APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Sept. 1995, p. 3400–3406 Vol. 61, No. 9
0099-2240/95/$04.00+0

Comparison of Phenotypical and Molecular Methods for the Identification of Bacterial Strains Isolated from a Deep Subsurface Environment

V. BOIVIN-JAHNS,[1] A. BIANCHI,[2] R. RUIMY,[1] J. GARCIN,[2] S. DAUMAS,[3] AND R. CHRISTEN[1]*

Centre National de la Recherche Scientifique and Station Zoologique, Observatoire Océanologique, Université Paris 6, 06230 Villefranche sur mer,[1] Microbiologie Marine, Centre National de la Recherche Scientifique, UPR 223, Campus de Luminy, 13288 Marseille Cedex 9,[2] and Société GRAM SA, 13856 Aix en Provence Cedex 3,[3] France

Received 6 February 1995/Accepted 27 June 1995

Seventy-four bacterial strains were freshly isolated from a mine gallery. Using these bacteria, we have investigated how a molecular identification based on the analysis of small subunit rDNA sequences would compare in terms of precision and reliability to a more classical comparison of phenotypical descriptions (100 morphological and physiological traits). Our data clearly showed that a phylogenetic analysis of small subunit rDNA sequences is more efficient than classical phenotypic methods for the identification of bacterial strains freshly isolated from a natural environment, because occurrences of misidentification are very much decreased by this method. The lack of rDNA sequences for many described species is probably the major cause of a few failures in molecular identification, as the completeness of the database of small subunit rDNA sequences holds much importance in the degree of uncertainty in such identifications.

Developments in microbial ecology and industrial biotechnology are severely hampered by the lack of a reliable identification system (6, 11). It is notoriously difficult and often unsatisfactory to try to identify fresh bacterial isolates from natural environments such as soils, lakes, and oceans. To obtain a conventional phenotypic description requires long and fastidious work, and identifications of novel isolates by the use of dichotomous keys or computerized numerical taxonomy do not warrant satisfactory identifications of species already described in reference manuals (14, 15, 18, 23, 24, 26–28) or appropriate propositions for actual new bacterial species and genera.

On the other hand, phylogenetic relationships among various organisms, and thus their identification, can be derived from the degree of relatedness of their genomes: two closely related organisms share many homologies, while distant organisms display many differences. Methods that use this approach comprise measurements of DNA relatedness over the entire genome, comparisons of restriction patterns, especially ribotyping, and comparative analyses of sequences of homologous genes (25). DNA-DNA relatedness and ribotyping are best suited for the identification of closely related species or strains within a single species (3, 10, 13, 30, 31). Presently, a direct comparison of rRNA sequences is probably the most powerful tool for the identification of many bacteria (25). Indeed, rRNA genes (rDNA) are present and expressed in all bacterial species, are truly homologous in all organisms, are easily sequenced, and now offer a large and ever increasing database of sequences and allow the identification of uncultured bacteria (1).

In this study, we have undertaken to investigate how a molecular identification, based on comparisons of rDNA sequences, would compare with a more conventional phenotypic method for the identification of newly isolated bacteria. A collection of 74 bacterial strains was therefore isolated from a mine gallery. By its size and originality this set of strains is representative of bacterial collections usually studied by microbial ecologists. For each strain, identifications independently obtained by each method were compared. We then considered reliability and precision of identifications, usefulness of the data in terms of understanding bacterial ecology, and also ease of use in a daily laboratory practice for the study of a taxonomically diversified bacterial collection.

MATERIALS AND METHODS

Origin of the bacterial collection. A collection of bacterial strains was isolated from a mine gallery at a depth of 224 m in the Boom clay formation (Mol, Belgium). Clay samples were obtained by coring in the gallery wall horizontally along a length of 18 m (17). Subsamples were collected aseptically in the central part of the cores between 5 cm and 18 m. Samples of interstitial waters were also collected by three piezometers located 3, 7, and 15 m down the bore hole. Finally, samples were taken from the air and from boring tools to analyze possible contaminations. All instruments for boring or collecting water had been sterilized before use.

Bacterial strain isolation. Strains were isolated from nutrient agar plates used for viable counts (Bacto Peptone, 5 g; Bacto agar, 15 g; synthetic salt solution of the Mol formation, 1,000 ml). The synthetic salt solution included $MgCl_2 \cdot 6H_2O$, 22 mg; KCl, 25 mg; Na_2SO_4, 30 mg; NaCl, 40 mg; $NaHCO_3$, 1,170 mg; and H_2O, 1,000 ml. Bacterial colonies were isolated from each agar plate under aerobic conditions at 20°C (the temperature of the site of Mol), using a random isolation grid (4). Comparisons were conducted on 74 bacterial strains as referenced in Table 1. Because it is not possible to know beforehand how such bacterial strains are phylogenetically related, each strain or set of strains grouped together by one method will be referred to hereinafter as an OTU (operational taxonomic unit) as defined by Sneath and Sokal (24).

Phenotypical description of the strains. A set of 100 phenetic characters was used (16). These features included cellular morphology, gram staining, colonial morphology, and pigmentation; the production of catalase, oxidase, phosphatase, urease, gelatinase, Tween esterase, DNase, esculinase, amilase, and lecithinase; the ability to oxidize or ferment glucose; the ability to grow at 4, 37, and 44°C; and the ability to reduce nitrate and nitrite. Nutritional tests, the ability to grow at different NaCl concentrations, and growth factor requirements were analyzed with microtiter plate cultures. Organic compounds tested as the sole source of carbon and energy were D-alpha alanine, serine, glutamate, aspartate, L-lysine, L-arginine, L-proline, L-tryptophan, methionine, L-ornithine, asparagine, glycine, D-arabinose, D-ribose, D-glucose, lactose, saccharose, starch, gluconate, *N*-acetylglucosamine, glycerol, mannitol, sorbitol, butyrate, caproate, acetate, propionate, citrate, alpha keto-glutarate, malonate, succinate, fumarate, adipate, lactate, glycolate, mandelate, benzoate, and sarcosine. Organic compounds were added

* Corresponding author. Phone: (33) 93 76 37 80. Fax: (33) 93 76 37 92. Electronic mail address: christen@ccrv.obs-vlfr.fr.

