International Journal of Systematic and Evolutionary Microbiology (2002), **52**, 91–100 Printed in Great Britain

Identification of coryneform bacteria and related taxa by Fourier-transform infrared (FT-IR) spectroscopy

Helene Oberreuter, Herbert Seiler and Siegfried Scherer

Microbial Ecology Group, Department of Basic Biological Sciences, Technical University of Munich, Weihenstephaner Berg 3, D-85350 Freising, Germany

Author for correspondence: S. Scherer. Tel: +49 8161 713516. Fax: +49 8161 714512. e-mail: siegfried.scherer@lrz.tum.de

An extensive Fourier-transform infrared (FT-IR) spectroscopy database for the identification of bacteria from the two suborders *Micrococcineae* and *Corynebacterineae* (*Actinomycetales*, *Actinobacteria*) as well as other morphologically similar genera was established. The database consists of averaged IR spectra from 730 reference strains, covering 220 different species out of 46 genera. A total of 192 species are represented by type strains. The identity of 352 reference strains was determined by comparative 16S rDNA sequence analysis and, if necessary, strains were reclassified accordingly. FT-IR frequency ranges, weights and reproducibility levels were optimized for this section of high-G+C Gram-positive bacteria. In an internal validation, 98·1% of 208 strains were correctly identified at the species level. A simulated external validation which was carried out using 544 strains from 54 species out of 16 genera resulted in a correct identification of 87·3% at the species level and 95·4% at the genus level. The performance of this identification system is well within the range of those having been reported in the literature for the identification of coryneform bacteria by phenotypical methods. Coryneform and related taxa display a certain degree of overlapping distribution of different taxonomical markers, leading to a limited differentiation capacity of non-genotypical identification methods in general. However, easy handling, rapid identification within 25 h starting from a single colony, a satisfactory differentiation capacity and low cost, render FT-IR technology clearly superior over other routine methods for the identification of coryneform bacteria and related taxa.

Keywords: coryneform bacteria, *Micrococcineae*, *Corynebacterineae*, FT-IR spectroscopy, identification

INTRODUCTION

Coryneform bacteria and related taxa occur almost everywhere on living and non-living matter in the environment, in soil, on cellulosic plant fibres (Lednická *et al*., 2000), on mammals, on smeared cheeses (e.g. Seiler, 1986; Schubert *et al*., 1996; Bockelmann *et al*., 1997; Kollöffel *et al*., 1997; Valdés-Stauber *et al*., 1997; Carnio *et al*., 1999), even in subsurface sediments (Crocker *et al*., 2000), and in Antarctic samples (Junge *et al*., 1998; Tindall *et al*., 2000). Several strains or species are classified as opportunistic or obligate human pathogens (for review see Funke *et al*., 1997a) as well as animal and plant pathogens (Holt *et al*., 1994). Identification of these high-G+C Gram-positive bacteria is therefore not only of ecological and technological, but also of medical relevance.

Among the methods applied for classification of coryneform bacteria and related taxa in recent years are analyses of polyamine patterns (Altenburger *et al*.,

Abbreviations: *D*, spectral distance; FISH, fluorescence *in situ* hybridization; FT-IR, Fourier-transform infrared; TGGE, temperature-gradient gel electrophoresis; UPGMA, unweighted pair group method algorithm.

A list of coryneform and related or morphologically similar species contained in the FT-IR spectral identification database is available in IJSEM Online (http://ijs.sgmjournals.org). A frequently updated reference spectra library can be obtained from the authors.

1997; Busse & Schumann, 1999) or fatty acids (Kämpfer & Kroppenstedt, 1996), numerical taxonomic analysis (Kämpfer *et al*., 1993) and assessment of the heterogeneity of partial 16S rRNA sequences by temperature-gradient gel electrophoresis (TGGE) (Felske *et al*., 1999). These methods have proven to be valuable but somewhat limited tools for characterization and differentiation of the organisms studied. Other methods have also been reported which placed more emphasis on the identification capability for coryneform genera: analysis of physiological characteristics by the BIOLOG Identification System (Lindenmann *et al*., 1995), the API (RAPID) Coryne database (Funke *et al*., 1997b) or the RapID CB Plus system (Funke *et al*., 1998), comparative 16S rDNA sequence analysis (Bockelmann *et al*., 1997), the use of genus-specific oligonucleotide probes (Kollöffel *et al*., 1997) and fluorescence *in situ* hybridization (FISH) as well as colony hybridization for the analysis of cheese-surface bacteria (Kollöffel *et al*., 1999). Not all of these methods have provided satisfactory results and they are either too laborious and time-consuming to be performed on a routine basis with respect to the molecular methods, or the costs per sample are high, which is the case with commercially available identification systems for the analysis of physiological properties.

