Base-specific fragmentation of amplified 16S rRNA genes analyzed by mass spectrometry: A tool for rapid bacterial identification

Friedrich von Wintzingerode*, Sebastian Böcker†, Cord Schlötelburg*, Norman H. L. Chiu†, Niels Storm‡, Christian Jurinke†, Charles R. Cantor†, Ulf B. Göbel*, and Dirk van den Boom†§

*Institut für Mikrobiologie und Hygiene, Universitätsklinikum Charité, Humboldt-Universität zu Berlin, Dorotheenstrasse 96, 10117 Berlin, Germany; ‡SEQUENOM, Mendelssohnstrasse 15D, 22761 Hamburg, Germany; and †SEQUENOM, 3595 John Hopkins Court, San Diego, CA 92121

Contributed by Charles R. Cantor, March 21, 2002

A rapid approach to the 16S rRNA gene (16S rDNA)-based bacterial identification has been developed that combines uracil-DNA-glycosylase (UDG)-mediated base-specific fragmentation of PCR products with matrix-assisted laser desorption ionization-time-of-flight mass spectrometry (MALDI-TOF MS). 16S rDNA signature sequences were PCR-amplified from both cultured and as-yet-uncultured bacteria in the presence of dUTP instead of dTTP. These PCR products then were immobilized onto a streptavidin-coated solid support to selectively generate either sense or antisense templates. Single-stranded amplicons were subsequently treated with uracil-DNA-glycosylase to generate T-specific abasic sites and fragmented by alkaline treatment. The resulting fragment patterns were analyzed by MALDI-TOF MS. Mass signals of 16S rDNA fragments were compared with patterns calculated from published 16S rDNA sequences. MS of base-specific fragments of amplified 16S rDNA allows reliable discrimination of sequences differing by only one nucleotide. This approach is fast and has the potential for high-throughput identification as required in clinical, pharmaceutical, or environmental microbiology. In contrast to identification by MS of intact whole bacterial cells, this technique allows for the characterization of both cultured and as-yet-uncultured bacteria.

Emerging antibiotic-resistant pathogens, the actual threat of bioterrorist attacks, and bioremediation or bioprospecting efforts all require fast and accurate identification of the bacteria involved. Conventional diagnoses rely on the characterization of phenotypic traits of pure cultures obtained from specimens after cultivation and isolation of bacteria on appropriate laboratory media. These culture-based methods are time-consuming especially for slow-growing pathogens such as mycobacteria and generally are restricted to yet-culturable bacteria, which constitute only a small fraction (1–10%) of the global bacterial diversity (1). In contrast, genotypic analyses such as comparative sequencing of PCR-amplified 16S rRNA genes (rDNAs) allow for the identification of both cultured and as-yet-uncultured bacteria (1). Currently, 16S rDNA sequences constitute the largest gene-specific data set, and the number of entries in generally accessible databases is increasing continually (currently ≈30,000), making 16S rDNA-based identification of unknown bacteria isolates more and more likely. Conventional 16S rDNA sequencing is expensive and time-consuming because of the enzymatic reactions and electrophoresis or chromatography steps involved. Hence, it still is not suited for massive parallel testing as required in clinical, pharmaceutical, or environmental microbiology. Compared with standard 16S rDNA sequencing, alternative nonelectrophoretic techniques such as sequencing by MS (2) or pyrosequencing (3) are faster but generate only short read lengths, limiting their utility in 16S rDNA-based bacterial identification.

Here we introduce the concept of base-specific fragmentation of PCR-amplified DNA followed by MS analysis of the resulting fragment pattern as a method for rapid 16S rRNA-based bacterial identification. To exemplify the concept, uracil-DNA-glycosylase-based generation of T-specific fragment patterns was combined with analysis by matrix-assisted laser desorption ionization–time-of-flight (MALDI-TOF) MS for identification of environmental pathogenic and nonpathogenic bacteria strains. The inherent accuracy of mass spectrometric analysis and speed of signal acquisition make this an attractive candidate for high-throughput microbial identification.

Materials and Methods

***Bordetella* Species and As-Yet-Uncultured Bacteria.** The following *Bordetella* species were used in this study: *Bordetella trematum* DSM 11334^T, *Bordetella avium* DSM 11332^T, obtained from the Deutsche Sammlung von Mikroorgansimen und Zellkulturen (Braunschweig, Germany), and *Bordetella petrii* DSM 12804^T, which was recently isolated in our laboratory (4). Cultivation of strains and PCR amplification of nearly full-length 16S rDNA (≈1,500 bp) was described previously (4). Yet-uncultured bacteria of anaerobic, organochlorine-reducing microbial consortia were represented by 16S rDNA sequence regions (≈350 or ≈1,500 bp) directly PCR-amplified and cloned from dechlorinating bioreactors (16S rDNA amplicons B and SHA; refs. 5 and 6).

MS on Base-Specific Fragmentation Patterns of Amplified 16S rDNA. *PCR.* Aliquots (1-μl) of 16S rDNA amplicons obtained from standard PCR in previous studies (5, 6) served as templates for PCR in the presence of dUTP instead of dTTP. For each individual DNA, PCR was performed in triplicate. Reactions (50-μl) contained 1× PCR reaction buffer, 5 pmol of biotinylated forward primer TPU1 (5′-AGAGTTTGATCMTGGCTCAG-3′, *Escherichia coli* 16S rRNA position 8–27 modified from Weisburg *et al.*, ref. 7), 15 pmol of nonbiotinylated reverse primer RTU334a (5′-TGCCACCCGTAGGTGTATGG-3′, *E. coli* positions 334–353, this study), used only for yet-uncultured bacteria B1-10 and B1–12, or RTU334b (5′ TGCCTCCCGTAGGAGTCTGG 3′, *E. coli* positions 334 to 353, this study), 1.5 mM $MgCl_2$, 200 μM dATP/dGTP/dCTP, 600 μM dUTP, and 1.25 units of AmpliTaq polymerase. The temperature profile consisted of 5 min initial denaturation at 95°C followed by 35 cycles of 93°C for 1 min, 60°C for 1 min, and 72°C for 1 min followed by a final extension at 72°C for 7 min.

