APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Nov. 2003, p. 6740–6749 Vol. 69, No. 11
0099-2240/03/$08.00+0 DOI: 10.1128/AEM.69.11.6740–6749.2003

Usefulness of *rpoB* Gene Sequencing for Identification of *Afipia* and *Bosea* Species, Including a Strategy for Choosing Discriminative Partial Sequences

Atieh Khamis, Philippe Colson, Didier Raoult, and Bernard La Scola*

Unité des Rickettsies, CNRS UPRESA 6020, Faculté de Médecine, Université de la Méditerrannée, 13385 Marseille Cedex 05, France

Received 23 June 2003/Accepted 18 August 2003

Bacteria belonging to the genera *Afipia* and *Bosea* are amoeba-resisting bacteria that have been recently reported to colonize hospital water supplies and are suspected of being responsible for intensive care unit-acquired pneumonia. Identification of these bacteria is now based on determination of the 16S ribosomal DNA sequence. However, the 16S rRNA gene is not polymorphic enough to ensure discrimination of species defined by DNA-DNA relatedness. The complete *rpoB* sequences of 20 strains were first determined by both PCR and genome walking methods. The percentage of homology between different species ranged from 83 to 97% and was in all cases lower than that observed with the 16S rRNA gene; this was true even for species that differed in only one position. The taxonomy of *Bosea* and *Afipia* is discussed in light of these results. For strain identification that does not require the complete *rpoB* sequence (4,113 to 4,137 bp), we propose a simple computerized method that allows determination of nucleotide positions of high variability in the sequence that are bordered by conserved sequences and that could be useful for design of universal primers. A fragment of 740 to 752 bp that contained the most highly variable area (positions 408 to 420) was amplified and sequenced with these universal primers for 47 strains. The variability of this sequence allowed identification of all strains and correlated well with results of DNA-DNA relatedness. In the future, this method could be also used for the determination of variability "hot spots" in sets of housekeeping genes, not only for identification purposes but also for increasing the discriminatory power of sequence typing techniques such as multilocus sequence typing.

Aquatic bacteria such as *Legionella*, *Pseudomonas*, *Stenotrophomonas*, *Burkholderia* spp., and *Acinetobacter* spp. may colonize hospital water supplies and have previously been shown to be causally associated with cases of nosocomial infections (23). Free-living amoebae have been shown to be a reservoir of pathogens, such as *Legionella* spp., *Burkholderia pickettii*, and *Cryptococcus neoformans* (2, 26). The most studied of amoeba-resisting bacteria (ARB) is *Legionella pneumophila*, the agent of Legionnaires' disease (27), which frequently results from exposure to contaminated aerosols. There are growing hints that additional ARB might be implicated in community-acquired pneumonia, including *Legionella*-like amoebal pathogens and members of the genus *Parachlamydia* (19). As part of the research into the diversity of bacterial agents associated with amoebae in hospital water supplies, we previously identified new α-proteobacteria belonging to the *Bradyrhizobiaceae* (13). Moreover, we demonstrated that patients with nosocomial pneumonia who were hospitalized in a public hospital where contaminated water was found had elevated titers of antibodies against these bacteria (14) and that patient seroconversion to *Bosea massiliensis* was frequent in patients hospitalized in intensive care units and was associated with the occurrence of ventilator-acquired pneumonia (17). Among the *Bradyrhizobiaceae*, bacteria of the genera *Bosea* and *Afipia* were the most frequently isolated. Due to the fastidiousness of these bacteria (3, 15, 16), identification is mostly based on 16S rRNA gene sequence (15, 16, 21). However, the 16S rRNA genes of these bacteria show very low variability: bacteria with only 1 base difference may belong to different species, as evidenced by DNA-DNA hybridization studies (15, 16, 25). To develop a molecular tool for both identification of cultured bacteria and detection from human samples, we decided to develop a sequence-based identification assay. Among the universal genes that can be used for this purpose, the RNA polymerase β-subunit-encoding gene (*rpoB*) was extensively used by our team for *Bartonella* spp. (22), *Staphylococcus* spp. (5), and *Enterobacteriaceae* (20), as well as for *Mycobacterium* (11) and *Legionella* spp. (12). The RNA polymerase β′ subunit is encoded by the *rpoC* gene. This gene has a low level of homology with *rpoB* and has been less studied for sequence-based identification. Herein we investigate the usefulness of *rpoB* sequencing for differentiation and identification of *Afipia* and *Bosea*. As *rpoB* is large (>4,000 bp), we also determined regions of variability in the sequence that are bordered by conserved sequences with the objective of designing universal primers for amplification of a small but discriminative sequence for routine *Afipia* and *Bosea* identification.

MATERIALS AND METHODS

Bacterial strains. The bacterial stains used in this study are listed in Table 1. These strains were routinely grown on buffered charcoal-yeast extract agar plates (bioMérieux, Marcy l'Étoile, France) as previously described (13).

***rpoB* gene amplification and sequencing.** The sequences of *rpoB* from the most closely related species of the studied bacteria were aligned in order to produce a consensus sequence. The chosen bacteria were *Sinorhizobium meliloti*, *Meso-*

* Corresponding author. Mailing address: Unité des Rickettsies, CNRS UPRESA 6020, Faculté de Médecine, Université de la Méditerrannée, 27 Blvd. Jean Moulin, 13385 Marseille Cedex 05, France. Phone: 33.04.91.38.55.17. Fax: 33.04.91.83.03.90. E-mail: bernard.lascola@medecine.univ-mrs.fr.

TABLE 1. List of the species for which *rpoB* partial or complete sequences were determined

