, *B. faecis* and *B. thetaiotaomicron*, and *B. ovatus* and *B. xylanisolvens*, changed in the trees constructed from the *hsp60* and *rpoB* genes as mentioned above. Possibly, in these genes there occurred a small number of accumulated nucleotide substitutions leading to a close relationship between species. Recombination was detected in the Ovatus, Plebeius, Uniformis and Vulgatus clades at least (Table 3), in agreement with the conflicting phylogenetic splits (parallelograms) observed on the basis of the SplitsTree4 program (Fig. 2). The recombination analysis suggests that genes responsible for different essential functions in the cell may be targets of recombination, but we cannot rule out the possibility that the recombination tests are providing false-positive results.

The species found in each clade occupy similar niches. For example, the species in the Coprosuis clade were isolated from rice-straw residue in a methanogenic reactor treating waste from cattle farms and swine-manure storage pits, and were probably of cattle and swine faeces origin. In addition, the DNA G+C contents reported of *B. coprosuis* (36.4 mol%) (Whitehead *et al.*, 2005) and *B. graminisolvens* (38.0 mol%) (Nishiyama *et al.*, 2009) are significantly lower than that of other species of genus *Bacteroides* (40–49 mol%) (Sakamoto & Benno, 2006). The Fragilis clade comprises three species which show a narrow DNA G+C content range. The species in this clade were mainly isolated from human clinical specimens. Among *Bacteroides* species, *B. fragilis* strains are opportunistic pathogens (Sears, 2009). The Pyogenes clade contains only the species isolated from animals, i.e. felines and swine. On the other hand, the species in other clades are isolated from a variety of sources, i.e. human, chinchilla, chicken, mouse and swine (Table 3).

Conclusions

This study demonstrates that MLSA of housekeeping genes is a valuable alternative technique for the identification and classification of species of the genus *Bacteroides*. Strains of the same species should share at least 97.5 % concatenated sequence similarity using the fragments of six genes. Although six-gene MLSA is a reliable method for elucidating taxonomic data to inform *Bacteroides* species identification, three-gene (*dnaJ*, *gyrB* and *hsp60*) MLSA was nearly as reliable and would be more feasible for routine use in a clinical reference microbiology laboratory. To our knowledge, this is the first report on the MLSA of species of the genus *Bacteroides*, although type strains were mainly analysed. The future inclusion of more reference strains will enhance the value of this methodology.

ACKNOWLEDGEMENTS

We thank Ms Natsuko Suzuki and Ms Misako Matsuda for their technical assistance. We also thank Dr Hidetoshi Shimodaira, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, for help in using CONSEL and Dr Radhey S. Gupta, Department of Biochemistry and Biomedical Science, McMaster University, for helpful comments on the manuscript. This work was supported by a research grant from the Institute for Fermentation, Osaka (IFO, Osaka, Japan), (2009–2011) and also by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science to M. S (no. 23580126).

REFERENCES

Ah-You, N., Gagnevin, L., Grimont, P. A. D., Brisse, S., Nesme, X., Chiroleu, F., Bui Thi Ngoc, L., Jouen, E., Lefeuvre, P. & other authors (2009). Polyphasic characterization of xanthomonads pathogenic to members of the *Anacardiaceae* and their relatedness to species of *Xanthomonas*. *Int J Syst Evol Microbiol* **59**, 306–318.

Bakir, M. A., Kitahara, M., Sakamoto, M., Matsumoto, M. & Benno, Y. (2006a). *Bacteroides intestinalis* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* **56**, 151–154.

Bakir, M. A., Kitahara, M., Sakamoto, M., Matsumoto, M. & Benno, Y. (2006b). *Bacteroides finegoldii* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* **56**, 931–935.

Bakir, M. A., Sakamoto, M., Kitahara, M., Matsumoto, M. & Benno, Y. (2006c). *Bacteroides dorei* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* **56**, 1639–1643.

Bilhère, E., Lucas, P. M., Claisse, O. & Lonvaud-Funel, A. (2009). Multilocus sequence typing of *Oenococcus oeni*: detection of two

subpopulations shaped by intergenic recombination. *Appl Environ Microbiol* **75**, 1291–1300.

Bruen, T. C., Philippe, H. & Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681.