Originally introduced by Naumann and co-workers (Helm *et al*., 1991; Naumann *et al*., 1994), Fourier-transform infrared (FT-IR) spectroscopy is a fast technique for classification and identification of micro-organisms (e.g. Holt *et al*., 1995; Goodacre *et al*., 1998; Kümmerle *et al*., 1998; Timmins *et al*., 1998). Moreover, it is a valuable tool for rapid screening of environmental isolates (Tindall *et al*., 2000). The method would gain more attraction if an extensive database were available for the identification of unknown strains, since FT-IR is a very cost-efficient technique which allows a rapid and simple identification of micro-organisms within 25 h (Kümmerle *et al*., 1998).

Actinomycetes have been investigated by FT-IR previously. Haag *et al*. (1996) have placed their emphasis on strains from the families *Thermomonosporaceae*, *Streptosporangiaceae*, *Micromonosporaceae* and *Streptomycetaceae* while including a few strains from different families from the *Corynebacterineae* as well. Klatte (1995) has studied numerous species from the two genera *Corynebacterium* and *Rhodococcus* in detail. However, these investigations have not resulted in comprehensive databases and did not allow us to conclude on the capacity of this method to identify unknown strains.

In 1997, Stackebrandt *et al*. established a novel hierarchic classification system for actinomycetes, the class of the *Actinobacteria* (high-G+C Gram-positive bacteria) (Stackebrandt *et al*., 1997). This taxonomy was recently updated to accommodate new taxa of this quickly expanding group of bacteria (Stackebrandt & Schumann, 2000). Based on this classification system, a spectral database consisting of reference strains from 46 genera out of the two suborders *Micrococcineae* and *Corynebacterineae* (order *Actinomycetales*) was established which allows the rapid and simple identification of isolates of this taxonomic group. Morphologically similar taxa such as strains from the genus *Staphylococcus* were included in the database for practical reasons. In order to check the identities of reference strains, comparative 16S rDNA sequence analysis was carried out for more than half of the reference strains.

METHODS

Strains. A total of 730 strains from international culture collections, commercial starter culture companies and other laboratories provided the reference material. Strains with questionable classification were identified by comparative 16S rDNA sequence analysis. The reference strains represent 220 species out of 46 genera from the two suborders *Micrococcineae* and *Corynebacterineae* (*Actinomycetales*, *Actinobacteria*) as well as morphologically similar genera – see the list of coryneform and related or morphologically similar species contained in the FT-IR spectral identification database in IJSEM Online (http://ijs.sgmjournals.org).

Sample preparation. Sample preparation was performed according to the methods of Kümmerle *et al*. (1998). The cells were incubated at 30 °C for 24 h (tryptone soya agar, containing 15·0 g tryptone, 5·0 g soya peptone, 5·0 g sodium chloride and 15·0 g agar per litre; Oxoid). Few strains yielding a clumpy suspension were subjected to ultrasonication for 2×10 s at 25% power with a Bandelin Sonopuls HD2200, probe MS72 (Bandelin Electronic) in order to improve spectral quality.

FT-IR spectroscopy. All spectra were recorded and evaluated according to the methods of Kümmerle *et al*. (1998) using an IFS-28B FT-IR spectrometer (Bruker). To diminish the difficulties arising from unavoidable baseline shifts and to improve the resolution of complex bands, the first derivation of the digitized original spectrum was used.

The adjustment of FT-IR parameters such as selection of relevant spectral frequency ranges, their weighting and so-called reproducibility levels was done based on Bruker (1996) and Kümmerle *et al*. (1998) but was extended in this work. In order to calibrate the spectral distance values obtained in the identification hit lists, the reproducibility levels were determined for each of the five relevant frequency ranges (spectral windows) W1 (3000–2800 cm^{-1}), W2 (1800–1500 cm^{-1}), W3 (1500–1200 cm^{-1}), W4 (1200–900 cm^{-1}) and W5 (900–700 cm^{-1}). For this procedure, eight species were used which were represented by between five and 42 strains each. The reproducibility levels normalize the different spectral reproducibility variation encountered in each frequency range. Their adjustment allows the threshold value for a valid identification of an isolate at the species level to be set at a spectral distance of 1·0–1·5, approximately. This implies that if the spectral distance between an isolate and the first hit of the identification hit list is higher than 1·5, the isolate cannot be identified. The reproducibility levels differ greatly between different frequency ranges, corresponding to the different degree of spectrum variation within the spectral windows. The subsequent selection of spectral windows and their weighting was done according to the results from the internal validation.