Immobilization of PCR products. To the crude PCR product, 50 μg of prewashed streptavidin-coated paramagnetic beads (Dynal, Oslo) in 45 μl of 2× BW buffer (10 mM Tris·HCl, pH 7.5/1 mM EDTA/2 M NaCl) were added and incubated for 20 min at room temperature.

Abbreviations: rDNA, rRNA gene; MALDI-TOF, matrix-assisted laser desorption ionization–time-of-flight; AAB, *Alcaligenes–Achromobacter–Bordetella*.

§To whom reprint requests should be addressed. E-mail: dvandenboom@sequenom.com.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MICROBIOLOGY

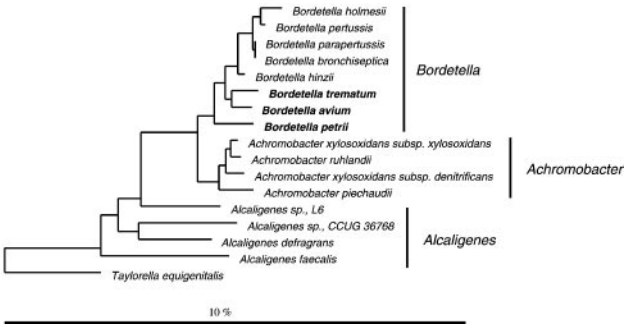

Fig. 1. 16S rDNA-based trees showing the phylogenetic relationship of bacteria of the *Alcaligenes–Achromobacter–Bordetella* (AAB) complex. (Scale bar, 10% estimated sequence divergence.)

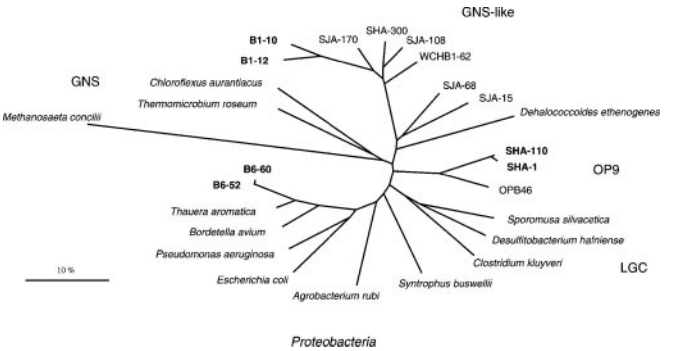

Fig. 2. 16S rDNA-based trees showing the phylogenetic relationship of yet-uncultured B and SHA bacteria from anaerobic bioreactors used for biological treatment of chlorinated compounds. The species used for MALDI/TOF MS analysis of base-specific 16S rDNA fragmentation patterns are shown in bold. GNS, green nonsulfur bacteria; GNS-like, domain-level phylogenetic cluster related to green nonsulfur bacteria; OP9, domain-level phylogenetic cluster OP9; LGC, Gram-positives with a low DNA G+C content. (Scale bar, 10% estimated sequence divergence.)

Preparation of single-stranded templates. Streptavidin-beads carrying immobilized PCR products were incubated with 0.1 N NaOH for 5 min at room temperature. After removing the supernatant, beads were washed three times with 50 μl of 10 mM Tris·HCl, pH 8.0.

Generation of nucleotide-specific abasic sites. Immobilized single-stranded biotinylated PCR products were redissolved in 15 μl of uracil-DNA-glycosylase buffer (60 mM Tris·HCl, pH 7.6/1 mM EDTA). Two units of uracil-DNA-glycosylase (Fermentas, Hanover, PA) were added, and the mixture was incubated for 45 min at 37°C. Beads were washed twice each with 25 μl of 10 mM Tris·HCl, pH 8.0.

Elution of products from solid-phase and phosphate backbone cleavage. To elute solid-phase bound single-stranded DNA, the beads were suspended in 12 μl of aqueous NH_3, incubated at 60°C for 10 min, and cooled to 4°C. The supernatant was transferred to a new tube and subsequently heated to 95°C for 10 min followed by 80°C for 11 min with an open lid.

MALDI-TOF MS analysis. Samples were analyzed from SpectroCHIP arrays. Fifteen nanoliters of analyte were deposited onto a SpectroCHIP (SEQUENOM) by using a nanopipette with a piezoelectric pump. All spectra were recorded by using a Bruker Biflex DE mass spectrometer (Bruker, Billerica, MA).

Sample processing (96- and 384-microtiterplate format) and analysis by MALDI-TOF MS can be completed within 2 h post-PCR by using the described protocol. With further advances in molecular biologic techniques, this time can be shortened substantially.

All experiments were carried out in triplicate to validate the reproducibility of the described principle.

Computational Analyses. Standard comparative 16S rDNA sequence analysis and reconstruction of phylogenetic trees were done by using the ARB software package (8). All mass spectra analysis and automated data interpretation were performed by using in-house software developed by SEQUENOM.