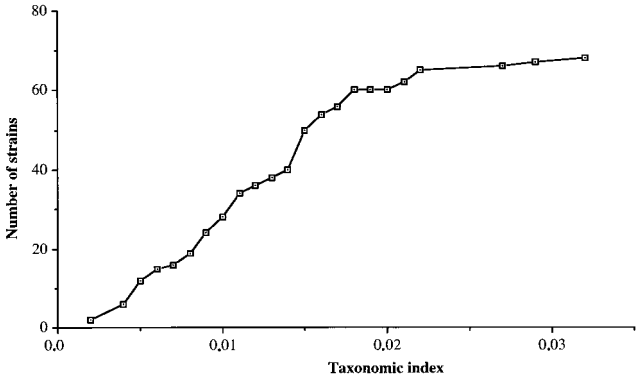

FIG. 1. Clustering curve obtained by plotting the cumulative numbers of strains clustered versus the corresponding taxonomic index values. This clustering curve can be used to determine a working phenon line setting apart clusters and individual strains considered separate OTUs. A cutting level can be chosen when the further clustering of a few strains would require a large increase of the overall taxonomic index. A change of the slope at around 0.015 indicates that above this value the aggregation of a few more strains would imply a drastic increase in intracluster heterogeneity. This value is a mean phenon line, but the most discriminative taxonomic index needs to be confirmed or adjusted for each cluster, as it depends upon the relative heterogeneity of each phenon.

to the synthetic saline solution (amended with NH_4Cl, 1 g liter^{-1}) at a final concentration of 0.1% (wt/vol), except for carbohydrates (which were added at a final concentration of 0.2% [wt/vol]), and supplemented after sterilization with a sterile solution of growth factors (16). Pure cultures were suspended into the mineral medium and 100 μl of this suspension was added with a multichannel pipette to each of the 96 wells of a microtitration plate containing 100 μl of medium. Each organic compound was tested in duplicate. For each strain, eight blank tests were performed in the unamended mineral medium. Bacterial growth was measured optically as optical density versus the average of the blank tests with a Titertek Multiskan apparatus after 24, 48, and 72 h of growth at 20°C.

Numerical comparison. In order to facilitate data handling, the phenotypes were coded in a binary form and were compared by the Adansonian principle of equally weighted characters (24) by means of a computerized program using the KHI2 coefficient and variance analysis (7). This numerical analysis led to the construction of a hierarchical tree in which strains were progressively clustered with decreasing overall similarity. The success of such numerical analysis depends mostly upon the choice of a cutting level for the hierarchical tree. A discriminating cutting level, assuming a compromise between optimal intracluster homogeneity and intercluster heterogeneity (19), can be determined from the clustering curve obtained by plotting the cumulative numbers of clustered strains against the corresponding taxonomic index values (Fig. 1). A cutting level can be chosen when a further clustering of a few strains would require a large increase of the overall taxonomic index (Fig. 1). Clusters appearing above this value are much less homogeneous than those formed below this value. It should be noted that the exact cutting value is finely tuned for each cluster, according to a detailed examination of the phenotypic descriptions. As an example, for cluster C12 (Fig. 2), the cutting level has been adjusted to 0.017 to include strain AC47. In fact, this strain differed from strains AG03, S203, S207, 1F15, and S208 by only a few nutritional characteristics, but it shared all the key characteristics used to define the genus *Micrococcus* (note that molecular analysis also included this strain in the same genus). On the other hand, a cluster at 0.015 would have grouped OTUs AC40, AG06, AG24, and AC43 (Fig. 2). However, such a cluster cannot be considered a single OTU, since it would include heterogeneity for two key characteristics (gram coloration and oxidase reaction). In this case our analysis led us to adjust the phenon line to a taxonomic index of 0.013 to distinguish four different OTUs belonging to four genera: *Acinetobacter*, *Micrococcus*, *Corynebacterium*, and *Moraxella*. Similar adjustments led to the clusterings shown in Fig. 2.

Conventional identification. By using their phenotypic descriptions, strains were tentatively identified to previously described taxonomic units according to the main reference manuals (14, 15, 18, 23, 24, 26–28). For two groups of higher rank, pseudomonads and the actinomycetes, identification at the genus level would have required the knowledge of specific characteristics not included in the standard description we used.

rDNA amplification and sequencing. Bacterial colonies were suspended in 200 μl of lysis mixture (10 mM Tris [pH 8.0], 1 mM EDTA, 1% Triton X-100) and boiled for 5 min. After a single chloroform extraction, 5 μl of supernatant was used to amplify the small subunit rRNA genes with two primers, corresponding respectively to positions 8 to 28 and 1491 to 1508 of the *Escherichia coli* small subunit rRNA sequence. The PCR conditions and sequencing of the PCR products were as described previously (21).

Phylogenetic analysis and alignment: general procedure. Partial rDNA sequences that extended from position 8 to position 703 of the *Escherichia coli* small subunit rDNA sequence were first obtained. This domain was chosen because it comprises altogether divergent parts that would clearly discriminate two bacteria that do not belong to the same genus and it includes conserved domains that allow phylogenetic analyses of more distant relationships. All sequence alignments and species selections were done with computer programs developed by us and available by anonymous ftp (ccrv.obs-vlfr.fr directory pub/christen/16S). A neighbor-joining algorithm like that developed by Saitou and Nei (22) was used. The program was rewritten to include inputs and outputs compatible with the ribosomal database and with other programs developed in our laboratory (running on 386 [and higher]-compatible personal computers). Finally, all trees were plotted with a Macintosh computer and a program (njplot) developed by M. Gouy (Unité Recherche Associée 243, Centre National de la Recherche Scientifique, Université Claude Bernard, Villeurbanne, France) that