Bryant, D. & Moulton, V. (2004). Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* **21**, 255–265.

Bui Thi Ngoc, L., Vernière, C., Jouen, E., Ah-You, N., Lefeuvre, P., Chiroleu, F., Gagnevin, L. & Pruvost, O. (2010). Amplified fragment length polymorphism and multilocus sequence analysis-based genotypic relatedness among pathogenic variants of *Xanthomonas citri* pv. *citri* and *Xanthomonas campestris* pv. *bilvae*. *Int J Syst Evol Microbiol* **60**, 515–525.

Chassard, C., Delmas, E., Lawson, P. A. & Bernalier-Donadille, A. (2008). *Bacteroides xylanisolvens* sp. nov., a xylan-degrading bacterium isolated from human faeces. *Int J Syst Evol Microbiol* **58**, 1008–1013.

Enersen, M., Olsen, I., van Winkelhoff, A. J. & Caugant, D. A. (2006). Multilocus sequence typing of *Porphyromonas gingivalis* strains from different geographic origins. *J Clin Microbiol* **44**, 35–41.

Felsenstein, J. (1985). Confidence limits of phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., Stackebrandt, E., Van de Peer, Y., Vandamme, P. & other authors (2005). Opinion: re-evaluating bacterial species. *Nat Microbiol Rev* **3**, 733–739.

Grosse-Herrenthey, A., Maier, T., Gessler, F., Schaumann, R., Böhnel, H., Kostrzewa, M. & Krüger, M. (2008). Challenging the problem of clostridial identification with matrix-assisted laser desorption and ionization-time-of-flight mass spectrometry (MALDI-TOF MS). *Anaerobe* **14**, 242–249.

Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.

Hanage, W. P., Fraser, C. & Spratt, B. G. (2006). Sequences, sequence clusters and bacterial species. *Philos Trans R Soc Lond B Biol Sci* **361**, 1917–1927.

Hayashi, H., Shibata, K., Bakir, M. A., Sakamoto, M., Tomita, S. & Benno, Y. (2007). *Bacteroides coprophilus* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* **57**, 1323–1326.

Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254–267.

Jousimes-Somer, H. R., Summanen, P., Citron, D. M., Baron, E. J., Wexler, H. M. & Finegold, S. M. (2002). *Wadworth-KTL Anaerobic Bacteriology Manual*, 6th edn. Belmont, CA: Star Publishing Co.

Keymer, D. P. & Boehm, A. B. (2011). Recombination shapes the structure of an environmental *Vibrio cholerae* population. *Appl Environ Microbiol* **77**, 537–544.

Kim, M.-S., Roh, S. W. & Bae, J.-W. (2010). *Bacteroides faecis* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* **60**, 2572–2576.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111–120.

Kitahara, M., Sakamoto, M., Ike, M., Sakata, S. & Benno, Y. (2005). *Bacteroides plebeius* sp. nov. and *Bacteroides coprocola* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* **55**, 2143–2147.

Kitahara, M., Tsuchida, S., Kawasumi, K., Amao, H., Sakamoto, M., Benno, Y. & Ohkuma, M. (2011). *Bacteroides chinchillae* sp. nov. and *Bacteroides rodentium* sp. nov., isolated from chinchilla (*Chinchilla lanigera*) faeces. *Int J Syst Evol Microbiol* **61**, 877–881.

Ko, K. S., Kuwahara, T., Haehwa, L., Yoon, Y.-J., Kim, B.-J., Lee, K.-H., Ohnishi, Y. & Kook, Y.-H. (2007). RNA polymerase β-subunit gene (*rpoB*) sequence analysis for the identification of *Bacteroides* spp. *Clin Microbiol Infect* **13**, 48–54.

Lan, P. T. N., Sakamoto, M., Sakata, S. & Benno, Y. (2006). *Bacteroides barnesiae* sp. nov., *Bacteroides salanitronis* sp. nov. and *Bacteroides gallinarum* sp. nov., isolated from chicken caecum. *Int J Syst Evol Microbiol* **56**, 2853–2859.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). CLUSTAL W and CLUSTAL_X version 2.0. *Bioinformatics* **23**, 2947–2948.