Species	Strain	Reference	GenBank accession no. for:		Size (bp) of complete *rpoB*
			16S rRNA	*rpoB*	
Afipia felis	B-91-007352^T (ATCC 53690)	3	M65248	AY242824	4,137
	B-91-007147 (ATCC 49714)	3			
	B-90-007209 (ATCC 49715)	3			
	B-90-007260 (ATCC 49716)	3			
Afipia felis genospecies A	76713^T (CIP 106335, CCUG 43109)	15	AF374383	AY242825	4,134
Afipia clevelandensis	B-91-007353^T (ATCC 49720)	10	M69186	AY242823	4,131
Afipia broomeae	B-91-007286^T (ATCC 49717)	3	U87761	AY242822	4,119
	B-91-007288 (ATCC 49718)	3		AY310373	
	B-91-007289 (ATCC 49719)	3			
Afipia genospecies 1	B-91-007287^T (ATCC 49721)	3	U87763	AY242828	4,119
Afipia genospecies 2	B-91-007290^T (ATCC 49722)	3	U87765	AY242829	4,122
Afipia genospecies 3	B-91-007291^T (ATCC 49723)	3	U87766	AY242827	4,125
Afipia genospecies 3-related strains	34626 (CIP 106343; CCUG 43110)	15	AF288303	AY242826	4,119
	34631	15			
Afipia birgiae	34632^T (CIP 106344, CCUG 43108)	15	AF288304	AY242821	4,131
Afipia massiliensis	34633^T (CIP 107022, CCUG 45153)	15	AY029562	AY242820	4,128
Bosea thiooxidans	BI-42^T (DSM 9653)	4	AF508803	AY242832	4,122
Bosea massiliensis	63287^T (CIP 106336, CCUG 43117)	16	AF288309	AY242837	4,113
	34649 (CIP 106337, CCUG 43116)	16	AF288307	AY242836	4,113
	Isolate 18	17			
	Isolate 21	17			
	Isolate 40	17		AY310370	
	Isolate 44	17			
	Isolate 72	17			
	Isolate 74	17			
	Isolate 77	17			
	Isolate 79	17			
	Isolate 95	17			
	Isolate 238	17			
	Isolate 286	17			
Bosea eneae	34614^T (CIP 106338, CCUG 43111)	16	AF288300	AY242835	4,119
	34617 (CIP 106342, CCUG 43112)	16	AF288305	AY242841	1,241
Bosea vestrisii	34635^T (CIP 106340, CCUG 43114)	16	AF288306	AY242834	4,119
	34620 (CIP 106341, CCUG 43113)		AF288302	AY242840	
	63286 (CIP 106339, CCUG 43115)		AF288308	AY242839	
Bosea minatitlanensis	AMX51^T (CIP 106457, ATCC 700918)	20	AF273081	AY242833	4,122
Bosea sp.	7F		AF531764	AY242838	4,122
Bradyrhizobium lioaningense	ESG2281^T (CIP 104858, ATCC 700350)	34	AF363132	AY242831	4,125
	Isolate 22	17		AY310371	
	Isolate 26	17			
	Isolate 27	17			
	Isolate 30	17			
	Isolate 67	17			
	Isolate 128	17			
	Isolate 234	17			
	Isolate 93	17		AY310372	
Bradyrhizobium japonicum	3I1b6 (CIP 106093, ATCC 10324)	33	S46916	AY242830	4,119

rhizobium loti, *Bartonella henselae*, and *Bartonella quintana* (GenBank accession numbers SME591787, AP002994, AF171070, and AF165994, respectively). The consensus sequence was used to generate primers that were used in PCRs, for genome walking (24), and for sequencing. Additional primers were selected from ongoing base sequence determinations. All primers used in this study are summarized in Table 2. Bacterial DNA was extracted from a heavy suspension of strains with the QIAamp blood kit (Qiagen, Hilden, Germany) according to manufacturer's recommendations. All PCR mixtures contained 2.5×10^{-2} U of *Taq* polymerase per μl; 1× *Taq* buffer; 1.8 mM $MgCl_2$ (Gibco BRL, Life Technologies, Cergy Pontoise, France); 200 μM concentrations of dATP, dCTP, dTTP, and dGTP (Boehringer Mannheim GmbH, Hilden, Germany); and 0.2 μM concentrations of all primers (Eurogentec, Seraing, Belgium). PCR mixtures were subjected to 35 cycles of denaturation at 94°C for 30 s, primer annealing for 30 s (at a temperature 5°C below the melting temperature [T_m] of the primer with the lowest T_m), and extension at 72°C for 2 min. Every amplification program began with a denaturation step of 95°C for 2 min and ended with a final elongation step of 72°C for 10 min. Complete determination of the *rpoB* sequence ends was achieved by using the sequences of both 3′ and 5′ ends of the gene and amplifying by PCR using the Universal GenomeWalker kit (Clontech Laboratories, Palo Alto, Calif.). Briefly, genomic DNA was digested with *Eco*RV, *Dra*I, *Pvu*II, *Stu*I, and *Sca*I. DNA fragments were ligated with a GenomeWalker adaptor, which had one blunt end and one end with a 5′ overhang. The ligation mixture of the adaptor and the genomic DNA fragments were used as a template for PCR. This PCR was performed with an adaptor primer supplied by the manufacturer and specific primers to walk downstream the DNA sequence. For the amplification, 1.5 U of ELONGASE (Boehringer Mannheim) was used with a mixture containing 10 pmol of each primer, 20 mM (each) deoxynucleoside triphosphate, 10 mM Tris-HCl, 50 mM KCl, 1.6 mM $MgCl_2$, and 5 μl of template in a final volume of 50 μl. Genome walking was performed with the Universal GenomeWalker kit according to the manufacturer's recommendations. Amplicons were purified for sequencing with a QIAquick spin PCR purification kit (Qiagen) by following the protocol of the supplier. Sequencing reactions were carried out with the reagents of the ABI Prism 3100 DNA sequencer (dRhod.Terminator RR Mix; Perkin-Elmer Applied Biosystems) by following the standard automated-sequencer protocol.

TABLE 2. Primers used for amplification and sequencing of the entire *rpoB* gene in this study