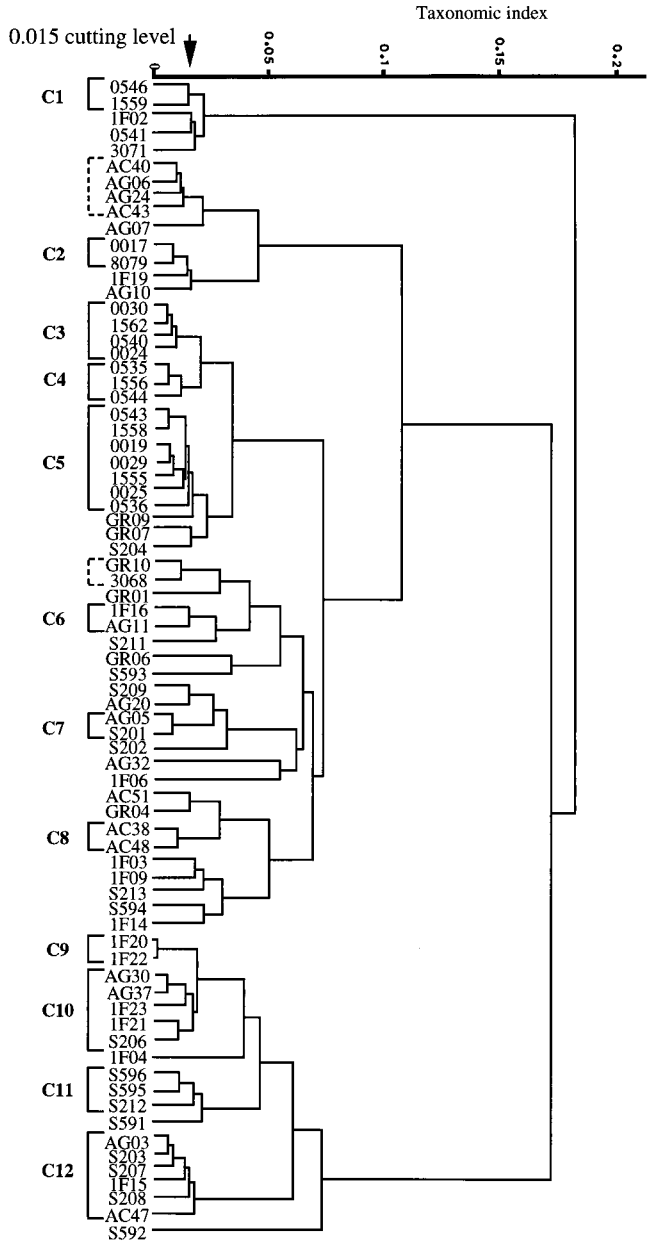

FIG. 2. Dendrogram showing a hierarchical comparison of the phenotypes for the 74 studied strains. Clusters retained around the cutting level of 0.015 are indicated. Two illegitimate groupings discussed in the text are indicated by dashed brackets. Tentative identifications of individual strains and clusters formed around the 0.015 taxonomical index were then made as described in the text.

TABLE 1. Comparison of phenotypic and molecular identifications for the 74 bacterial strains used in this study[a]