Comparative 16S rDNA sequence analysis. A total of 352 reference strains of questionable identity were subjected to comparative 16S rDNA sequence analysis. The strains were lysed and an amplification PCR of the almost complete 16S rDNA molecule was performed according to the methods of von Stetten *et al*. (1998). Two universal 16S rDNA-binding primers were used for the amplification PCR: 5′f (5′-AGAGTTTGATCCTGGCTCA-3′; positions 8–26 in the *Escherichia coli* numbering system, Brosius *et al*., 1978) and 3′r (5′-CGGCTACCTTGTTACGAC-3′; *E. coli* numbering 1511–1493). The PCR protocol started with a denaturation step for 5 min at 95 °C, followed by 30 cycles of denaturation for 20 s at 95 °C, primer annealing for 40 s at 55 °C and elongation for 2 min at 72 °C each. A final elongation step was added at 72 °C for 5 min. After PCR amplification, the DNA was purified using the QIAquick PCR purification kit (Qiagen) according to the instructions of the manufacturer, followed by a PEG precipitation of the purified product according to the methods of Facius *et al*. (1999). After purification, the samples were subjected to a cycle-sequencing PCR according to the methods of Facius *et al*. (1999), using the ThermoSequenase fluorescent labelled primer cycle-sequencing kit with 7-deaza-dGTP (Amersham Pharmacia Biotech). DMSO (11 %, v/v) and formamide (7 %, v/v) were added to facilitate the cycle-sequencing PCR. Fluorescently labelled primers 5′f, 3′r, 699R [5′-GGGTTG(AGT)GCTCGTT-3′; *E. coli* numbering 1114–1100] or 609V [5′-TTAGATACCCT(AG)GTAGT-3′; *E. coli* numbering 788–804] were used for the cycle-sequencing PCR (primers 699R and 609V: Ludwig & Strunk, http://www.biol.chemie.tu-muenchen.de/pub/ARB/documentation/ARB.ps). Independently of the primer applied, the protocol for the cycle-sequencing PCR started with a first denaturation step for 5 min at 88 °C, followed by 25 cycles of denaturation for 30 s at 88 °C, 25 s at the primer specific annealing temperature and elongation for 3 min at 50 °C. A final elongation step at 50 °C for 5 min was added. Annealing temperatures for the different cycle sequencing primers used were as follows: 42 °C for 5′f, 45 °C for 3′r, 37 °C for 699R and 35 °C for 609V. Sequencing was performed on a LI-COR sequencer (MWG Biotech), typically yielding sequence lengths of approximately 800–1000 bases per run.

Identification of the strains was accomplished by comparison of the partial sequences with sequences from databases such as GenBank (Altschul *et al*., 1997), the RDP database (Maidak *et al*., 2000) and the ARB database (Ludwig & Strunk, see above). According to Stackebrandt & Goebel (1994), a similarity of more than 97 % strongly points to species identity between query sequence and retrieved database sequences, although there are exceptions to this rule (e.g. Fox *et al*., 1992; Lechner *et al*., 1998). Identification of a sequence in the ARB software environment was accomplished by addition of the aligned query sequence into a validated and optimized tree based on aligned 16S rDNA sequences by maximum-parsimony analysis while keeping its topology constant (Ludwig *et al*., 1998).

Internal validation. Performing an internal validation means that newly recorded single spectra from strains already available in the spectral library are tested against the database. Randomly selected single spectra of 208 strains from 208 species out of 41 genera provided the test set for this internal validation, representing a cross-section of the coryneform bacteria and related organisms. The reference set consisted of average spectra of the complete spectral library of 730 strains. The average spectra from the 208 test strains were included in the reference set which did not comprise the single spectra of the test set. The test set spectra were then identified by the reference set and the result was determined at the strain, species and genus level by evaluating the relationship between the test spectrum and the first hit in the identification hit list. The result was counted as a correct identification at the strain level if the test spectrum was identified with a spectral distance value $D < 1{\cdot}5$ by its corresponding average spectrum. In case of an identification of the test spectrum by an average spectrum of a strain belonging to the same species, the result corresponded to a correct identification at the species level. Finally, if at least the genus matched, the result was counted as a correct identification at the genus level. The results of identification of all test spectra were averaged to obtain the so-called percentage correctness of identification. This parameter was determined for each frequency range W1–W5 separately and subsequently also for combinations of the five ranges altogether or only spectral windows W2–W5. The quality of identification obtained for each frequency range separately specified their eventual weighting. For comparison, the correctness of identification was also determined for the parameters used for the identification of yeasts by FT-IR spectroscopy by Kümmerle *et al*. (1998).

Simulated external validation. In this case, in contrast to an internal validation, the strains tested against the database are not comprised in the spectral library. A total of 544 strains from 54 species with at least three strains per species from 16 genera were used for this quasi-external validation for proof of principle. The average spectrum of one strain was excluded from the spectral library and this strain was then tested against the remaining database. If the first hit belonged to an average spectrum of a strain from the same species and the spectral distance value D was below 1·5, the result was counted as a correct identification at the species level. If the identification corresponded to a different species from the same genus, the result represented a correct identification at the genus level. In case of a spectral distance value $D > 1{\cdot}5$, the strain was counted as not-identified. A misidentification was noted if a strain was identified incorrectly. The results of the identification of all test strains belonging to the same species was averaged to obtain the mean correctness of identification for each species considered.