Results

An ≈350-bp region of 16S rDNAs corresponding to *E. coli* position 8–353 were PCR-amplified from each of three *Bordetella* species and six yet-uncultured bacteria. The phylogenetic affiliations of the *Bordetella* species and yet-uncultured bacterial species used in this study are shown in Figs. 1 and 2.

Fig. 3 depicts a representative mass spectrum of a 16S rDNA fragmentation pattern specific for the as-yet-uncultured bacterium B6-52. To facilitate interpretation of the spectrum, all expected mass signals of terminal and internal fragments calculated from the database reference sequence by "virtual" cleavage at all positions of thymine and uracil, respectively, have been cross-referenced with a number corresponding to the mass signals listed in Fig. 4.

Fig. 4 displays all mass signals that could be detected automatically in this spectrum correlating them with their expected sequence (provided in the description column), location within the untreated PCR amplicon (described as @position), reference mass, and chemical nature at the 3′ end. The latter refers to the fact that under the described conditions, the phosphate backbone cleavage is a sequential process that can proceed via hydrolysis and elimination reactions (D.v.d.B., N.S., R. Worl, S. Kammerer, A. Braun, S.B., A. Ruppert, and C.J., unpublished results). Our protocol was optimized to generate 90% of the products as the hydrolysis product. All side products generated during phosphate backbone cleavage, however, can be identified unambiguously from their molecular masses. They are classified as "Side" in Fig. 4 and usually are much lower in intensity but nevertheless are considered during the simulation so that they do not interfere with unambiguous discrimination of strains.

The informative mass signal pattern for strain B6-52 consists of 32 expected terminal and internal cleavage products. All the corresponding mass signals were observed experimentally in the MALDI-TOF mass spectrum and thus allowed unambiguous identification. Twenty additional mass signals listed in Fig. 4, seemingly increasing the complexity of the mass spectrum (as evident from the comparison of dotted lines with actual mass signals), have been identified unambiguously as side products of the phosphate backbone cleavage and thus do not interfere with automated interpretation of the mass spectrum. One further mass signal separately classified in Fig. 4 is called "Last TR", which refers to the 3′-terminal fragment of the PCR amplicon (the last fragment in the amplicon). A known feature of *Taq* polymerases is their tendency for nontemplated addition of nucleotides to the 3′ end. We observed this phenomenon. In addition to the mass signal corresponding to the 3′ fragment, which usually is the lowest in intensity, we also observed a fragment suggesting the addition of dAMP (3′ fragment + 313.2 Da) and dCMP (3′ fragment + 289.2 Da). The addition of dUMP or dGMP was less frequent, and the overall degree of nontemplated addition varied with the sequence context. All of these nontemplated additions have been included automatically in identification of fragments and thus are listed correspondingly. In the spectrum depicted in Fig. 3, three types of "Last" fragments have been identified (+U, +C, and +A), whereas the nonelongated fragment is barely visible (marked by the unreferenced dotted line preceding peak number 39). Of course the

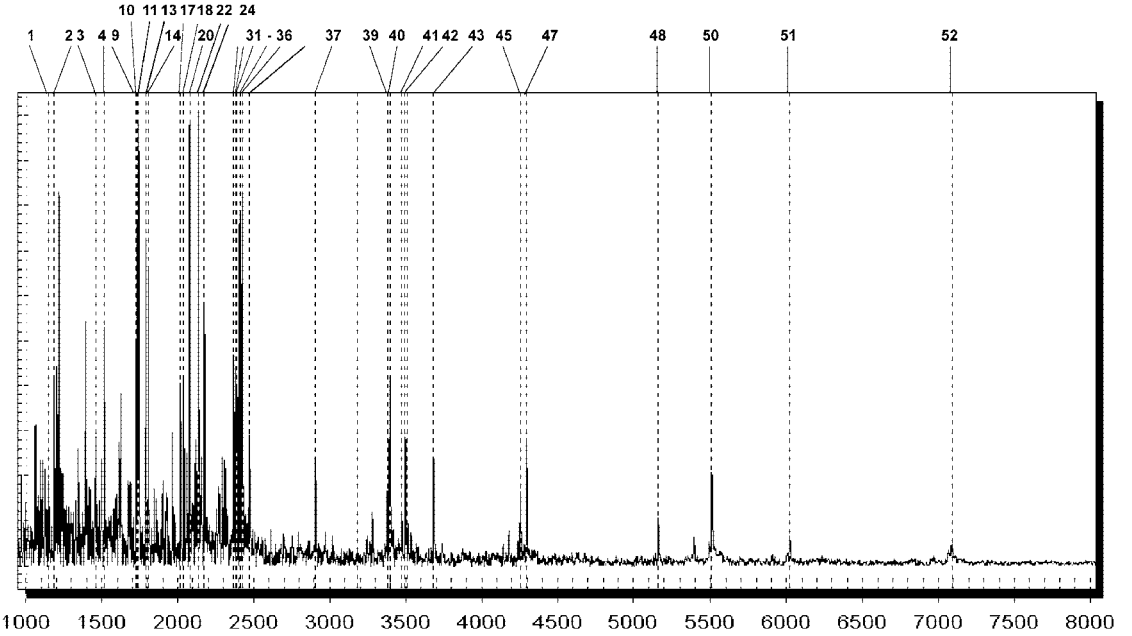

Fig. 3. Mass spectrum displaying all signals generated by T-specific fragmentation of the 16S rDNA amplicon of strain B6-52. All expected mass signals corresponding to main cleavage products calculated from the reference sequence are marked with a dotted line. To simplify identification, the expected mass signals have been cross-referenced with the corresponding peak numbers in Fig. 4. All expected mass signals have been detected and allowed unambiguous identification of this bacteria strain. One nonreferenced dotted line overlays a low-intensity signal, which corresponds to the last fragment of the amplicon. Because of a nontemplated addition at the 3′ end of the PCR amplicon, this mass signal is shifted by the mass of the respective nucleotide added (see also Fig. 4).

removal of uracil also proceeds when dUMP is added nontemplated at the 3′ end of the PCR product and the mass signal of the corresponding product thus is not grouped with the two other fragments (cross-compare peaks 39, 41, and 42 in Figs. 3 and 4).