Strain	PI 1 (genus or group)	Change[b]	PI 2 (genus or group)	MI 1 (genus)	MI 2 (genus)	Mismatch[c]	Comparison[d]
AC38	*Aerococcus*	Morphology	*Arthrobacter*	*Arthrobacter*			+
AC40	*Acinetobacter*			*Acinetobacter*			+
AC43	*Micrococcus*	Morphology	*Arthrobacter*	*Corynebacterium*	*Rhodococcus*	Morphology	−
AC47	*Staphylococcus*	Glucose fermentation	*Micrococcus*	*Micrococcus*			+
AC48	*Aerococcus*	Morphology	*Arthrobacter*	*Arthrobacter*			+
AC51	*Azomonas*	GRAM, Man	*Arthrobacter*	*Arthrobacter*			+
AG03	*Micrococcus*			*Micrococcus*			+
AG05	*Bacillus*			*Bacillus*			+
AG06	*Moraxella*-like		*Moraxella*-like	*Acinetobacter*		Oxidase	−
AG07	Actinomycete		Actinomycete	?	*		{(+)}
AG10	Actinomycete		Actinomycete	*Rhodococcus*			{+}
AG11	Pseudomonad			*Pseudomonas*			{+}
AG20	*Bacillus*			*Bacillus*			+
AG24	*Moraxella*			*Gamma Proteobacteria*	*Moraxella*		+
AG30	*Staphylococcus*			*Staphylococcus*			+
AG32	*Corynebacterium*			*Corynebacterium*			+
AG37	*Staphylococcus*			*Staphylococcus*			+
GR01	Pseudomonad			*Pseudomonas*			{+}
GR04	*Azomonas*	Oxidase	Pseudomonad	*Pseudomonas*			{+}
GR06	Pseudomonad			*Pseudomonas*			{+}
GR07	*Alcaligenes*	Glucose, Oxidase	Pseudomonad	*Pseudomonas*			{+}
GR09	Pseudomonad		Pseudomonad	*Pseudomonas*			{+}
GR10	Pseudomonad			*Pseudomonas*			{+}
OO17	*Paracoccus*	Morphology	*Moraxella*-like	*Alcaligenes*	*	Motility	−
OO19	*Alcaligenes*	Motility	*Moraxella*-like	*Alcaligenes*		Motility	−
OO24	*Moraxella*-like		*Moraxella*-like	*Alcaligenes*		Motility	−
OO25	*Acinetobacter*	Oxidase	*Moraxella*-like	*Alcaligenes*		Motility	−
OO29	*Moraxella*-like		*Moraxella*-like	*Alcaligenes*		Motility	−
OO30	*Moraxella*-like		*Moraxella*-like	*Alcaligenes*		Motility	−
O535	*Alcaligenes*	Motility	*Moraxella*-like	[*Streptomyces*]	*	Morphology	(−)
O536	*Alcaligenes*	Motility	*Moraxella*-like	*Alcaligenes*		Motility	−
O540	*Alcaligenes*	Motility	*Moraxella*-like	*Alcaligenes*		Motility	−
O541	Actinomycete		Actinomycete	*Staphylococcus*		Morphology	{−}
O543	*Alcaligenes*	Motility	*Moraxella*-like	*Alcaligenes*		Motility	−
O544	*Alcaligenes*	Motility	*Moraxella*-like	*Alcaligenes*		Motility	−
O546	Actinomycete			[*Streptomyces*]	*		{(+)}
S201	*Bacillus*			*Bacillus*			+
S202	*Bacillus*			*Bacillus*			+
S203	*Staphylococcus*	Glucose fermentation	*Micrococcus*	*Micrococcus*			+
S204	*Staphylococcus*		*Staphylococcus*	*Alcaligenes*	*	Morphology	−
S206	*Staphylococcus*			*Staphylococcus*			+
S207	*Staphylococcus*	Glucose fermentation	*Micrococcus*	*Micrococcus*			+
S208	*Staphylococcus*	Glucose fermentation	*Micrococcus*	*Micrococcus*			+
S209	*Corynebacterium*		*Corynebacterium*	[*Arthrobacter*/*Rothia*]	*	Morphology	(−)
S211	*Moraxella*-like		*Moraxella*-like	*Gamma Proteobacteria*	*; *Moraxella*		+
S212	*Staphylococcus*			*Staphylococcus*			+
S213	*Arthrobacter*		*Arthrobacter*	*Corynebacterium*		Glucose fermentation	−
S591	*Staphylococcus*			*Staphylococcus*			+
S592	*Staphylococcus*			*Staphylococcus*			+
S593	*Staphylococcus*			*Staphylococcus*			+
S594	*Staphylococcus*			*Staphylococcus*			+
S595	*Staphylococcus*			*Staphylococcus*			+
S596	*Staphylococcus*			*Staphylococcus*			+
1555	*Alcaligenes*	Motility	*Moraxella*-like	*Alcaligenes*		Motility	−
1556	*Alcaligenes*	Motility	*Moraxella*-like	*Alcaligenes*		Motility	−
1558	*Alcaligenes*	Motility	*Moraxella*-like	*Alcaligenes*		Motility	−
1559	Actinomycete			*Nocardioides*			{+}
1562	*Alcaligenes*	Motility	*Moraxella*-like	*Alcaligenes*		Motility	−
3068	*Alcaligenes*	Motility	*Moraxella*-like	*Alcaligenes*		Motility	−
3071	Actinomycete			*Nocardioides*			{+}
8079	*Paracoccus*	Morphology	*Moraxella*-like	*Gamma Proteobacteria*	*; *Moraxella*		+
1F02	Actinomycete		*Actinomycetes*	?	*		{(+)}
1F03	*Staphylococcus*			*Staphylococcus*			+
1F04	*Staphylococcus*			[*Rhizobium*/*Agrobacterium*]	*		(−)
1F06	*Arthrobacter*			[*Clavibacter*]	*		(+)
1F09	*Arthrobacter*		*Arthrobacter*	*Staphylococcus*		Morphology	−
1F14	*Acinetobacter*	Oxidase	*Moraxella*-like	*Pseudomonas*		Motility	−

Continued on following page

TABLE 1—*Continued*

Strain	PI 1 (genus or group)	Change[b]	PI 2 (genus or group)	MI 1 (genus)	MI 2 (genus)	Mismatch[c]	Comparison[d]
1F15	*Staphylococcus*	Glucose fermentation	*Micrococcus*	*Micrococcus*			+
1F16	Pseudomonad			*Pseudomonas*			{+}
1F19	*Acinetobacter*			*Acinetobacter*			+
1F20	*Staphylococcus*			*Staphylococcus*			+
1F21	*Staphylococcus*			*Staphylococcus*			+
1F22	*Staphylococcus*			*Staphylococcus*			+
1F23	*Staphylococcus*			*Staphylococcus*			+

[a] Results of the initial phenotypic identifications (PI) and molecular identifications (MI) appear in columns PI 1 and MI 1, respectively. The two identification processes were repeated for a number of bacteria, and the results have been reported in columns PI 2 and MI 2. For the second molecular identification, strains that were sequenced again are indicated by an asterisk. When a genus could not be reliably determined from rRNA sequences, either the most closely related genus is indicated in brackets or a question mark shows that the bacterium had a deep lineage that did not cluster with any other bacteria for which an rRNA sequence was available. Four OTUs were reassigned but only because more rDNA data had become available at that time (identical sequences were obtained in all cases).

[b] The key parameter that changed between the first and the second phenotypic descriptions.

[c] The phenotypical characteristic not compatible with the molecular determination.

[d] +, genus agreement; −, genus disagreement; (+), genus not determined from rDNA sequences but agreement above the genus level; (−), genus not determined from rDNA sequences but disagreement above the genus level; {+}, genus not determined by the phenotype but agreement above the genus level; {−}, genus not determined by the phenotype but disagreement above the genus level.

allows transformation of a formal tree representation (in Newick's format) into MacDraw drawings.

All sequences were entered in our rRNA database containing about 3,000 aligned sequences belonging to the presently known bacterial phylums; the sequences were aligned within domains of low divergence, which are easy to align among all bacteria. The diversity of the isolated bacteria was first determined by calculating evolutionary distances with only those sequence regions for which all strains could be unambiguously aligned (conserved regions), and a phylogenetic tree was constructed from dissimilarity matrices by a neighbor-joining method. This first phylogenetic analysis that included representatives of major known phyla of bacteria allowed the attribution of each newly isolated bacterium to one of these phyla. Identification at the genus level was achieved by further phylogenetic analyses within more divergent domains after progressive rounds of alignment, phylogenetic analysis, and taxonomic assignment. Such procedures resulted in an identification at some level of taxonomy. If an OTU clustered within a monophyletic unit of species that had the same genus, this genus was attributed to the OTU. The same attribution was made when an OTU formed a robust monophyletic cluster with a previously referenced bacterial genus for which only a single representative species had been sequenced. In some cases, either because an OTU branched deeply or because it did not cluster strongly with any other referenced bacteria, a genus identification could not be made (see Discussion). In such a case, each OTU was nevertheless attributed to a taxonomical unit of higher rank (in terms of rDNA taxonomy) that also comprised a number of well-known genera for which small subunit rDNA sequences were available.