RESULTS AND DISCUSSION

Composition of the identification database

The established database consists of IR average spectra from 730 reference strains belonging to 220 different coryneform species and phylogenetically or morphologically related taxa out of 46 genera, 40 of them from the two suborders *Micrococcineae* and *Corynebacterineae* (*Actinomycetales*, *Actinobacteria*) (see the list in IJSEM Online). A total of 192 species (87 %) are represented by type strains. Each average spectrum was calculated from at least three matching single spectra.

Several genera were not included in the database since they did not allow cultivation under the required culture conditions. These are the single-species genera *Frigoribacterium* (Kämpfer *et al*., 2000), *Cryobacterium* (Suzuki *et al*., 1997) and *Renibacterium* (Sanders

& Fryer, 1980) due to their obligate psychrophily. Furthermore, the genus *Bogoriella* (Groth *et al*., 1997) and the genus *Rarobacter* (Yamamoto *et al*., 1988) were excluded for obligate alkaliphily and for requirement of catalase, haemoglobin or haemin for growth in air, respectively. The genera *Tetrasphaera* (Maszenan *et al*., 2000) and *Mycobacterium* Lehmann and Neumann 1896 are not represented due to their extremely slowly growing species.

On the other hand, strains from morphologically similar genera such as *Staphylococcus* were included to enable safe identification results of Gram-positive cocci independent of their G+C content, in particular, since staphylococci are widely distributed in the environment and have frequently been isolated from cheeses (Bockelmann *et al*., 1997; Irlinger *et al*., 1997; Carnio *et al*., 2000).

Not considering species from the above-mentioned excluded genera, there are approximately 210 species from the two suborders *Micrococcineae* and *Corynebacterineae* validly described so far and this number is growing continuously. A total of 180 of these species are included in the database, which corresponds to a coryneform coverage of 85%.

Comparative 16S rDNA sequence analysis

After having included several hundred strains in the database, it became apparent from FT-IR cluster analyses that, sometimes, strains carrying completely different species and/or genus names grouped together at a very high level of spectral similarity. In order to check the identity of these questionable strains, comparative 16S rDNA sequence analysis was carried out for 352 strains. Partial sequences of typical sequence lengths between 800 and 1000 nucleotides were determined and identified using GenBank's BLAST, the RDP and the ARB databases. Generally, there was good agreement between the results obtained from the hit lists of GenBank and the RDP database. As a result, 224 (64%) of these strains were reclassified according to the sequence analysis while the other 128 strains (36%) proved to be correctly classified. Fig. 1(a) shows an example for a number of strains clustering at a very high level of spectral similarity while carrying many different genus and species denotations, some of which have never been validly described (names within quotation marks). After comparative 16S rDNA sequence analysis, each of these strains was identified as *Rhodococcus erythropolis* (Fig. 1b). Comparative sequence analysis and, if necessary, strain reclassification was carried out in order to calibrate the spectral library for subsequent determination of its quality of identification.

According to Stackebrandt & Goebel (1994), homologies of more than 97% strongly point to species identity between query sequence and database sequences, although exceptions to this rule have been reported (e.g. Fox *et al*., 1992). However, a more

Fig. 1. Dendrogram of spectrometrically similar reference strains. Average linkage, correlation with normalized to reproducibility level. Frequency ranges with weights and reproducibility levels: 3000–2800 cm^{-1}/0·8/3·3; 1800–1500 cm^{-1}/0·8/5; 1500–1200 cm^{-1}/0·9/20; 1200–900 cm^{-1}/0·9/33; 900–700 cm^{-1}/0·9/116. Unweighted pair group method algorithm (UPGMA). (a) Classification of reference strains before comparative 16S rDNA sequence analysis. (b) Partial reclassification after sequence analysis.

thorough characterization of the strains including DNA–DNA hybridization and determination of physiological and biochemical properties was not possible due to the vast number of strains dealt with. For this reason, further efforts to differentiate strains reclassified as *Arthrobacter ramosus*/*Arthrobacter pascens* or *Staphylococcus vitulinus*/*Staphylococcus pulvereri* which are indistinguishable by means of comparative 16S rDNA analysis have not been undertaken (see the list in IJSEM Online).

The need to carefully check the identities of reference strains on a regular basis has also been noted by Felske *et al*. (1999) in the course of assessing the heterogeneity of partial 16S rRNA sequences of coryneform strains by TGGE. These authors explained the aberrant behaviour of some strains by the so-called 'human factor', which refers to the handling of strains in laboratories over long periods of time. Moreover, coryneform bacteria have often been misclassified in the past (Altenburger *et al*., 1997). Kollöffel *et al*. (1997) have also reported irregularities after investigating reference strains.