Primer name	Primer sequence (5′–3′)	Position[a]	T_m (°C)
B-STAR1	GAGGAACAACATGGTCAATT	−10	56
STAR-WLK-F	GCDCGTCADTTTGCGTCT	−143	56
STAR-FW2	AATGGRRRCMACGATGG	−13	52
159F	CGYARRCGYGTACGCAAGTT	30	62
226R	GATCGTAGGATGCCTTCTGAA	97	62
228F	GAAGGCATCCTACGATCAGTT	100	62
240F	TAYGAYCAGTTCCTSATGGT	110	58
350R	CCATGTAGACRTCCTGCTCCTTGAT	376	72
460R	TCGTCGATATCGAACACGAT	328	58
517F	GACGTCTACATGGGCGATAT	387	60
524F	CATGGGCGATATGCCTTTAAT	395	60
558F	GAGTTCGACGCCAAGGACAT	584	62
BOS558F	AGTTCGACGCCAAGGACAT	585	58
BOS1190F	CATGTTCCAGTCGCTGTTCT	1180	60
1192F	CATGTTCCAGTCGCTGTTCTT	1180	62
1192R	GAAGAACAGCGACTGGAACAT	1181	62
1500F	AAGGGCGARATCGACGACAT	1347	60
1500R	ATGTCGTCGATYTCGCCCTT	1347	60
BOS1574F	TCGCAGTTCATGGACCAGA	1566	58
BOS1574R	TCTGGTCCATGAACTGCGA	1566	58
1700F	TCGCAGCTSTCGCAGTTCAT	1557	62
1700R	GTCCATGAACTGCGASAGCT	1561	62
1844F	CCGATTGAAACGCCGGAAGG	1710	64
1844R	CCTTCCGGCGT(TC)TC(AG)ATCGG	1710	66
1875R	TTGCGAGCGAGTTGATGAGA	1743	60
2204F	CTGATGGGCTCGAACATGCA	2070	62
2207R	GCTGCATGTTCGAGCCCAT	2074	60
2213R	CCTGACGCTGCATGTTCGA	2080	60
2216R	CGGCCTGACGCTGCATGTT	2083	62
BOS2080F	TGATGGGCTCGAACATGCA	2071	58
BOS2080R	TGCATGTTCGAGCCCATCA	2071	58
2260R	CGACTTGGTCGGRTCGAGAT	2249	64
2400F	CAACGTGCTCGTCGCGTTCA	2420	64
2400R	TGAACGCGACGAGCACGTTG	2420	64
2530F	GATCCACATCGAGGAATTCGAA	2520	64
2530R	CTTCGAATTCCTCGATGTGGAT	2521	64
2540F	CGGTCCGTCGACCGATCT	2385	60
BOS2560R	ACACGTTCGGAATGTCGCGCGT	2581	70
BOS2590F	AAGAGGCGCTGAAGAACCTCGA	2605	68
BOS2590R	TCGAGGTTCTTCAGCGCCTCTT	2605	68
2860F	CCGATGACGCCGGAAGAAA	2710	60
2860R	GCTTTTCTTCCGGCGTCAT	2713	58
BOS3030R	CACCACTGCGAGCGTGGGAA	3043	66
BOS3100R	TTCTTCGACTCATCGTACTGCT	3116	66
3177F	ATCCGCGCTCGCAGTGGT	3044	60
3184F	CCSGGCGTGATGAAGATGGTCAA	3199	72
3190F	GCGTGATGAAGATGGTCAARGTCT	3203	72
3210F	TCGTCGCGGTGAAGCGCAAGAT	3227	70
3320R	GCTGMASCTTYTCGACCTT	3163	58
3550F	GCATGAAYGTCGGBCAGAT	3386	58
3740F	GGCTACATCTAYATGCTSAAGCTG	3754	72
3745WR	GGTGCAGCTTSAGCATRTAGATGTA	3757	76
3920F	GTGACSGTGGGCTAATCTAYAT	3745	72
4010R	GCTGGGTRACGAGCGAGTA	3822	64
4026R	AGWGCCCARACCTCCATTT	3887	60
4078R	TCCAYTTSACYGTCAGCAT	3940	60
STOP-WR	CTCCTHGCGTGCCGATHGT	−98	62
STOP-ST-R	GATCGTCACCGGCAGCAA	−80	58
B-STOP1	GCCGAAAAGGTTCATGACCT	−85	60
B-STOP2	CGSCTTABTCBGCBGCCT	4124	60
A-STOP1	TACTCDGCHGCCTCBGASGT	4120	64
Br3200F	TGAAGATGGTCAAGGTCTTCGT	3208	64
Br3950R	GTCCGACTTSACHGTCAGCAT	3940	64

[a] Position relative to the *A. felis rpoB* sequence.

***rpoB* sequence analysis.** The nucleotide sequences of the *rpoB* gene fragments obtained were processed into sequence data with Sequence Analysis software (Applied Biosystems), and partial sequences were combined into a single consensus sequence with Sequence Assembler software (Applied Biosystems). Multiple sequence alignments were made, and percentages of similarity among the different species with *rpoB* and the 16S rRNA gene were obtained, with CLUSTALW (28) on the EMBL-EBI World Wide Web server (http://www.ebi.ac.uk/clustalw/). Phylogenetic trees were obtained from DNA sequences by three different methods: neighbor joining, maximum parsimony, and maximum likelihood (6). Bootstrap replicates were performed in order to estimate the node reliability of the phylogenetic trees obtained. Bootstrap values were obtained from 1,000 trees generated randomly with SEQBOOT in the PHYLIP software package.

Strategy for determination of discriminative partial sequences. To search for parts of sequences with high variability bordered by conserved regions, we created a simple analysis tool on Microsoft Excel 97 software that analyzes, reveals, and graphically represents variability along nucleotide sequences. This program (SVARAP, for sequence variability analysis program) can analyze and simultaneously process sets of up to 100 sequences of a maximal length of 4,000 nucleotides (hypertext link "Téléchargement" at the URL http://ifr48.free.fr/recherche/jeu_cadre/jeu_rickettsie.html). All sequences of our set of sequences (including the sequence used as an outgroup) were aligned with ClustalX, version 1.8 (29). The program calculates the consensus nucleotide (defined as the most frequent nucleotide at a site in the studied set of sequences), the absolute number of each of four nucleotides (G, A, C, and T) or the number of deletions or insertions, and their frequencies (percentages). The variability is considered the proportion of sequences for which the nucleotide at a position is different from the nucleotide found in the consensus sequence generated from the set of studied sequences. It is generated by the following formula: 100 − the maximum frequency for each of the four nucleotides at a given position. The program also calculates the number of nucleotides of different nature that are present at a given site. All these data are available in different sheets in tables or plotted in graphical windows. The data are then processed to calculate for a window of 60 nucleotides median, mean, and highest and lowest variability, with standard deviations.

After this analysis was done, the most variable area in *rpoB* was identified, and a primer pair designed from the border conserved area was used for PCR amplification of this area. PCR conditions that incorporated this consensus primer pair (Br3200F-Br3950R; Table 2) were those described above. These primers were used for amplification of the hypervariable region for all the strains for which complete *rpoB* sequences were previously determined and 27 additional strains (Table 1). Amplified fragments were then sequenced with the same primers under conditions described above.

Nucleotide sequence accession numbers. GenBank accession numbers for 16S rRNA and *rpoB* sequences obtained in this study are listed in Table 1.