Nucleotide sequence accession numbers. rRNA sequences have been submitted to the EMBL data bank under accession numbers X86572 to X86653.

RESULTS

Phenotypic identification. The Adansonian analysis was not used for a direct taxonomical purpose but only to facilitate data handling and strain gathering; therefore, conventional and numerical methods led to the description of the same set of genera. The use of conventional identification keys led to the identification of nine taxonomical groups, including seven genera: *Acinetobacter*, *Moraxella*-like, *Arthrobacter*, *Corynebacterium*, *Bacillus*, *Staphylococcus*, *Micrococcus*, pseudomonads, and the actinomycetes (see Materials and Methods). The clustering curve (Fig. 1) indicates a decrease of cluster homogeneity above the 0.015 taxonomic index value. At this level we distinguished 12 clusters and 34 individual strains (Fig. 2) we can suppose to be 46 different OTUs. These 46 OTUs were identified as the gram-negative rods (31 strains corresponding to 18 OTUs) consisting of the *Moraxella*-like genus (21 strains, 9 OTUs), pseudomonads (8 strains, 7 OTUs), and the *Acinetobacter* genus (2 strains, 2 OTUs); the gram-positive rods (13 strains distributed into 11 OTUs) consisting of the *Arthrobacter* (7 strains, 6 OTUs), *Corynebacterium* (2 strains, 2 OTUs), and *Bacillus* (4 strains, 3 OTUs) genera; the gram-positive cocci (23 strains corresponding to 11 OTUs) consisting of the *Staphylococcus* genus (17 strains, 10 OTUs) and the *Micrococcus* genus (6 strains, 1 OTU); and the actinomycetes group (7 strains, 6 OTUs).

Clusters that formed around the 0.015 level appeared phenotypically homogeneous in relation to the conventional identification. Illegitimate gatherings appeared exceptional: one cluster, for example, contained four strains, three of which were gram negative and one of which was gram positive, and another cluster contained two strains described as gram-negative rods, one being motile and oxidase positive, the other one being motile and oxidase negative (Fig. 2). These misassociations could be due to an overweighting of nutritional characteristics in the phenotypic description. Because phenetic and molecular identifications did not correspond for many OTUs, a second phenetic identification was undertaken for a number of strains. The identifications resulting from these analyses of phenetic data are summarized in Table 1.

Molecular identification. A phylogenetic analysis (Fig. 3) that included all partial sequences showed (i) the existence of major clusters and (ii) very close relationships among some OTUs. A closer scrutiny of OTUs that clustered very tightly often showed that they had identical sequences even in the more divergent domains of partial sequences. Remaining OTUs were either the single representative of a unique taxon or they clustered in monophyletic taxa of seemingly related OTUs. The deeper branches of Fig. 3 represented three of the major phyla of bacteria: *Firmicutes*, high G+C; *Firmicutes*, low G+C; and *Proteobacteria*. OTUs belonging to the phylum *Proteobacteria* fell into three (α, β, and γ) of the five major groups (α, β, γ, δ, and ε) presently recognized within this phylum. Almost the entire length of the small subunit rDNA was sequenced for 23 of these OTUs. This was undertaken (i) for a selection of OTUs that represented some of the major clusters determined in the preceding analysis and (ii) for some of the seemingly strongly related OTUs. OTUs for which nearly a full-length sequence was obtained are indicated in Fig. 3. Full-length sequences confirmed that OTUs that had partial identical sequences also had identical full-length sequences.

More detailed phylogenetic analyses were then undertaken as shown in Fig. 4. The results of similar analyses undertaken for every OTU are summarized in Table 1. No significant difference in tree resolution was observed when partial se-