Further examples of typical mass spectra are shown in Figs. 5 and 6. Base-specific fragmentation of *B. avium*, *B. trematum*, and *B. petrii* 16S rDNA amplicons resulted in unique mass spectra despite the high sequence similarities of 97.3–99.1%. To avoid identification problems associated with single species-specific rDNA mass peaks (see *Discussion*), only patterns of mass peaks were used. Experimentally determined mass signal patterns were compared with mass signal patterns calculated from 50 published 16S rDNA sequences including 13 closely related species of the *Alcaligenes–Achromobacter–Bordetella* (AAB) complex (9). Bacteria then were ranked according to the number of nonidentical (additional or missing) mass peaks. The resulting theoretical peak patterns for the *Bordetella* species under study are shown in Table 1. For better visualization, only those peaks were included that are not present in all bacteria of the list (discriminating peaks). Compared with *B. avium* the closest species, *Bordetella hinzii*, showed four additional as well as two missing peaks. For *B. trematum* the most similar mass spectrum again is that of *B. hinzii* (three additional and two missing peaks). *B. petrii* clearly is more separated from other AAB species as compared with the former two *Bordetella* species. The more closely related species, *B. trematum*, showed six additional and six missing peaks. All calculated mass pattern differences have been validated experimentally.

For *Bordetella pertussis*, *Bordetella parapertussis*, and *Bordetella bronchiseptica* identical mass spectra were calculated because of 100% sequence identity within the 16S rDNA sequence region used for MS. This result is consistent with previous studies concluding that these organisms should be considered as strains or subspecies of a single species rather than separate species (9).

MALDI-TOF MS analysis of base-specific 16S rDNA fragmentation patterns also was tested with further yet-uncultured bacteria (SHA and B) represented by 16S rDNAs directly PCR-amplified and cloned from dechlorinating bioreactors (5, 6). These six selected rDNA sequences displayed pairwise sequence similarities of 92.0 (B1-10 and B1-12), 95.9 (SHA-1 and SHA-110), and 99.4% (B6-52 and B6-60; Fig. 2). Base-specific fragmentation and subsequent MS analysis generated specific spectra for all rDNA amplicons. Even strains B6-52 and B6-60, which differ in only two nucleotide positions, could be identified unambiguously on the basis of their unique mass signal pattern (Fig. 7). A comparison of experimentally observed mass spectra with the above-mentioned data set of 50 simulated mass spectra resulted in unambiguous identification of all SHA and B sequences (data not shown).

Discussion

The goal of this study was to combine MS with 16S rDNA-based genotyping to develop a rapid technique for bacterial identification. The described technique avoids the time-consuming and expensive steps inherent in standard 16S rDNA sequencing. Previous work has shown the potential of MS in bacterial genotyping. In those studies MS was used to distinguish bacterial species by size determination of PCR-amplified specific marker genes (10, 11). However, for accurate bacterial identification this approach is limited, because length heterogeneities of specific marker genes do not provide sufficient discriminatory power. Instead, the technique described here makes use of sequence variations among 16S rDNAs of different species as revealed by base-specific fragmentation of 16S rDNA amplicons before MALDI-TOF MS.

To allow 16S rDNA amplification from a wide range of bacterial species, primers were selected that targeted highly conserved regions of 16S rDNA. Because of its length, a full analysis of 16S rDNA would generate too many cleavage fragments and complicate mass spectrometric analysis. To keep the mass signal complexity within the capability limits of conventional MALDI-TOF MS, 16S rDNA sequence regions rather