RESULTS

Determination of *rpoB* sequences in *Afipia*, *Bosea*, and *Bradyrhizobium* species. The *rpoB* sequences varied in length, the longest being that of *Afipia felis*, with 4,137 bp, and the shortest being that of *B. massiliensis*, with 4,113 bp (Table 1). The percentage of homology between different species ranged from 83 to 97% (Table 3). It was always lower for *rpoB* than for 16S ribosomal DNA (rDNA), even for species not well discriminated by 16S rRNA gene sequences (Table 3). In the *Afipia* genus, *A. birgiae* and *A. massiliensis*, which have 99% homology with 16S rRNA gene sequences, have only 96% homology in *rpoB*. Nearly all members of the genus *Bosea* that have homologies above 98% for the 16S rRNA gene have homologies that range from 90 to 92% in *rpoB*. The exceptions in the genus are *B. eneae* and *B. vestrisii*, which have only 97% homology in *rpoB* but whose 16S rRNA gene sequences differ by only 1 position. The phylogenetic trees constructed with the different methods have the same topology except for the relations between *Bradyrhizobium* spp. and the group of the three *Afipia* genospecies. Bacteria of the genus *Bosea* form a group independent from *Afipia* (Fig. 1). *B. vestrisii* and *B. eneae* are separated from other *Bosea* species, whereas the recently described *B. minatitlanensis* is closely related with *B. thiooxidans*. *Bosea* sp. strain 7F appears as a well-separated species. In the group of *Afipia*, a cluster that contains *A. massiliensis*, *A. birgiae*, *A. broomeae*, *A. clevelandensis*, *A. felis*, and *A. felis* genospecies A is well separated from other species with high bootstrap values. The two *Bradyrhizobium* spp. cluster together as *Afipia* genospecies 1 and 2. The positions of *Afipia* genospecies 3 strains vary with the technique used to construct the tree and are never supported by high bootstrap values.

Strain identification with discriminative partial sequences. Study of sequence variability allowed detection of four highly variable sequences bordered by conserved regions (Fig. 2). These regions were between positions 481 and 1141, 1741 and 2041, 2881 and 3241, and 3361 and 3841. As the last region was the most variable (no. 4 in Fig. 2), especially the central part of 408 to 420 bp from position 3380 to position 3800 of the sequence of *A. felis* (AY242824), taken as reference, we designed a consensus primer pair (Br3200F-Br3950R; Table 2) that allowed amplification of a 740- to 752-bp fragment that contains the 408- to 420-bp hypervariable region in all species. Sizes of the amplified fragment and hypervariable region vary according to the species. The hypervariable region was determined for all the strains for which a complete *rpoB* sequence was determined and for 27 additional strains, 3 of *A. felis*, 2 of *A. broomeae*, 1 of *B. eneae*, 2 of *B. vestrisii*, 11 of *B. massiliensis*, and 8 of *Bradyrhizobium liaoningense* (Table 1). With the exception of those for *Afipia* genospecies 1 and 2, the percentages of homology observed with the 408- to 420-bp partial *rpoB* were always lower than those observed with the complete sequence (Table 4). Interestingly, homology between *B. eneae* and *B. vestrisii* was lowered to 96%. Among strains belonging to the same species the homology for this fragment ranged from 98 to 100%. The homologies between strains 34614 and 34617^T of *B. eneae*, between strains 34620 and 34635^T of *B. vestrisii*, between strain 34649 and isolates 18 and 21 of *B. massiliensis*, and between all strains of *A. felis* were 100%. The homology was 98% between strain 34649, strain 63287^T, and isolates 40 to 286 of *B. massiliensis*, which had the same sequence; between strain B-91-007286^T of *A. broomeae* and the two other strains of this species that had the same sequence; and between strains 34635^T and 63286 of *B. vestrisii*. The homology between strain ESG2281^T and isolate 93 of *B. japonicum* was 99%, and that between ESG2281^T and all other isolates of *B. japonicum* that shared the same partial sequence was 98%. The trees constructed by using the hypervariable region have the same topologies as those obtained with the complete sequence, but bootstrap values are lower and the distribution of some *Bosea* spp. is modified (Fig. 3). However, different species remain clearly differentiated.

DISCUSSION

The description of new bacterial species is currently based on results of DNA-DNA hybridization and phenotypic characters, so-called polyphasic classification data (8, 32). This method of classification has two major drawbacks: the difficulty of performing DNA-DNA hybridization, which is an expensive, technically complex, and labor-intensive procedure, and the scarcity of reproducible and distinguishable phenotypic char-

TABLE 3. Percent homology observed between *Afipia* and *Bosea* species according to the gene analyzed

Species[a]	% Homology for 16S rRNA gene/complete *rpoB* sequence with species[a]:																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	100/100	99/94	98/90	97/90	96/89	96/89	97/88	95/88	95/88	98/90	98/90	92/83	92/81	92/82	92/83	92/84	92/83	91/82	97/89	97/89
2		100/100	98/90	97/89	96/89	96/89	97/88	96/88	96/88	98/89	98/89	92/83	92/81	92/81	92/83	92/84	92/83	92/83	97/89	98/89
3			100/100	97/93	96/88	96/88	97/88	97/88	97/88	98/93	97/93	92/82	92/81	92/81	92/83	92/83	92/83	91/82	97/89	98/90
4				100/100	96/88	96/88	96/89	96/89	96/89	99/94	98/94	92/82	92/81	92/82	92/83	91/83	92/83	92/82	97/90	97/90
5					100/100	99/97	96/90	95/88	95/88	96/88	96/87	91/84	91/83	91/84	91/84	91/84	91/84	90/84	95/91	96/91
6						100/100	96/91	95/89	95/89	96/88	96/87	91/84	91/84	91/84	90/85	90/85	90/85	90/84	95/91	95/91
7							100/100	97/89	97/89	97/89	97/88	92/83	91/83	91/83	91/84	91/84	91/84	91/82	98/90	98/90
8								100/100	100	96/88	96/88	91/82	91/82	91/82	90/83	90/83	91/83	91/82	96/89	97/89
9									100/100	96/88	96/88	91/82	91/82	91/82	90/83	90/83	91/83	91/82	96/89	97/89
10										100/100	99/96	92/82	92/81	92/81	92/83	92/83	92/83	92/82	97/89	97/90
11											100/100	92/82	92/80	92/81	92/83	92/82	92/82	92/82	97/89	97/90
12												100/100	98/92	98/92	99/91	99/91	99/94	98/92	92/84	92/84
13													100/100	99/98	99/91	99/91	98/91	98/90	92/83	92/83
14														100/100	98/91	98/91	98/91	98/90	92/83	92/84
15															100/100	99/97	98/92	98/91	92/84	92/84
16																100/100	98/92	98/91	91/84	92/84
17																	100/100	98/91	91/85	91/84
18																		100/100	91/83	91/83
19																			100/100	99/96
20																				100/100

[a] Species numbering is as follows: 1, *A. felis*; 2, *A. felis* genospecies A; 3, *A. clevelandensis*; 4, *A. broomeae*; 5, *Afipia* genospecies 1; 6, *Afipia* genospecies 2; 7, *Afipia* genospecies 3; 8, *Afipia* genospecies 3-related strain (34626); 9, *Afipia* genospecies 3-related strain (34631); 10, *A. birgiae*; 11, *A. massiliensis*; 12, *B. thiooxidans*; 13, *B. massiliensis* (63287); 14, *B. massiliensis* (34649); 15, *B. eneae* (34614); 16, *B. vestrisii* (34635^T); 17, *B. minatitlanensis*; 18, *Bosea* sp. strain 7F; 19, *Bradyrhizobium liaoningense*; 20, *Bradyrhizobium japonicum*.