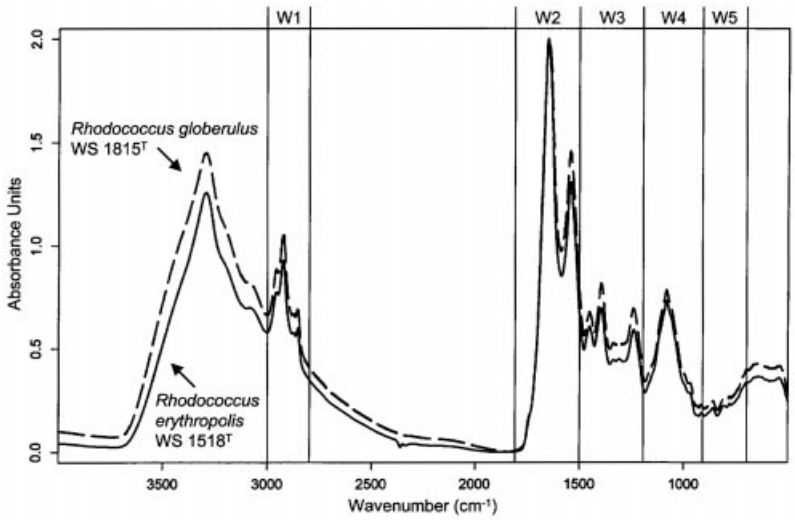

Fig. 2. Normalized averaged absorbance spectra of the type strains from two example species and frequency ranges W1–W5.

Internal validation

After the determination of the so-called reproducibility levels obtained for each spectral window, the different frequency ranges' capacity for correct identification was determined by running identifications of a spectra test set against the complete database separately for each of the five spectral windows. The overall correctness of identification obtained for each frequency range (data not shown) was used for a permutation of different spectral window combinations in order to obtain those yielding highest values for correct identification. The combination comprising all five spectral windows was found to be optimal. In this internal validation, 93·9, 98·1 and 99·5 % of the 208 test strains were correctly identified at the strain, species and genus level, respectively. These values are optimized, meaning that when in four cases a single spectrum was not correctly identified by its corresponding average spectrum, several further single spectra were recorded and included to form a new average spectrum in order to increase the representation of the spectral variety of the same strain in the database. The parameters found to be optimal for the identification of coryneform bacteria by FT-IR spectroscopy are listed in the legend for Fig. 1. Fig. 2 displays the normalized averaged IR spectra of the type strains from two example species together with the five relevant frequency ranges.

In contrast, performing this internal validation with the same strain set but using instead the optimized parameters of the spectral database for yeasts as described by Kümmerle *et al.* (1998) resulted in mean values for correct identification at strain, species and genus level of 90·6, 94·3 and 96·7 %, respectively. The comparison between 'yeast parameters' and 'coryneform parameters' clearly shows that it is worthwhile determining the optimal settings for each group of organisms separately and that adjustments proven to be optimal for one group do not necessarily represent the optimum for another group (see also Helm *et al.*, 1991).

Simulated external validation

The quality of the database was further assessed by carrying out a simulated external validation. One strain was singled out and the library was built up without the average spectrum of this strain. Subsequently, the strain was identified by the spectral library and the correct identification was checked at the species and genus level. A total of 544 strains from 54 species containing a minimum number of three strains per species were used. As a result, 87·3 % of the strains were correctly identified at the species level and 95·4 % at the genus level; 12·7 % did not result in a correct identification at the species level, 1·3 % of which were not identified while 11·4 % were misidentified.

While some species yield 100 % correctness of identification even if being represented only by a low number of strains per species, the general trend shows a correlation between the number of strains per species in the library and the percentage of correct identification obtained (Fig. 3, shading added for illustration reasons only). Roughly, the more strains a species is represented by in the identification database, the more likely is a correct identification at the species level. Likewise, the probability of a misidentification decreases with an increasing number of strains (data not shown). This observation can be explained with an increased coverage of the intraspecies variety in case of a high number of strains per species. The number of non-identified strains, i.e. spectra that were identified with a spectral distance $D > 1{\cdot}5$, equals zero for most of the species under investigation. Seven species out of