Spectrum B6-52, Peak Nr.	Detected Mass signal in Da	Reference Mass signal in Da	Reference Type	Description
1	1146,91	1146,73	MAIN	5p-GCA-3hos @108; 5p-AGC-3hos @234
2	1186,50	1186,75	MAIN	5p-GAG-3hos @101
3	1460,11	1459,94	MAIN	5p-CAGA-3hos @228
4	1516,16	1515,96	MAIN	5p-AGGG-3hos @245
5	1610,03	1608,98	SIDE	5p-GCCCA-3p @120
6	1625,43	1624,98	SIDE	5p-CGCGC-3p @151; 5p-CGGCC-3p @74
7	1673,67	1673,03	SIDE	5p-GCAAG-3p @49
8	1689,91	1689,20	SIDE	5p-GGGAC-3p @306
9	1724,51	1725,10	MAIN	5p-GCCCA-3hos @120
10	1732,89	1733,12	MAIN	5p-ACACA-3hos @43
11	1740,99	1741,09	MAIN	5p-CGCGC-3hos @151; 5p-CGGCC-3hos @74
12	1770,91	1771,13	SIDE	5p-GCAAG-3bos @49
13	1788,99	1789,15	MAIN	5p-GCAAG-3hos @49
14	1804,96	1805,14	MAIN	5p-GGGAC-3hos @306
15	1899,13	1898,16	SIDE	5p-ACGCCC-3p @167
16	1963,06	1962,21	SIDE	5p-GAACGC-3p @23
17	2013,75	2014,28	MAIN	5p-ACGCCC-3hos @167
18	2037,72	2038,31	MAIN	5p-ACCGCA-3hos @160
19	2060,07	2060,32	SIDE	5p-GAACGC-3bos @23
20	2078,10	2078,33	MAIN	5p-GAACGC-3hos @23
21	2116,48	2116,34	SIDE	5p-AGCGGG-3bos @276; 5p-GGCGAG-3bos @83
22	2134,15	2134,35	MAIN	5p-AGCGGG-3hos @276; 5p-GGCGAG-3hos @83
23	2156,73	2156,36	SIDE	5p-GGGGGA-3bos @130
24	2173,96	2174,38	MAIN	5p-GGGGGA-3hos @130
25	2252,22	2251,40	SIDE	5p-CGCGCAA-3p @205
26	2275,10	2275,42	SIDE	5p-AAAGGCC-3p @250
27	2291,88	2291,42	SIDE	5p-CGGAACG-3p @112
28	2307,77	2307,42	SIDE	5p-GGCGGCA-3p @30
29	2349,17	2349,50	SIDE	5p-CGCGCAA-3bos @205
30	2355,48	2355,47	SIDE	5p-GAGAGGA-3p @285
31	2367,17	2367,52	MAIN	5p-CGCGCAA-3hos @205
32	2383,30	2383,52	MAIN	5p-CGGAGCC-3hos @197
33	2390,63	2391,54	MAIN	5p-AAAGGCC-3hos @250
34	2407,18	2407,54	MAIN	5p-CGGAACG-3hos @112
35	2423,16	2423,54	MAIN	5p-GGCGGCA-3hos @30
36	2471,72	2471,59	MAIN	5p-GAGAGGA-3hos @285
37	2906,02	2905,86	MAIN	5p-CCGCCACAC-3hos @296
38	3280,71	3279,05	SIDE	5p-GGCGAACGGG-3p @90
39	3377,57	3379,17	LAST	5p-ACGGGAGGCA-3hos @332
40	3395,37	3395,17	MAIN	5p-GGCGAACGGG-3hos @90
41	3473,55	3472,26	LAST	5p-ACGGGAGGCA-C-3OH @332
42	3496,60	3496,28	LAST	5p-ACGGGAGGCA-A-3OH @332
43	3684,55	3684,35	MAIN	5p-GGAGCGGCCGA-3hos @214
44	4179,85	4178,66	SIDE	5p-AACGCGGCGAAAG-3p @137
45	4255,54	4254,75	MAIN	5p-ACCAAGGCGACGA-3hos @258
46	4276,37	4276,76	SIDE	5p-AACGCGGCGAAAG-3bos @137
47	4294,70	4294,78	MAIN	5p-AACGCGGCGAAAG-3hos @137
48	5162,48	5162,33	MAIN	5p-GAGACACGGCCCAGAC-3hos @312
49	5489,47	5489,53	SIDE	5p-CGAACGGCAGCACGGGC-3bos @55
50	5507,34	5507,54	MAIN	5p-CGAACGGCAGCACGGGC-3hos @55
51	6020,04	6020,87	MAIN	5p-GAGGGGGAAAGCGGGGGA-3hos @174
52	7062,53	7063,33	MAIN	Biotinylated PCR primer

Fig. 4. Detected mass signals for a B6-52 fragmentation pattern corresponding to the mass spectrum displayed in Fig. 3. For all signals, the composition of the corresponding cleavage fragments was assigned automatically. Their exact sequence and location within the uncleaved PCR amplicon is provided in the description column. Different reference fragments were used in congruence with the possible cleavage reaction products. Main, main cleavage product; Side, side products of phosphate backbone cleavage, which occur with significantly lower intensity but can be assigned unambiguously; Last, products at the 3′ end of the amplicon, which often are elongated by one nucleotide because of 3′-terminal transferase activity of *Taq*; hos, hydrolysis products of phosphate backbone cleavage (main product); bos, β-elimination products of phosphate backbone cleavage. Fragments ending on a 3′-terminal phosphate group generated by δ elimination are marked with 3 P (both termed Side).

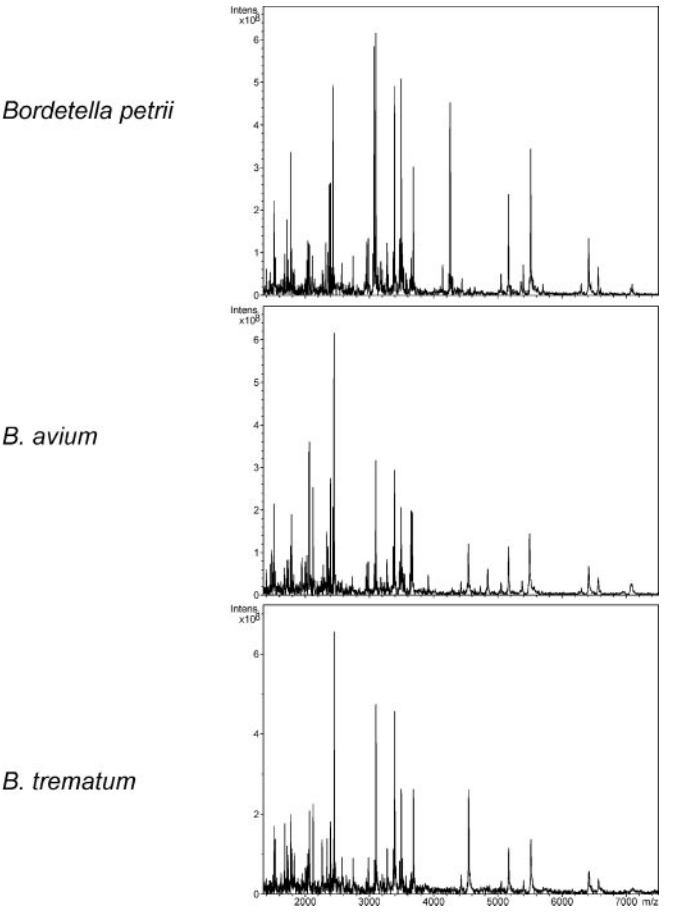

Fig. 5. Mass spectra of base-specific 16S rDNA fragments of *B. petrii*, *B. avium*, and *B. trematum*. For display, mass spectra were smoothed by a five-point average (Savitzky–Golay method) and baseline-subtracted.