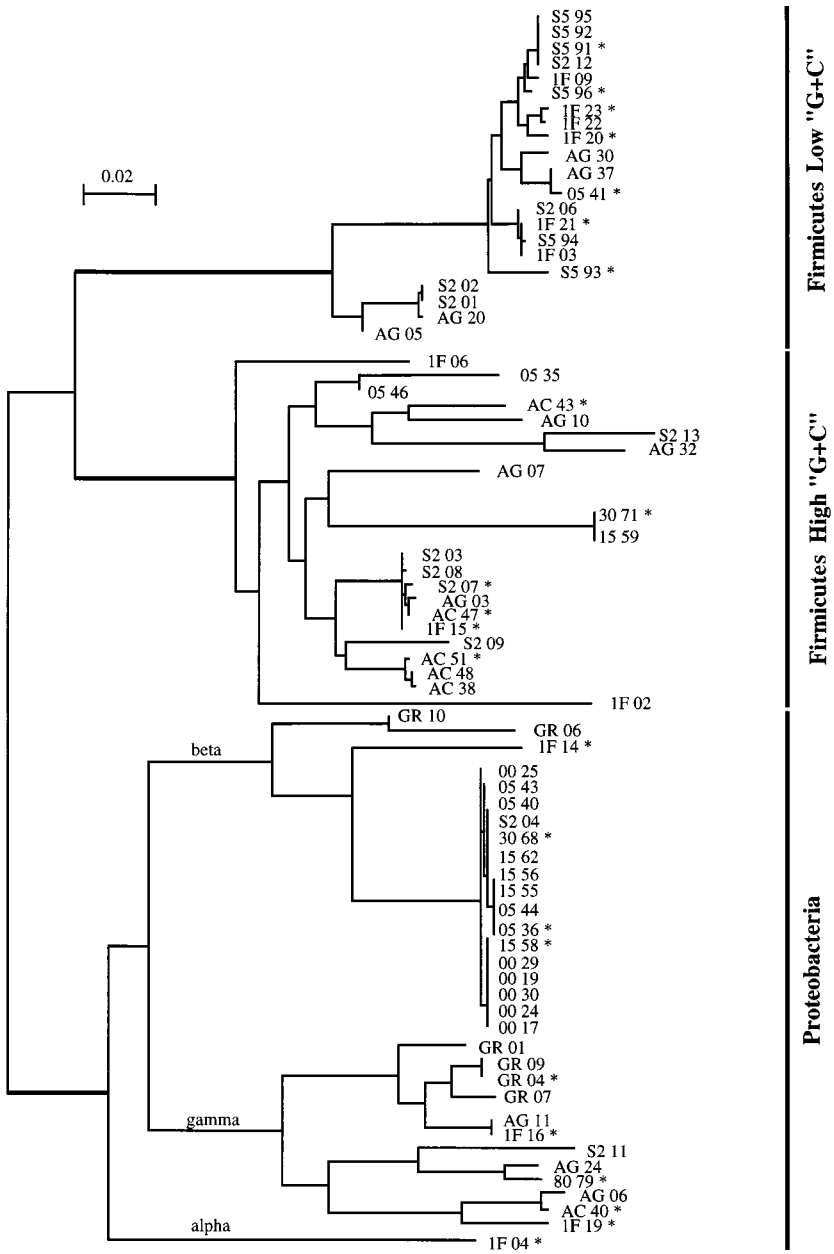

FIG. 3. Molecular phylogeny of the 74 OTUs studied, according to a neighbor-joining analysis of partial small subunit rDNA sequences (unrooted tree). OTUs for which full-length sequences have been obtained are marked by asterisks. Major taxonomic divisions are indicated. See the text for further details.

quences instead of full-length sequences were used for a determination of the genus.

Since molecular and phenotypic identifications conflicted for a large number of OTUs, a second molecular identification was undertaken for several of these OTUs. Bacteria that were submitted to a second analysis were freshly taken from the cultures in the laboratory doing the phenotypic identification, therefore ruling out the possibility that sequences and phenotypes could have been determined for different strains. All steps of the molecular procedure were undertaken a second time: DNA was extracted, amplified, and sequenced (partial sequences). This task was accomplished for 11 OTUs and always provided identity between the first and the second sequences (Table 1).

DISCUSSION

Following the completion of a first procedure of identification, and when the two methods had been conclusive for the delineation of a genus, 42 discrepancies resulted from the comparisons of the two datum sets. For 26 strains a second phenetic analysis led to the determination of a genus that often differed from that deduced from the first analysis. By contrast, the second analysis using the molecular method was consistent, except in four cases in which a more precise identification could be obtained, because additional sequences had become available in the rDNA database. Following the completion of this study, results from both the phenotypical and the molecular analyses were in agreement for the genus determination of

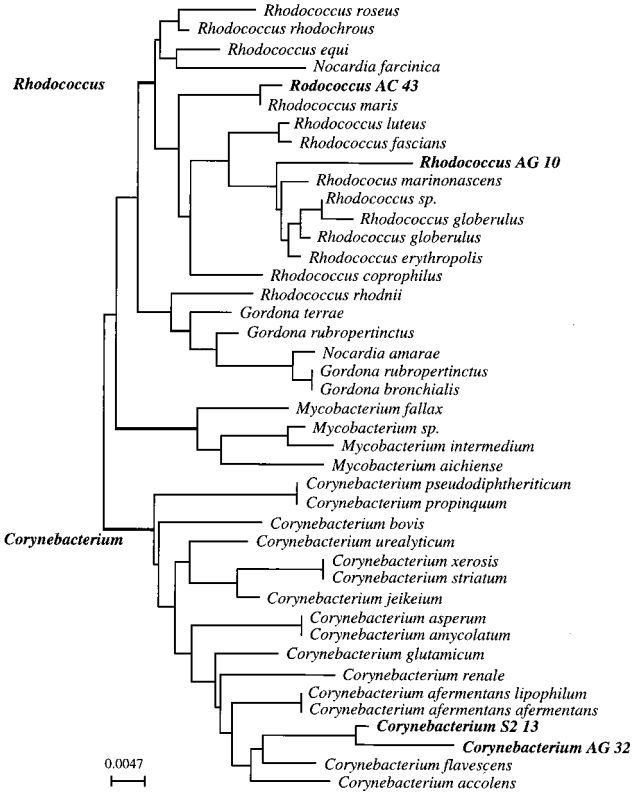

FIG. 4. Example of a detailed phylogenetic analysis that allowed a determination at the genus level. Four OTUs (AC43, AG10, S213, AG32) are clustered within the CMN group of the phylum *Firmicutes*, high G+C. The two OTUs AC43 and AG10 were included within a monophyletic group composed entirely of *Rhodococcus* species. They were thus identified as members of the *Rhodococcus* genus. For similar reasons, OTUs S213 and AG32 were identified as members of the genus *Corynebacterium*. Other similar detailed analyses led to determinations as indicated in Table 1.

34 OTUs (Table 1). For 15 further strains, identifications were compatible (the genus identified by one method was included in the taxonomic unit of higher rank obtained by the other method). This means that between 45 and 66% of these newly isolated strains could probably be correctly identified with a standard phenotypic description and the subsequent use of conventional identification keys. For 22 strains, phenetic and molecular identifications differed. For these strains, even after accurate verifications of phenetic descriptions and rDNA sequences, two different identifications were confirmed and remained incompatible. Incompatibilities in identifications were mostly due to strain motility, but other characteristics were also contradictory (Table 1).