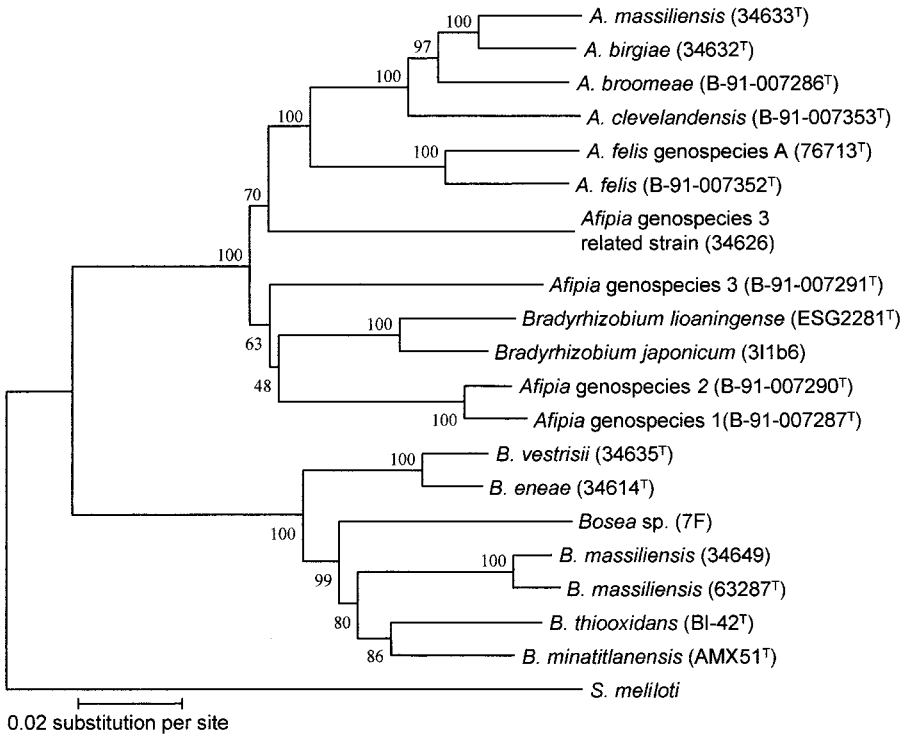

FIG. 1. Dendrogram representing phylogenetic relationships of *Afipia* and *Bosea* by the neighbor-joining method. The tree was derived from alignment of complete *rpoB* sequences. The support of each branch, as determined from 1,000 bootstrap samples, is indicated by the value at each node (in percent). *Sinorhizobium meliloti* was used as an outgroup.

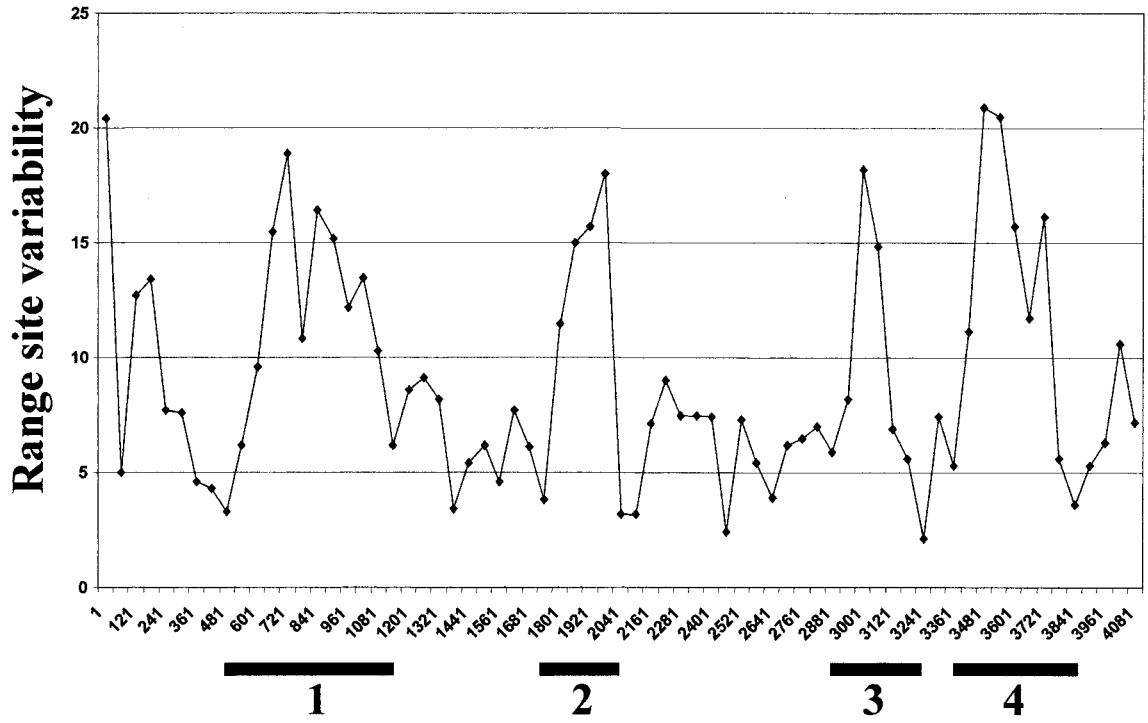

FIG. 2. Graphical representation of range site variability (*y* axis) in *rpoB* sequences of species studied per window of 60 nucleotides (*x* axis: position). Hypervariable regions bordered by conserved regions are numbered from 1 to 4.