Lee, C. S. & Lee, J. (2010). Evaluation of new *gyrB*-based real-time PCR system for the detection of *B. fragilis* as an indicator of human-specific fecal contamination. *J Microbiol Methods* **82**, 311–318.

Leon, A., Pronost, S., Fortier, G., Andre-Fontaine, G. & Leclercq, R. (2010). Multilocus sequence analysis for typing *Leptospira interrogans* and *Leptospira kirschneri*. *J Clin Microbiol* **48**, 581–585.

Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452.

Liu, C., Song, Y., McTeague, M., Vu, A. W., Wexler, H. & Finegold, S. M. (2003). Rapid identification of the species of the *Bacteroides fragilis* group by multiplex PCR assays using group- and species-specific primers. *FEMS Microbiol Lett* **222**, 9–16.

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K. & other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140–3145.

Margos, G., Vollmer, S. A., Cornet, M., Garnier, M., Fingerle, V., Wilske, B., Bormane, A., Vitorino, L., Collares-Pereira, M. & other authors (2009). A new *Borrelia* species defined by multilocus sequence analysis of housekeeping genes. *Appl Environ Microbiol* **75**, 5410–5416.

Martens, M., Delaere, M., Coopman, R., De Vos, P., Gillis, M. & Willems, A. (2007). Multilocus sequence analysis of *Ensifer* and related taxa. *Int J Syst Evol Microbiol* **57**, 489–503.

Martens, M., Dawyndt, P., Coopman, R., Gillis, M., De Vos, P. & Willems, A. (2008). Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). *Int J Syst Evol Microbiol* **58**, 200–214.

Mellmann, A., Cloud, J., Maier, T., Keckevoet, U., Ramminger, I., Iwen, P., Dunn, J., Hall, G., Wilson, D. & other authors (2008). Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria. *J Clin Microbiol* **46**, 1946–1954.

Nagy, E., Maier, T., Urban, E., Terhes, G., Kostrzewa, M. & ESCMID Study Group on Antimicrobial Resistance in Anaerobic Bacteria (2009). Species identification of clinical isolates of *Bacteroides* by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry. *Clin Microbiol Infect* **15**, 796–802.

Nishiyama, T., Ueki, A., Kaku, N., Watanabe, K. & Ueki, K. (2009). *Bacteroides graminisolvens* sp. nov., a xylanolytic anaerobe isolated from a methanogenic reactor treating cattle waste. *Int J Syst Evol Microbiol* **59**, 1901–1907.

Pascual, J., Macián, M. C., Arahal, D. R., Garay, E. & Pujalte, M. J. (2010). Multilocus sequence analysis of the central clade of the genus *Vibrio* by using the 16S rRNA, *recA, pyrH, rpoD, gyrB, rctB* and *toxR* genes. *Int J Syst Evol Microbiol* **60**, 154–165.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Mol Biol Evol* **25**, 1253–1256.

R Development Core Team (2009). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051–07–0, URL http://www.R-project.org.

Redondo, M. C., Arbo, M. D., Grindlinger, J. & Snydman, D. R. (1995). Attributable mortality of bacteremia associated with the *Bacteroides fragilis* group. *Clin Infect Dis* **20**, 1492–1496.

Rivas, R., Martens, M., de Lajudie, P. & Willems, A. (2009). Multilocus sequence analysis of the genus *Bradyrhizobium*. *Syst Appl Microbiol* **32**, 101–110.

Robert, C., Chassard, C., Lawson, P. A. & Bernalier-Donadille, A. (2007). *Bacteroides cellulosilyticus* sp. nov., a cellulolytic bacterium from the human gut microbial community. *Int J Syst Evol Microbiol* **57**, 1516–1520.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.

Sakamoto, M. & Benno, Y. (2006). Reclassification of *Bacteroides distasonis, Bacteroides goldsteinii* and *Bacteroides merdae* as *Parabacteroides distasonis* gen. nov., comb. nov., *Parabacteroides goldsteinii* comb. nov. and *Parabacteroides merdae* comb. nov. *Int J Syst Evol Microbiol* **56**, 1599–1605.

Sakamoto, M. & Ohkuma, M. (2010). Usefulness of the *hsp60* gene for the identification and classification of Gram-negative anaerobic rods. *J Med Microbiol* **59**, 1293–1302.