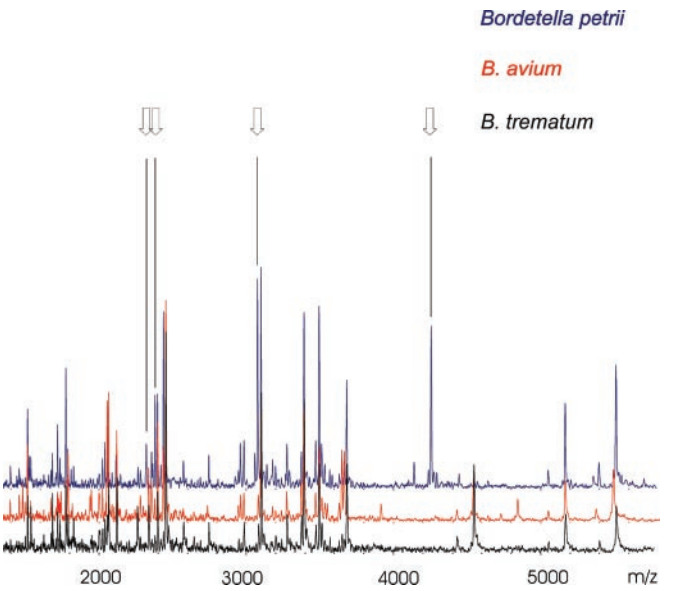

Fig. 6. Overlay of mass spectra of *B. petrii* (blue), *B. avium* (red), and *B. trematum* (black). Peaks specific for *B. petrii* are marked by the arrows. For display, mass spectra were smoothed by a five-point average (Savitzky–Golay method) and baseline-subtracted.

than full-length genes were PCR-amplified. The forward primer TPU1 in combination with reverse primers RTU334a or RTU334b amplified a sequence region corresponding to position 8–353 of the *E. coli* 16S rRNA. This region was chosen because it includes highly variable areas V1 and V2 (12), allowing discrimination even among closely related bacterial species (13, 14). As described, MALDI-TOF MS analysis of base-specific fragmented 16S rDNA amplicons was tested on two classes of bacteria for which 16S rDNA sequencing currently is the method of choice for identification.

(*i*) *Bordetella* species that belong to a group of closely related pathogenic and environmental bacteria including members of the genera *Alcaligenes* and *Achromobacter* (AAB complex; Fig. 1). At present, comparative 16S rDNA sequencing

Table 1. Ranking of base-specific 16S rDNA fragmentation mass patterns of AAB species according to the number of identical and nonidentical mass peaks compared with the mass spectrum of *B. avium*

	Calculated mass, Da																																							
Organism†	1,187	1,436	1,709	1,725	1,741	2,054	2,062	2,070	2,102	2,134	2,368	2,392	2,448	2,464	2,745	2,970	3,082	3,395	3,411	3,652	3,668	3,684	4,255	4,528	4,544	4,560	4,624	5,162	5,492	5,507	5,524	5,636	5,853	5,965	6,021	6,230	6,407	6,559	6,792	7,064	7,088
Bav	■			■				■		■				■				■		■	■				■			■	■								■	■		■	
Bhi	■			■	▨	▨		■		■				■				■		■		▨			■			■			▨						■	■		■	
Btr	■			■		▨		■		■				■				■		■		▨			■			■		▨						▨	■			■	
Bper/br/par	■			■	▨	▨		■		■	▨			■				■		■		▨	▨			▨					▨						■	■		■	
Bho	■		▨	■	▨			■		■	▨			■				■		■		▨	▨				▨	■					▨					■		■	
Acru	■	▨				▨				■			▨			▨		■		■		▨			■			■	■					▨			■			■	
Acxx	■	▨				▨				■			▨			▨		■		■		▨			■			■	■					▨			■				▨
Acxd	■	▨				▨				■			▨			▨		■			■	▨			■			■	■					▨			■				▨
Bpe				■		▨		■				▨	▨				▨	■		■		▨	▨					■		▨							■	■			▨
Acpi	■	▨					▨			■		▨	▨					■			■				■			■	■			▨					■				▨
Afa	■					▨			▨				▨		▨	▨	▨		▨			▨		▨				■							▨		■		▨	■	

*Only discriminating mass peaks >1,186 Da are shown. Mass peaks identical to those of *B. avium* are marked in black. Additional mass peaks are marked in grey.
†Bav, *B. avium*; Bhi, *B. hinzii*; Btr, *B. trematum*; Bper/br/par, *B. pertussis*/*B. bronchiseptica*/*B. parapertussis*; Bho, *B. holmesii*; Acru, *A. ruhlandii*; Acxx, *Achromobacter xylosoxidans* subspecies *xylosoxidans*; Acxd, *A. xylosoxidans* subspecies *denitrificans*; Bpe, *Bordetella petrii*; Acpi, *A. piechaudii*; Afa, *Alcaligenes faecalis.*

is the most reliable method for the identification of AAB species because of the lack of specific phenotypic traits (9).