TABLE 4. Comparison of the percent homology observed between *Afipia* and *Bosea* species according to the size of the *rpoB* gene studied

Species[a]	% Homology for complete *rpoB* sequence/partial hypervariable 420-bp *rpoB* sequence with species[a]:																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	100/100	94/94	90/87	90/83	89/78	89/79	88/79	88/83	88/83	90/85	90/83	83/79	81/76	82/77	83/77	ND[b]/77	84/79	83/78	82/77	89/85	89/84
2		100/100	90/84	89/85	89/78	89/78	88/78	88/83	88/83	89/82	89/83	83/76	81/74	81/75	83/74	ND/75	84/76	83/75	83/77	89/83	89/82
3			100/100	93/84	88/81	88/81	88/81	88/81	88/81	93/90	93/89	82/77	81/76	81/77	83/76	ND/76	83/77	83/78	82/76	89/84	89/85
4				100/100	88/79	88/78	89/77	89/80	89/80	94/85	94/87	82/73	81/71	82/72	83/73	ND/73	83/73	83/72	82/72	90/82	90/81
5					100/100	97/97	90/85	88/78	88/78	88/80	87/80	84/74	83/74	84/74	84/74	ND/74	84/75	84/74	84/73	91/85	91/85
6						100/100	91/86	89/79	89/79	88/81	87/80	84/74	84/74	84/75	85/75	ND/75	85/76	85/75	84/73	91/85	91/85
7							100/100	89/81	89/81	89/80	88/77	83/75	83/75	83/77	84/76	ND/75	84/76	84/74	82/75	90/83	90/83
8								100/100	100	88/81	88/80	82/72	82/73	82/74	83/75	ND/74	83/75	83/73	82/75	89/81	89/81
9									100/100	88/81	88/80	82/72	82/73	82/74	83/75	ND/74	83/75	83/73	82/75	89/81	89/81
10										100/100	96/91	82/77	81/75	81/76	83/76	ND/77	83/77	83/77	82/75	89/83	90/84
11											100/100	82/74	80/73	81/74	83/75	ND/75	82/75	82/74	82/74	89/83	90/84
12												100/100	92/87	92/87	91/88	ND/89	91/89	94/92	92/84	84/77	84/77
13													100/100	98/98	91/90	ND/89	91/90	91/90	90/83	83/76	83/75
14														100/100	91/90	ND/90	91/91	91/92	90/83	83/77	84/77
15															100/100	ND/96	97/96	92/89	91/83	84/77	84/77
16																100/100	ND/98	ND/89	ND/83	ND/77	ND/78
17																	100/100	92/90	91/83	84/78	84/78
18																		100/100	91/83	85/79	84/78
19																			100/100	83/76	83/75
20																				100/100	96/95
21																					100/100

[a] Species numbering is as follows: 1, *A. felis*; 2, *A. felis* genospecies A; 3, *A. clevelandensis*; 4, *A. broomeae*; 5, *Afipia* genospecies 1; 6, *Afipia* genospecies 2; 7, *Afipia* genospecies 3; 8, *Afipia* genospecies 3-related strain (34626); 9, *Afipia* genospecies 3-related strain (34631); 10, *A. birgiae*; 11, *A. massiliensis*; 12, *B. thiooxidans*; 13, *B. massiliensis* (63287); 14, *B. massiliensis* (34649); 15, *B. eneae* (34614); 16, *B. vestrisii* (63286); 17, *B. vestrisii* (34635^T); 18, *B. minatitlanensis*; 19, *Bosea* sp. strain 7F; 20, *Bradyrhizobium liaoningense*; 21, *Bradyrhizobium japonicum.* The sequence homologies for *B. eneae* strain 34617 are not presented in this table as they are identical to those of strain 34614.

[b] ND, not done.

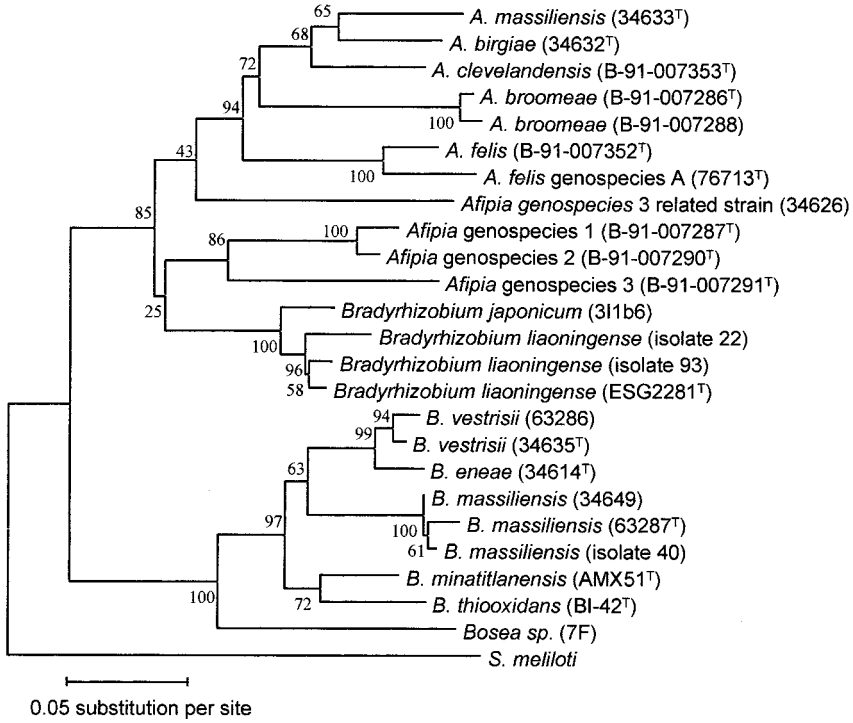

FIG. 3. Dendrogram representing phylogenetic relationships of *Afipia* and *Bosea* by the neighbor-joining method. The tree was derived from alignment of partial *rpoB* sequences. The support of each branch, as determined from 1,000 bootstrap samples, is indicated by the value at each node (in percent). *Sinorhizobium meliloti* was used as an outgroup.