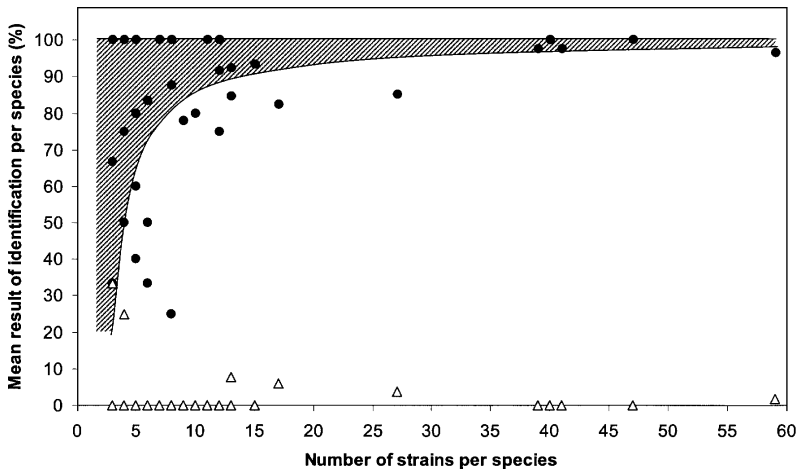

Fig. 3. External validation using 544 strains belonging to 54 species from 16 genera. One datapoint displays the mean result of identification over all strains per considered species *vs* number of strains per species in the spectral library. ●, Correct identification at the species level; △, non-identified (spectral distance $D > 1{\cdot}5$). Shading added for illustration reasons only.

54, however, contained strains that could not be identified. This indicates that the variability range of some species is better represented in the reference library than that of others, independent of the number of strains available in the spectral database. In order to allow the identification of future isolates which might show a higher spectral similarity to these outlier strains than to the other 'core' strains of a particular species, it is sensible to retain the non-identified strains in the database. These isolates' subsequent addition as reference spectra contributes to a better representation of the corresponding species' diversity in the library.

Evaluation of identification quality by comparison of results with those described in the literature

Several publications report on the identification capacity of various methods dealing with different groups of micro-organisms (Table 1). The identification of coryneform bacteria and related genera has been evaluated by using the BIOLOG Identification System (Biolog), API Coryne database (bioMérieux) and the RapID CB Plus system (Remel), three commercially available identification kits based on the analysis of physiological properties. The latter two identification databases contain only a limited number of species, most of which are clinically relevant coryneform bacteria. The percentage values of correct identification achieved in external validations differ greatly between these systems, yielding best results for the API Coryne database (Table 1).

Several studies have analysed the capacity for identification of micro-organisms by FT-IR spectroscopy. Most studies have only performed internal validations where the percentage of correct identification is generally higher than in the case of an external validation. Moreover, the size of the spectral libraries under investigation has been rather limited except for the databases reported by Kümmerle *et al*. (1998) and by this study.

When comparing published identification results, it is therefore necessary to compare the size and composition of the databases evaluated as well as the size and composition of the set of test organisms used for validation. Moreover, the kind of validation performed (internal *vs* external) must be noted carefully since the results of correct identification will largely be influenced by the presence or absence of the tested strains in the library. The effects of the composition and size of both database and test set on the identification quality on the one hand is hardly separable from the effect of the kind of validation performed on the other hand with the limited data available. Generally, an internal validation performed with a small heterogeneous database (i.e. only few strains per genus and genera from different phyla) (Table 1; Helm *et al*., 1991) will yield much higher values for correct identification than an external validation performed on a rather homogeneous database (i.e. closely related taxa) due to a higher percentage of misidentified strains in the latter (e.g. Funke *et al*., 1998). Misidentification occurs if a test strain is less similar to other strains of its own species than to a strain of a different species.

Results are also influenced by the intrataxa separation of the particular group of organisms studied: in case of comparable heterogeneity of databases and similar heterogeneity of the test sets (i.e. numbers of strains per species or genus in the same order of magnitude), a highly different result of correct species allocation in an external validation points to a different intrataxa separability between the organisms analysed, e.g. between yeasts and coryneform bacteria (Kümmerle *et al*., 1998 *vs* this study). Our results seem to indicate that yeast taxa can be considered more distinctive than

Table 1. Literature survey: identification capacity of various methods used for identifying different groups of micro-organisms

Abbreviations: st, strain(s); sp, species; gen, genus/genera; lev, level; non-id, non-identified (FT-IR spectral distance $D > 1{\cdot}5$); misid, misidentified at the species level.