In conventional methods, misidentifications are frequently due to the well-known uncertainties of phenotypical descriptions. Some of the features used to describe bacterial strains in a numerical analysis are characteristics that may be influenced by cultivation conditions. Many differences have been observed in the characteristics of fresh isolates compared with those of strains stored in culture collections, perhaps because fresh isolates from natural environments are unadapted to nutrient-rich culture conditions. Plasmid DNA might be responsible for encoding some of the traits used for diagnosis (2); as plasmid DNA may be lost in storage, resulting differences in characteristics may explain why environmental isolates often do not cluster with named reference strains according to numerical taxonomy (2, 7, 12). Long-term regulations of some adaptive genes are also possible by means of the movements of transposons, for example (5, 29). A striking example can also be found among *Photobacterium* spp. in which flagella are present in one strain but absent in another strain of the same species (9). When a dichotomous identification is attempted, the misinterpretation of only one of these major phenotypical tests leads to an erroneous identification at the genus level. A numerical comparison and clustering of strains based on a probabilistic approach decreases the influence of a few (or even unique) aberrant results (19), but as shown in this study, major problems are still encountered. On the other hand, since the rDNA genes are always present and since their sequences are probably very little affected by changes in the ecological niche in which bacteria have evolved or adapted, none of the previous problems is to be taken into account. Seven strains could not be identified to any previously referenced genus, a failure that can be explained either because that particular strain belongs to an already-referenced genus for which small subunit ribosomal sequences are still unavailable or because the organism under consideration is indeed a new genus.

Our data confirm that for the identification of bacterial strains isolated from a natural environment, a molecular procedure based upon comparisons of small subunit rDNA sequences clearly appears more efficient than the use of the classical phenotypic methods, because the occurrences of misidentification were very much decreased. A molecular identification, still a sophisticated methodology for many laboratories specialized in ecology, requires sequencing to be done, but this procedure is strictly identical for all groups of bacteria. With the advent of automation, this could soon be easier and less expensive than the diverse conventional techniques (photonic and electronic microscopy, culture media, chemical analyses of cell wall composition, G+C content [moles percent], analysis of metabolic pathways, etc.) necessary to obtain a reliable phenotypic identification.

Another advantage of molecular methods lies in the process of identification itself. Whether phenetic or molecular, identification deals with the process of allocating every organism to a current and previously defined taxon by matching relevant characteristics to those of an appropriate database containing the same information for known taxa (11). When the organism under consideration is sufficiently different from any referenced organism, a new taxon can be created according to well-established criteria. By the conventional approach, identification to an already-described genus is usually the rule, as the phenotypic characteristics of many genera are often vague enough to accommodate different bacteria. The use of molecular methods has indeed demonstrated that some conventionally described genera are in fact aggregates of bacteria that should be reassigned or divided in different genera (8, 9, 20, 21). Molecular identification also proceeds by matching a sequence to a database. When the sequence is identical to that of a reference strain already sequenced, determination at the genus level is easy. When there is no close match between the new sequence and any sequence of the database, the advantage is that it is possible to use the methods of molecular phylogeny to determine precisely to which bacteria a new organism can be closely related. When the organism under consideration clusters within a group of bacteria that do belong to the same genus, it is usually assumed that an identification at the genus level has been obtained. Data analysis is more problematic when a monophyletic cluster is obtained that groups bacteria belonging to different genera and when more precise in-group phylogenetic relationships are difficult to determine. It is now often assumed that taxonomy for this group of bacteria is poor. Finally, molecular phylogeny can attribute very deep branch-

ing to the organism under consideration and no strong relationship with any already-referenced species. Whenever genus attribution is unclear or difficult, additional data (such as chemotaxonomy) are necessary to complement the rRNA sequence in order to solve the problem. An important advantage of molecular methods is that the lack of a determination at the genus level does not preclude attribution at a higher taxonomic level, such as the determination of a phylum, class, and family. Finally, there remains some doubt as to whether rDNA sequence identity is sufficient to assess whether two bacteria belong to the same species. rDNA sequences probably do not always contain enough information to ascertain taxonomy at the species level, and additional studies of phenotypes and DNA-DNA hybridizations are therefore necessary.

In conclusion, the lack of rDNA sequences for many described species is probably the major cause of failure in a molecular identification. The present database includes comparatively few sequences of organisms isolated from the environment and for which an extended phenotypic description has been accomplished. For these reasons it is obvious that phenetic and rDNA sequences cannot be considered exclusive elements of strategies to characterize novel isolates of natural environments. Indeed, if we had sequenced only the small subunit rDNA molecules for the 74 strains, we would have obtained good identifications at the genus level but not very good indications of their metabolic capacities, as shown by the departure of many strains from the repertoire classically described for each of these genera.

ACKNOWLEDGMENTS

This work was supported by the CNRS and by grants of the European Community as part of the project Archimede-Argiles and of the Direction de la Recherche et des Etudes Doctorales. V.B. was a recipient of a CIFRE fellowship.