(*ii*) Yet-uncultured bacteria of selected anaerobic, organochlorine-reducing microbial consortia (Fig. 2). Currently, these bacteria are represented only by their 16S rDNA, which was PCR-amplified directly and cloned from bioreactors used for the biological treatment of chlorinated compounds (5, 6).

Despite high sequence similarities of 92.0 to 99.4%, MALDI-TOF MS analysis of base-specific 16S rDNA fragments resulted in unique mass spectra for each of the bacteria strains. We based the identification of strains on a comparison of mass peak patterns, because identification based on single species-specific mass peaks is limited; the number of species-specific unique mass peaks will decrease with a growing number of reference spectra/species. An analysis based on single species-specific peaks could lead to an erroneous assignment of an acquired sample spectrum as one of the reference species, especially if they randomly have the one species-specific peak in common.

We were interested also in the potential of this method for clustering and classification of unknown bacteria. As a preliminary assessment, we used the simulated peak pattern to construct most-parsimonious phylogenies of the 13 studied bacteria. The peak patterns in Table 1 were considered as a binary code, and the complete space of phylogenies was searched for the most-parsimonious trees that could explain the observed sequence information. This calculation was done by using the program PENNY, which is part of the PHYLIP software package (15). *Alcaligenes faecalis* was used as the outgroup. In total, 10 most-parsimonious trees were found, and the strict consensus of these most-parsimonious trees can be found in Fig. 8 (being artificially rooted at the branching point between its outgroup and the remaining species). All branching points found in the phylogenetic tree of Fig. 8 are comparable to the phylogenetic relationships described earlier, and unresolved edges (such as the placement of *B. petrii*) correspond to speciation events that are not completely decided in microbiology either, which suggests that at least for this example almost all the "phylogenetic information" contained in the DNA sequences of the examined species (see Figs. 1 and 2) can be found also in their fragmentation spectra. It therefore might be feasible to cluster and classify new isolates based on the fragmentation pattern.

This technique uses well established genotypic markers, the 16S rDNA or regions thereof, allowing universal identification of both cultured and as-yet-uncultured bacteria (1, 8). The fact that only a minor part of the bacterial diversity can be cultivated in the laboratory and the increasing importance of yet-uncultured bacteria in both medical and environmental microbiology (1, 16) makes MS analysis of 16S rDNA specific fragmentation patterns superior to protein- or whole cell-based identification by MS requiring *in vitro* cultivation of bacteria before analysis. Similar to many other phenotypic identification methods, the latter approach is prone to growth-dependent variations resulting in

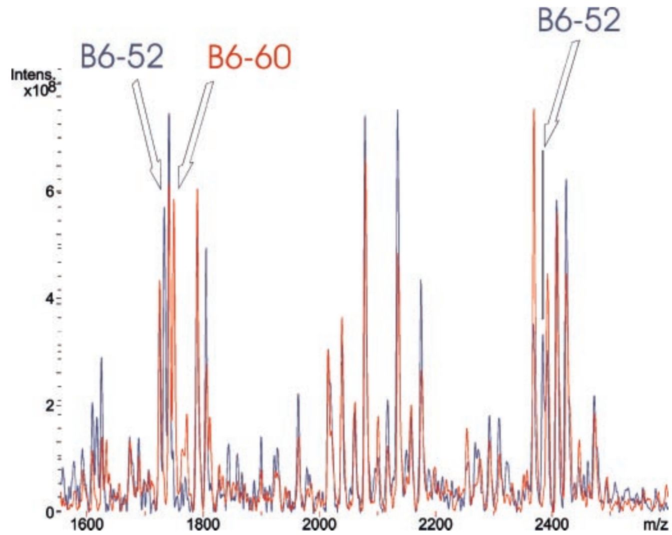

Fig. 7. Overlay of base-specific 16S rDNA fragmentation mass spectra of yet-uncultured bacteria B6-52 and B6-60 (B6-52 spectrum is indicated in blue, and B6-60 is indicated in red). Partial 16S rDNA sequences (354-bp) differ in only two nucleotide positions. Fragmentation patterns of the respective bacteria differ in three mass signals. B6-52 shows base-specific fragmentation mass signals at 1,733.1 and 2,383.5 Da, both missing in B6-60. In contrast, B6-60 shows a fragmentation mass signal at 1,749.1 Da, not present in the respective B6-52 spectrum. The arrows indicate the described mass peaks allowing discrimination of the strains.

MICROBIOLOGY

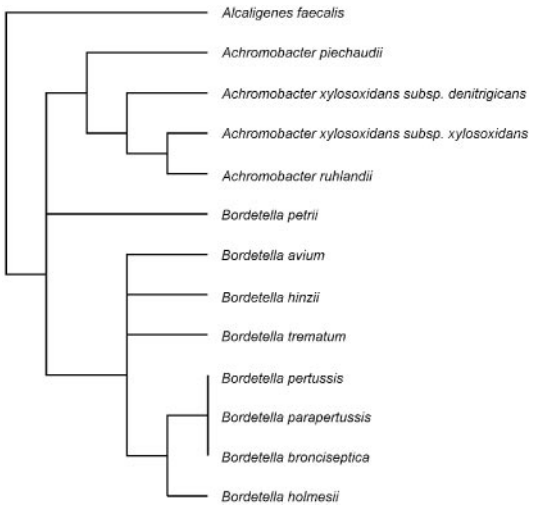

Fig. 8. Most-parsimonious tree data based on the mass signal patterns generated by T-specific fragmentation of 16S rRNA PCR amplicons.