acteristics for several bacterial species. The development of gene amplification and sequencing, especially that of the 16S rRNA gene sequences, has simplified the identification and the detection of fastidious bacteria, especially those lacking distinguishable phenotypic characteristics. However, as previously described for several species, including *Bacillus* spp. (1, 7), the 16S rDNA gene alone is not variable enough to allow confident discrimination between different species in some genera. This is the case for bacteria that belong to the genus *Afipia* and *Bosea* which we recently described (15, 16). For example, *A. felis* and *A. felis* genospecies A represent two distinct genospecies on the basis of DNA-DNA hybridizations and phenotypic data such as susceptibility to antibiotics, sodium dodecyl sulfate-polyacrylamide gel electrophoresis profile, and whole-cell fatty acid composition (8, 32), but they exhibit levels of 16S rRNA gene sequence similarity of 99.9%. As the comparison of 16S rDNA gene sequences is not sensitive enough for the reliable delineation of several species, comparison of sequences from a more divergent part of the genome, such as the *rrs-rrl* intergenic spacer, is more suitable, and this approach has been used for other members of the α-proteobacterium subgroup, including *Nitrobacter* spp. (9) and *Bradyrhizobium* (30). Our data, based on *rpoB* sequences of these bacteria, confirm that this gene is probably polymorphic enough to replace or supplement the 16S rRNA gene for definitive identification of *Afipia* and *Bosea* bacteria, as the two closest bacteria by 16S rRNA gene comparisons, with 1 different position (<0.1%), differ by at least 3% with *rpoB*. The results of *rpoB* sequencing support our proposal for removing *Afipia* genospecies 1 and 2 from the genus *Afipia* but still do not allow definition of the positions of *Afipia* genospecies 3 and related strains (15). The *rpoB* sequences of *A. felis* and *A. felis* genospecies A that have homology of only 94% are in agreement with results of DNA-DNA hybridization and clearly confirm that these are different species. Sequencing *rpoB* could also help classify, without the use of DNA-DNA hybridization, some isolates that are misidentified as *Afipia* based on 16S rRNA gene sequencing in the GenBank database. The 16S rRNA gene sequences given for *Afipia* genospecies 8 and 9 are in fact those of *Bosea* spp. (16).

The major drawback of *rpoB* sequencing is that the length of the gene (>4,000 bp) does not allow routine molecular identification or detection in clinical samples. For this purpose, we developed a simple tool that allowed determination of regions with high variability flanked by conserved areas. This tool allowed the design of universal primers for amplification and sequencing of a 740- to 752-bp fragment containing a hypervariable region of 408 to 420 bp for identification of all species tested in the phylum. Moreover, the percentages of homology observed in this partial sequence analysis correlate well with results of DNA-DNA hybridization (Table 5). With this partial sequence, a percentage of homology ≥98% ensures that two bacterial isolates belong to the same species whereas a percentage ≤96% indicates that they belong to two different species. *A. felis* and *A. felis* genospecies A, which are two genospecies on the basis of DNA-DNA hybridization results (45%), appear also as two genospecies by partial *rpoB* sequence comparison (94%). The development of partial *rpoB* sequencing allows the quick and accurate identification of bacteria in the genera *Afipia* and *Bosea* and detection of potential new species that will be used for surveys of hospital water system coloni-

TABLE 5. Comparison between DNA-DNA relatedness and percent homology in the *rpoB* hypervariable region for *Bosea* spp. and some *Afipia* spp.

Species	Strain	% DNA-DNA relatedness/% homology with strain									
		BI-42^T	63287^T	34649	34614^T	34635^T	63286	B-91-007352^T	76713	34632^T	34633^T
B. thiooxidans	BI-42^T	100/100									
B. massiliensis	63287^T	8/87	100/100								
	34649	14/87	71/98	100/100							
B. eneae	34614^T	16/88	14/90	15/90	100/100						
B. vestrisii	34635^T	13/89	12/90	12/91	31/96	100/100					
	63286	12/89	17/89	14/90	39/96	70/98	100/100				
A. felis	B-91-007352^T							100/100			
A. felis gsp A	76713							45/94	100/100		
A. birgiae	34632^T							6/85	8/82	100/100	
A. massiliensis	34633^T							7/83	9/83	40/91	100/100

zation and detection of human infection. Last, the procedure for designing PCR primers for amplification of hypervariable areas may be used in primer design for multilocus sequence typing (MLST). MLST is a typing method based on sequence comparisons of multiple loci (18). In this technique, partial sequences of housekeeping genes are determined and used to construct matrices that allow analysis of genetic relationships among isolates of a single species (18, 31). The number of alleles observed by using a given sequence is almost directly proportional to the number of polymorphic sites in the sequence (31). Actually, partial sequences are chosen randomly. Thus, in order to increase the number of alleles without increasing the sizes of determined sequences, it seems important to determine the most-variable regions in a given set of sequences. The SVARAP tool we propose herein could be useful for this purpose.

ACKNOWLEDGMENTS

We are indebted to J. S. Dumler for reviewing the manuscript and S. Ouattara for providing the *B. minatitlanensis* strain.