Sakamoto, M., Suzuki, M., Umeda, M., Ishikawa, I. & Benno, Y. (2002). Reclassification of *Bacteroides forsythus* (Tanner *et al.* 1986) as *Tannerella forsythensis* corrig., gen. nov., comb. nov. *Int J Syst Evol Microbiol* **52**, 841–849.

Sakamoto, M., Kitahara, M. & Benno, Y. (2007). *Parabacteroides johnsonii* sp. nov., isolated from human faeces. *Int J Syst Evol Microbiol* **57**, 293–296.

Sakamoto, M., Suzuki, N. & Benno, Y. (2010). *hsp60* and 16S rRNA gene sequence relationships among species of the genus *Bacteroides* with the finding that *Bacteroides suis* and *Bacteroides tectus* are heterotypic synonyms of *Bacteroides pyogenes*. *Int J Syst Evol Microbiol* **60**, 2984–2990.

Sawabe, T., Kita-Tsukamoto, K. & Thompson, F. L. (2007). Inferring the evolutionary history of vibrios by means of multilocus sequence analysis. *J Bacteriol* **189**, 7932–7936.

Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504.

Sears, C. L. (2009). Enterotoxigenic *Bacteroides fragilis*: a rogue among symbiotes. *Clin Microbiol Rev* **22**, 349–369.

Shah, H. N., Olsen, I., Bernard, K., Finegold, S. M., Gharbia, S. & Gupta, R. S. (2009). Approaches to the study of the systematics of anaerobic, Gram-negative, non-sporeforming rods: current status and perspectives. *Anaerobe* **15**, 179–194.

Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. *Syst Biol* **51**, 492–508.

Shimodaira, H. & Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* **16**, 1114–1116.

Shimodaira, H. & Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* **17**, 1246–1247.

Snydman, D. R., Jacobus, N. V., McDermott, L. A., Ruthazer, R., Golan, Y., Goldstein, E. J. C., Finegold, S. M., Harrell, L. J., Hecht, D. W. & other authors (2007). National survey on the susceptibility of *Bacteroides fragilis* group: report and analysis of trends in the United States from 1997 to 2004. *Antimicrob Agents Chemother* **51**, 1649–1655.

Song, Y. L., Liu, C. X., McTeague, M. & Finegold, S. M. (2004). ''*Bacteroides nordii*'' sp. nov. and ''*Bacteroides salyersae*'' sp. nov. isolated from clinical specimens of human intestinal origin. *J Clin Microbiol* **42**, 5565–5570.

Song, Y., Liu, C., Bolanos, M., Lee, J., McTeague, M. & Finegold, S. M. (2005). Evaluation of 16S rRNA sequencing and reevaluation of a short biochemical scheme for identification of clinically significant *Bacteroides* species. *J Clin Microbiol* **43**, 1531–1537.

Stackebrandt, E., Frederiksen, W., Garrity, G. M., Grimont, P. A. D., Kämpfer, P., Maiden, M. C. J., Nesme, X., Rosselló-Mora, R., Swings, J. & other authors (2002). Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**, 1043–1047.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24**, 1596–1599.

Tanigawa, K. & Watanabe, K. (2011). Multilocus sequence typing reveals a novel subspeciation of *Lactobacillus delbrueckii*. *Microbiology* **157**, 727–738.

Thompson, F. L., Gevers, D., Thompson, C. C., Dawyndt, P., Naser, S., Hoste, B., Munn, C. B. & Swings, J. (2005). Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis. *Appl Environ Microbiol* **71**, 5107–5115.

Thompson, F. L., Gomez-Gil, B., Vasconcelos, A. T. R. & Sawabe, T. (2007). Multilocus sequence analysis reveals that *Vibrio harveyi* and *V. campbellii* are distinct species. *Appl Environ Microbiol* **73**, 4279–4285.

Whitehead, T. R., Cotta, M. A., Collins, M. D., Falsen, E. & Lawson, P. A. (2005). *Bacteroides coprosuis* sp. nov., isolated from swine-manure storage pits. *Int J Syst Evol Microbiol* **55**, 2515–2518.

Yamamoto, S. & Harayama, S. (1995). PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl Environ Microbiol* **61**, 1104–1109.

Edited by: D. A. Mills