problems with reproducibility (17, 18). In contrast, MALDI-TOF MS analysis of base-specific 16S rDNA fragmentation patterns (as well as most other DNA-based identification methods) is not biased by the cultivation conditions (17, 18). The absence of any time-consuming electrophoresis or chromatography step is an important advantage of MS analysis when compared with conventional 16S rDNA sequencing or rDNA fingerprinting by restriction analysis (19). Other sequencing techniques such as Sanger sequencing by MS (2) or pyrosequencing (3) have been proposed recently as viable alternatives to gel- or capillary-DNA sequencing. However, their short read lengths of 20–30 nucleotides strongly limit the use for 16S rDNA-based bacterial identification, because additional typing steps are required, e.g., the use of taxon-specific PCRs (20). Compared with these techniques, the principle of base-specific fragmentation of PCR amplicons combines the possibility of scanning larger sequence regions for bacterial typing with the speed and accuracy of MALDI-TOF MS.

The experimental procedure required for PCR amplification and subsequent base-specific fragmentation is rather simple and could make this method an ideal candidate for high-throughput identification in clinical diagnostics and environmental microbiology. It has great potential for bacterial and fungal pathogens that cause problems in routine conventional diagnostics, e.g., mycobacteria, *Chlamydia*, and dermatophytes. The MS-based identification approach is not restricted to 16S rDNA but can be expanded to other genotypic markers, e.g., typing of viral strains or typing of MDRR regions, which will broaden its applicability further.

Recently, other methods for generating base-specific fragmentation patterns of PCR amplicons have been described. They require either a subsequent post-PCR primer extension reaction to introduce a cleavable nucleotide (21) or rely on transcription of the PCR amplicon in RNA molecules that are cleaved by the means of a base-specific RNase (R. Hartmer, N.S., S.B., C.P. Rodi, F. Hillenkamp, C.J., and D.v.d.B., unpublished results). Because they follow the same principle, they as well could be applied to bacterial typing, and further studies may verify their suitability.

1. Amann, R. I., Ludwig, W. & Schleifer, K. H. (1995) *Microbiol. Rev.* **59,** 143–169.
2. Nordhoff, E., Luebbert, C., Thiele, G., Heiser, V. & Lehrach, H. (2000) *Nucleic Acids Res.* **28,** e86.
3. Ronaghi, M. (2000) *Anal. Biochem.* **286,** 282–288.
4. von Wintzingerode, F., Schattke, A., Rösick, U., Siddiqui, R., Göbel, U. B. & Gross, R. (2001) *Int. J. Syst. Evol. Microbiol.* **51,** 1257–1265.
5. von Wintzingerode, F., Landt, O., Ehrlich, A. & Göbel, U. B (2000) *Appl. Environ. Microbiol.* **66,** 549–557.
6. Schlötelburg, C., von Wintzingerode, F., Hauck, R., Hegemann, W. & Göbel, U. B. (2000) *Int. J. Syst. Evol. Microbiol.* **50,** 1505–1511.
7. Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J., (1991) *J. Bacteriol.* **173,** 697–703.
8. Ludwig, W., Strunk, O., Klugbauer, S., Klugbauer, N., Weizenegger, M., Neumaier, J., Bachleitner, M. & Schleifer, K. H. (1998) *Electrophoresis* **19,** 554–568.
9. Gerlach, G., von Wintzingerode, F., Middendorf, B. & Gross, R. (2001) *Microbes Infect.* **3,** 61–72.
10. Hurst, G. B., Weaver, K., Doktycz, M. J., Buchanan, M. V., Costello, A. M. & Lidstrom, M. E. (1998) *Anal. Chem.* **70,** 2693–2698.
11. Muddiman, D. C., Wunschel, D. S., Liu, C., Pasa-Tolic, L., Fox, K. F., Fox, A., Anderson, G. A. & Smith, R. D. (1996) *Anal. Chem.* **68,** 3705–3712.
12. Van de Peer, Y., Chapelle, S. & De Wachter, R. (1996) *Nucleic Acids Res.* **24,** 3381–3391.
13. Kataoka, M., Ueda, K., Kudo, T., Seki, T. & Yoshida, T. (1997) *FEMS Microbiol. Lett.* **151,** 249–255.
14. Kullen, M. J., Sanozky-Dawes, R. B., Crowell, D. C. & Klaenhammer, T. R. (2000) *J. Appl. Microbiol.* **89,** 511–516.
15. Felsenstein, J. (1993) PHYLIP, Phylogeny Inference Package (Dept. of Genetics, Univ. of Washington, Seattle), Version 3.5c.
16. Relman, D. A. (1999) *Science* **284,** 1308–1310.
17. Lay, J. O. J. (2000) *Trends Anal. Chem.* **19,** 507–516.
18. van Baar, B. L. M. (2000) *FEMS Microbiol. Rev.* **24,** 193–219.
19. Heyndrickx, M., Vauterin, L., Vandamme, P., Kersters, K. & De Vos, P. (1996) *J. Microbiol. Methods* **26,** 247–259.
20. Monstein, H., Nikpour-Badr, S. & Jonasson, J. (2001) *FEMS Microbiol. Lett.* **199,** 103–107.
21. Shchepinov, M. S., Denissenko, M. F., Smylie, K. J., Worl, R. J., Leppin, A. L., Cantor, C. R. & Rodi, C. P. (2001) *Nucleic Acids Res.* **29,** 3864–3872.