REFERENCES

1. **Ash, C., J. A. Farrow, M. Dorsch, E. Stackebrandt, and M. D. Collins.** 1991. Comparative analysis of *Bacillus anthracis*, *Bacillus cereus*, and related species on the basis of reverse transcriptase sequencing of 16S rRNA. Int. J. Syst. Bacteriol. **41:**343–346.
2. **Barker, J., and M. R. W. Brown.** 1994. Trojan horses of the microbial world: protozoa and the survival of bacterial pathogens in the environment. Microbiology **140:**1253–1259.
3. **Brenner, D. J., D. G. Hollis, C. W. Moss, C. K. English, G. S. Hall, J. Vincent, J. Radosevic, K. A. Birkness, W. F. Bibb, F. D. Quinn, B. Swaminathan, R. E. Weaver, M. W. Reeves, S. P. O'Connor, P. S. Hayes, F. C. Tenover, A. G. Steigerwalt, B. A. Perkins, M. L. Daneshvar, B. C. Hill, J. A. Washington, T. C. Woods, S. B. Hunter, T. D. Hadfield, G. W. Ajello, A. F. Kaufmann, D. J. Wear, and J. D. Wenger.** 1991. Proposal of *Afipia* gen. nov. with *Afipia felis* gen. nov. sp. nov. (formerly the cat scratch bacillus), *Afipia clevelandensis* sp. nov. (formerly the Cleveland clinic foundation strain), *Afipia broomeae* sp. nov., and three unnamed genospecies. J. Clin. Microbiol. **29:**2450–2460.
4. **Das, S. K., A. K. Mishra, B. J. Tindall, F. A. Rainey, and E. Stackebrandt.** 1996. Oxidation of thiosulfate by a new bacterium, *Bosea thiooxidans* (strain BI-42) gen. nov., sp. nov.: analysis of phylogeny based on chemotaxonomy and 16S ribosomal DNA sequencing. Int. J. Syst. Bacteriol. **46:**981–987.
5. **Drancourt, M., and D. Raoult.** 2002. *rpoB* gene sequence-based identification of *Staphylococcus* species. J. Clin. Microbiol. **40:**1333–1338.
6. **Felsenstein, J.** 1989. PHYLIP—phylogeny inference package (version 3.2). Cladistics **5:**164–166.
7. **Fox, G. E., J. D. Wisotzkey, and P. J. Jurtshuk.** 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. Syst. Bacteriol. **42:**166–170.
8. **Grimont, P. A. D.** 1988. Use of DNA reassociation in bacterial classification. Can. J. Microbiol. **34:**541–546.
9. **Grundmann, G., M. Neyra, and P. Normand.** 2000. High-resolution phylogenetic analysis of NO_2-oxidizing *Nitrobacter* species using the *rrs-rrl* IGS sequence and *rrl* genes. Int. J. Syst. Evol. Microbiol. **50:**1893–1898.
10. **Hall, G. S., K. Pratt-Rippin, and J. A. Washington.** 1991. Isolation of agent associated with cat scratch disease bacillus from pretibial biopsy. Diagn. Microbiol. Infect. Dis. **14:**511–513.
11. **Kim, B. J., S. H. Lee, M. A. Lyu, S. J. Kim, G. H. Bai, G. T. Chae, E. C. Kim, C. Y. Cha, and Y. H. Kook.** 1999. Identification of mycobacterial species by comparative sequence analysis of the RNA polymerase gene (*rpoB*). J. Clin. Microbiol. **37:**1714–1720.
12. **Ko, K. S., H. K. Lee, M. Y. Park, K. H. Lee, Y. J. Yun, S. Y. Woo, H. Miyamoto, and Y. H. Kook.** 2002. Application of RNA polymerase beta-subunit gene (*rpoB*) sequences for the molecular differentiation of *Legionella* species. J. Clin. Microbiol. **40:**2653–2658.
13. **La Scola, B., L. Barrassi, and D. Raoult.** 2000. Isolation of new α proteobacteria and *Afipia felis* from hospital water supplies by direct plating and amoebal co-culture procedures. FEMS Microbiol. Ecol. **34:**129–137.
14. **La Scola, B., L. Mezi, J. P. Auffray, Y. Berland, and D. Raoult.** 2002. Intensive care unit patients are exposed to amoeba associated pathogens. Infect. Control Hosp. Epidemiol. **23:**462–465.
15. **La Scola, B., M. N. Mallet, P. A. D. Grimont, and D. Raoult.** 2002. Description of *Afipia birgiae* sp. nov. and *Afipia massiliensis* sp. nov. and recognition of *Afipia felis* genospecies A. Int. J. Syst. Evol. Microbiol. **52:**1773–1782.
16. **La Scola, B., M. N. Mallet, P. A. D. Grimont, and D. Raoult.** 2003. Description of *Bosea eneae* sp. nov., *Bosea massiliensis* sp. nov. and *Bosea vestrisii* sp. nov., three novel species isolated from hospital water supplies and emendation of the genus *Bosea* (Das 1996). Int. J. Syst. Evol. Microbiol. **53:**15–20.
17. **La Scola, B., I. Boyadjiev, G. Greub, A. Khamis, C. Martin, and D. Raoult.** 2003. Amoebae-resisting bacteria and ventilator-associated pneumonia. Emerg. Infect. Dis. **9:**815–821.
18. **Maiden, J. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Rusel, 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.
19. **Marrie, T. J., D. Raoult, B. La Scola, R. J. Birtles, and E. de Carolis.** 2001. *Legionella*-like and other amoebal pathogens as agents of community-acquired pneumonia. Emerg. Infect. Dis. **7:**1026–1029.
20. **Mollet, C., M. Drancourt, and D. Raoult.** 1997. *rpoB* sequence analysis as a novel basis for bacterial identification. Mol. Microbiol. **26:**1005–1011.
21. **Ouattara, A. S., E. A. Assih, S. Thierry, J. L. Cayol, M. Labat, O. Monroy, and H. Macarie.** 2003. *Bosea minatitlanensis* sp. nov., a strictly aerobic bacterium isolated from an anaerobic digester. Int. J. Syst. Evol. Microbiol. **53:**1247–1251.
22. **Renesto, P., J. Gouvernet, M. Drancourt, V. Roux, and D. Raoult.** 2001. Use of *rpoB* gene analysis for detection and identification of *Bartonella* species. J. Clin. Microbiol. **39:**430–437.
23. **Rutala, W. A., and D. J. Weber.** 1997. Water as a reservoir of nosocomial pathogens. Infect. Control Hosp. Epidemiol. **18:**609–616.
24. **Siebert, P. D., A. Chenchik, D. E. Kellogg, K. A. Lukyanov, and S. A. Lukyanov.** 1995. An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Res. **23:**1087–1088.
25. **Stackebrandt, E., and B. M. Goebel.** 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present definition of bacteriology. Int. J. Syst. Bacteriol. **44:**846–849.
26. **Steenbergen, J. N., H. A. Shuman, and A. Casadevall.** 2001. *Cryptococcus neoformans* interactions with amoebae suggest an explanation for its virulence and intracellular pathogenic strategy in macrophages. Proc. Natl. Acad. Sci. USA **98:**15245–15250.
27. **Stout, J. E., and V. L. Yu.** 1997. Legionellosis. N. Engl. J. Med. **337:**682–687.

28. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
29. **Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins.** 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:**4876–4882.
30. **van Berkum, P., and J. J. Fuhrmann.** 2000. Evolutionary relationships among the soybean *Bradyrhizobia* reconstructed from 16S rRNA gene and internally transcribed spacer region sequence divergence. Int. J. Syst. Evol. Microbiol. **50:**2165–2172.
31. **Viscidi, R. P., and J. C. Demma.** 2003. Genetic diversity of *Neisseria gonorrhoeae* housekeeping genes. J. Clin. Microbiol. **41:**197–204.
32. **Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. L. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Trüper.** 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. **37:**463–464.
33. **Willems, A., and M. D. Collins.** 1992. Evidence of close relationship between *Afipia* (the causative organism of cat scratch disease) *Bradyrhizobium japonicum* and *Blastobacter denitrificans.* FEMS Microbiol. Lett. **96:**241–246.
34. **Xu, L. M., Z. Cui, J. Li, and H. Fan.** 1995. *Bradyrhizobium liaoningense* sp. nov. isolated from root nodule of soybeans. Int. J. Syst. Microbiol. **45:**706–711.