Method	Organisms studied	Composition of identification database (average distribution of st/sp and st/gen)	Validation performed	Composition of test organisms for validation	Result of validation	Reference
BIOLOG Identification System 3.50 (BIOLOG)	Asporogenous, aerobic Gram-positive rods	Approx. 250 sp, 35 genera (7·1 sp/gen), 96 sp and 21 gen of which are high-G+C Gram-positive bacteria (4·6 sp/gen)	External	174 st, 42 sp, 12 gen (4·2 st/sp, 14·5 st/gen)	34 % correct at sp lev, 7 % non-id, 58 % misid	Lindenmann *et al.* (1995); Biolog (2000)
API Coryne database 2.0 (bioMérieux)	Clinically relevant coryneform bacteria and related taxa as well as *Listeria* spp.	Approx. 48 sp, 15 genera (3·2 sp/gen), 44 sp and 13 gen of which are high-G+C Gram-positive bacteria (3·4 sp/gen)	External	407 st, 54 sp, 20 gen (8 st/sp, 20 st/gen)	90·5 % correct at sp lev, 5·6 % non-id, 3·8 % misid	Funke *et al.* (1997b); bioMérieux (2000)
RapID CB Plus system (Remel)	Clinically relevant coryneform bacteria and related taxa as well as *Listeria* spp.	Approx. 52 sp, 14 genera (3·7 sp/gen), 44 sp and 12 gen of which are high-G+C Gram-positive bacteria (3·7 sp/gen)	External	378 st, 49 sp, 16 gen (8 st/sp, 23 st/gen)	80·9 % correct at sp lev, 3·2 % non-id, 15·7 % misid	Funke *et al.* (1998); Remel (2000)
FT-IR	Actinomycetes	46 st, 41 sp, 14 gen (1·1 st/sp, 3·3 st/gen)	Internal	39 st from identification database	89 % correct at st lev, no further information provided	Haag *et al.* (1996)
FT-IR	Low-G+C Gram-positive bacteria (*Staphylococcus* spp., *Clostridium* spp., *Streptococcus* spp.), *γ-Proteobacteria* (*Aeromonas* spp., *Pseudomonas* spp. and *Enterobacteriaceae*)	97 st, 42 sp, 17 gen (2·3 st/sp, 5·7 st/gen)	Internal	72 st from identification database	83·3 % correct at st lev, 94·4 % correct at sp lev, 97·2 % correct at gen lev	Helm *et al.* (1991)
FT-IR	Yeasts	332 st, 74 sp, 18 gen (4·5 st/sp, 18 st/gen)	External	717 st, 36 sp, 11 gen (20 st/sp, 65 st/gen)	97·5 % correct at sp lev, 0·8 % non-id, 1·7 % tentatively identified	Kümmerle *et al.* (1998)
FT-IR	Coryneform bacteria and phylogenetically or morphologically related genera	730 st, 220 sp, 46 gen (3·3 st/sp, 16 st/gen)	Internal	208 st, 208 sp, 41 gen from identification database (1·0 st/sp, 5·1 st/gen)	93·9 % correct at st lev, 98·1 % correct at sp lev, 99·5 % correct at gen lev	This study
			External	544 st, 54 sp, 16 gen (10 st/sp, 34 st/gen	87·3 % correct at sp lev, 95·4 % correct at gen lev, 1·3 % non-id, 11·4 % misid	

coryneform taxa since the percentage of misidentified isolates was found to be much lower. It appears that, in contrast to yeast taxa, coryneform taxa seem to be rather contiguous, exhibiting a certain amount of overlap of different taxonomical characteristics. Taxa merging into each other lacking rather discrete boundaries at the species or genus level will easily be confused, resulting in high percentages of misidentified isolates. This hypothesis, drawn from comparison between identification results of two phylogenetically well-separated kinds of micro-organisms by the same technique (FT-IR spectroscopy), is further supported by the observation that a high percentage of misidentification has also been noted for the identification of coryneform bacteria by physiological properties (cf. Table 1). Reference to the literature reveals that even though a wide spectrum of chemotaxonomical differentiation methods have been applied to the classification of actinobacterial taxa, differentiation down to the species level could not be achieved satisfactorily in all cases, sometimes not even to the genus level. Each of these techniques allows differentiation between some taxa while they fail to separate others (Kämpfer & Kroppenstedt, 1996; Altenburger *et al*., 1997; Busse & Schumann, 1999; Felske *et al*., 1999). Stackebrandt *et al*. (1997) have noted that the rich chemical, morphological and physiological diversity of phylogenetically closely related genera makes the description of families and higher taxa so broad that they become meaningless for the description of the enclosed taxa. Most main actinobacterial lines of descent, described as orders, suborders and families, are not well separated and the statistical significance of branching points is low. Because of the lack of common properties of phylogenetic significance shared by most taxa of a higher taxon, their delineation from each other is somewhat arbitrary and artificial (Stackebrandt & Schumann, 2000). Therefore, even though actinobacterial taxa form a systematically well-derivable group, due to their overlapping (or even contradictory) distribution of taxonomical characters, the possibilities for their differentiation by non-genotypical methods remain somewhat limited.

Examples for a certain degree of interference between different taxonomical characteristics and, as a result, diverging taxonomical opinions can be taken from the on-going discussion as to the separation or unification of different genera such as *Microbacterium*/*Aureobacterium* (Rainey *et al*., 1994; Takeuchi & Hatano, 1998), *Cellulomonas*/*Oerskovia* (Stackebrandt *et al*., 1982; Stackebrandt & Prauser, 1994) or *Corynebacterium*/*Turicella* (Funke *et al*., 1994; Pascual *et al*., 1995; Ruimy *et al*., 1995). These genera are or appear to be phylogenetically intermixed but can be differentiated by means of other taxonomic properties.