REFERENCES

1. **Amann, R., W. Ludwig, and K.-H. Schleifer.** 1994. Identification of uncultured bacteria: a challenging task for molecular taxonomists. ASM News **60:**360–365.
2. **Austin, B.** 1988. Marine microbiology, p. 222. Cambridge University Press, Cambridge.
3. **Aznar, R., W. Ludwig, and K.-H. Schleifer.** 1993. Ribotyping and randomly amplified polymorphic DNA analysis of *Vibrio vulnificus* biotypes. Syst. Appl. Microbiol. **16:**303–309.
4. **Bianchi, M., and A. Bianchi.** 1982. Statistical sampling of bacterial strains and its use in bacterial diversity measurements. Microb. Ecol. **8:**61–69.
5. **Blot, M., B. Hauer, and G. Monnet.** 1994. The Tn5 bleomycin resistance gene confers improved survival and growth advantage on *Escherichia coli*. Mol. Gen. Genet. **242:**595–601.
6. **Bull, A. T., M. Goodfellow, and J. H. Slater.** 1993. Biodiversity as a source of innovation in biotechnology. Annu. Rev. Microbiol. **46:**219–257.
7. **Delabre, M., A. Bianchi, and M. Veron.** 1973. Étude critique des méthodes de taxonomie numérique. Application à une classification de bactéries aquacoles. Ann. Inst. Pasteur (Paris) **124A:**489–506.
8. **Gauthier, G., M. Gauthier, and R. Christen.** Phylogenetic analysis of the genera *Alteromonas Shewanella*, and *Moritella* using genes coding for small-subunit rDNA sequences and division of the genus *Alteromonas* into two genera, *Alteromonas* (emended) and *Pseudoalteromonas*, gen. nov., and proposal of twelve new species combinations. Int. J. Syst. Bacteriol., in press.
9. **Gauthier, G., B. Lafay, R. Ruimy, V. Breittmayer, J. L. Nicolas, M. Gauthier, and R. Christen.** 1995. Small-subunit rRNA sequences and whole DNA relatedness concur for the reassignment of *Pasteurella piscicida* (Snieszko et al.) Janssen and Surgalla to the genus *Photobacterium* as *Photobacterium damsela* subsp. *piscicida* comb. nov. Int. J. Syst. Bacteriol. **45:**139–144.
10. **Gomez-Lus, P., B. S. Fields, R. F. Benson, W. T. Martin, S. P. O'Connor, and C. M. Black.** 1993. Comparison of arbitrarily primed polymerase chain reaction, ribotyping, and monoclonal antibody analysis for subtyping *Legionella pneumophila* serogroup 1. J. Clin. Microbiol. **31:**1940–1942.
11. **Goodfellow, M., and A. G. O'Donnell.** 1993. Roots of bacterial systematics, p. 3–54. *In* M. Goodfellow and A. G. O'Donnell (ed.), Handbook of new bacterial systematics. Academic Press, London.
12. **Grimes, D. J., D. Jacobs, D. G. Swartz, P. R. Brayton, and R. R. Colwell.** 1993. Numerical taxonomy of gram-negative, oxidase-positive rods from carcharhinid sharks. Int. J. Syst. Bacteriol. **43:**88–98.
13. **Grimont, F., and P. A. D. Grimont.** 1986. Ribosomal ribonucleic acid gene restriction patterns as potential taxonomic tools. Ann. Inst. Pasteur Microbiol. **137B:**165–175.
14. **Holt, J. G., N. R. Krieg, P. H. A. Sneath, J. T. Staley, and S. T. Williams (ed.).** 1994. Bergey's manual of determinative bacteriology, p. 787. Williams & Wilkins, Baltimore.
15. **Krieg, N. R., and J. G. Holt (ed.).** 1984. Bergey's manual of systematic bacteriology, vol. 1, p. 964. Williams & Wilkins, Baltimore.
16. **Martin, Y. P., and M. A. Bianchi.** 1980. Structure, diversity and catabolic potentialities of aerobic heterotrophic bacterial populations associated with continuous cultures of natural marine phytoplankton. Microb. Ecol. **5:**265–279.
17. **Merceron, T., M. N. Hermin, A. Bianchi, V. Boivin, and R. Christen.** Microbial investigations within the ARCHIMEDE-CLAY project concerning the understanding of the water chemistry in clay environment. Submitted for publication.
18. **O'Brien, M., and R. R. Colwell.** 1987. Characterization tests for numerical taxonomy studies, p. 69–104. *In* R. Grigorova and R. R. Colwel (ed.), Methods in microbiology. Academic Press, London.
19. **Priest, F. G., and S. T. Williams.** 1993. Computer-assisted classification, p. 361–381. *In* M. Goodfellow and A. G. O'Donnell (ed.), Handbook of new bacterial systematics. Academic Press, London.
20. **Ruimy, R., P. Boiron, V. Boivin, and R. Christen.** 1994. A phylogeny of the genus *Nocardia* deduced from the analysis of small-subunit ribosomal DNA sequences, including transfer of *Nocardia amarae* to the genus *Gordona* as *Gordona amarae* comb. nov. FEMS Microbiol. Lett. **123:**261–268.
21. **Ruimy, R., V. Breittmayer, P. Elbaze, B. Lafay, O. Boussemart, M. Gauthier, and R. Christen.** 1994. Phylogenetic analysis and assessment of the genera *Vibrio*, *Photobacterium*, *Aeromonas*, and *Plesiomonas* deduced from small-subunit rRNA sequences. Int. J. Syst. Bacteriol. **44:**416–426.
22. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
23. **Sneath, P. H. A., N. S. Mair, M. E. Sharpe, and H. G. Holt (ed.).** 1986. Bergey's manual of systematic bacteriology, vol. 2, p. 1599. Williams & Wilkins, Baltimore.
24. **Sneath, P. H. A., and R. R. Sokal.** 1973. Numerical taxonomy, p. 573. W. H. Freeman & Co., San Francisco.
25. **Stackebrandt, E., and M. Goodfellow.** 1991. Nucleic acid techniques in bacterial systematics, p. 329. John Wiley & Sons Ltd., Chichester, England.
26. **Staley, J. T., M. P. Bryant, N. Pfennig, and J. G. Holt (ed.).** 1989. Bergey's manual of systematic bacteriology, vol. 3, p. 1697. Williams & Wilkins, Baltimore.
27. **Starr, M. P., H. Stolp, H. G. Trüper, A. Balows, and H. G. Schlegel.** 1981. The prokaryotes. A handbook on habitats, isolation, and identification of bacteria, vol. I, p. 1102. Springer-Verlag, Berlin.
28. **Starr, M. P., H. Stolp, H. G. Trüper, A. Balows, and H. G. Schlegel.** 1981. The prokaryotes. A handbook on habitats, isolation, and identification of bacteria, vol. II, p. 1103–2284. Springer-Verlag, Berlin.
29. **Stevens, A. M., N. B. Shoemaker, L.-Y. Li, and A. A. Salyers.** 1993. Tetracycline regulation of genes on *Bacteroides* conjugative transposons. J. Bacteriol. **175:**6134–6141.
30. **Tarkka, E., H. Ahman, and A. Siitonen.** 1994. Ribotyping as an epidemiologic tool for *Escherichia coli*. Epidemiol. Infect. **112:**263–274.
31. **Van Steenbergen, T. J., C. Menard, C. J. Tijhof, C. Mouton, and J. De Graaff.** 1993. Comparison of three molecular typing methods in studies of transmission of *Porphyromonas gingivalis*. J. Med. Microbiol. **39:**416–421.