Taxonomically overlapping reference species are likely to result in an FT-IR hit list confusion and misidentification of test strains. Thus, an extensive representation of intraspecies and intragenus variety by a substantial number of reference strains is indeed critical for a high percentage value of correct species identification, but will not decrease confusion within a highly contiguous group of organisms. On the other hand, with regard to the behaviour of a single strain, e.g. in epidemiological analyses, a good coverage of intraspecies variety by a large number of reference strains will decrease the probability for differentiation of this particular strain from the others (i.e. yield low values of correct identification at the strain level in general), especially if the strains are very similar, but will yield high values of correct species identification. From the data presented in Fig. 3, we expect that, on average, approximately five to ten different strains per species in a coryneform FT-IR database are appropriate in order to achieve a reasonable identification capacity, depending on the spectral variability within the species. A considerable advantage of working with an FT-IR spectral library for identification is the users' possibility to influence the size and composition of the database themselves, i.e. by continuously adding strains in order to keep the reference library up-to-date or to include own sets of micro-organisms.

CONCLUSION

The FT-IR spectral database is a valuable tool for the rapid, simple and cost-efficient identification of coryneform strains from a variety of sources. Comparison of identification results reported in the literature showed that an overall correct identification of more than 95% at the species level has not been achieved by any non-genetical method so far. Therefore, our results again confirm that coryneform bacteria are a group of organisms whose taxa are not well separated by phenotypical characteristics and, therefore, are difficult to be identified properly. The identification values obtained in this work are well within the range of those having been reported in the literature for identification of these high-G+C Gram-positive bacteria. In order to obtain an identification at the species level with a reasonable probability, each species should be represented by an approximate number of five to ten strains in the database. For improvement of the library's capability for correct identification of coryneform isolates, the number of strains should be increased constantly both for intraspecific representation and for coverage of relevant taxa in general.

ACKNOWLEDGEMENTS

The authors wish to thank Wolfgang Ludwig for providing the ARB software environment together with the ARB 16S rDNA database and for giving valuable information whenever necessary. We are grateful to Lyudmila Evtushenko (VKM, Russia), Ingrid Groth (Hans-Knöll-Institut für Naturstoff-Forschung e.V., Jena, Germany), Klaus P. Schaal (Rheinische Friedrich-Wilhelms-Universität, Bonn, Germany) and Karin Schubert (Ludwig Maximilians Universität, Munich, Germany) for supplying us with strains from their collections and to the companies Chr. Hansen GmbH, Rhodia GmbH, SKW Biosystems GmbH and Wisby GmbH & Co. KG (all Germany) for supplying us

with starter culture strains. Thanks are due to Joachim Charzinski for helping us with the LINUX operating system. The technical assistance of Sven Illgner, Patrizia Hägele and Louise Arnold is greatly appreciated. Special thanks are due to Cornelia Fischer for excellent technical assistance. This work was supported by the FEI (Forschungskreis der Ernährungsindustrie e.V., Bonn), the AiF (Arbeitskreis für industrielle Forschung) and the Ministry of Economics and Technology, project no. 11627N.

REFERENCES

Altenburger, P., Kämpfer, P., Akimov, V. N., Lubitz, W. & Busse, H.-J. (1997). Polyamine distribution in actinomycetes with group B peptidoglycan and species of the genera *Brevibacterium*, *Corynebacterium*, and *Tsukamurella*. *Int J Syst Bacteriol* **47**, 270–277.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402.

Biolog (2000). Biolog GP Data Base. Release 3.50. Hayward, CA: Biolog.

bioMérieux (2000). API Coryne database 2.0. Marcy-l'Étoile: bioMérieux.

Bockelmann, W., Krusch, U., Engel, G., Klijn, N., Smit, G. & Heller, K. J. (1997). The microflora of Tilsit cheese. Part 1. Variability of the smear flora. *Nahrung* **41**, 208–212.

Brosius, J., Palmer, M. L., Kennedy, P. J. & Noller, H. F. (1978). Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. *Proc Natl Acad Sci U S A* **75**, 4801–4805.

Bruker (1996). *Working with the IFS* 28/*B – Sample Preparation*, *Measurement and Evaluation of the FT-IR Spectra of Microorganisms*, Version 2.2. Ettlingen: Bruker.

Busse, H.-J. & Schumann, P. (1999). Polyamine profiles within genera of the class *Actinobacteria* with LL-diaminopimelic acid in the peptidoglycan. *Int J Syst Bacteriol* **49**, 179–184.

Carnio, M. C., Eppert, I. & Scherer, S. (1999). Analysis of the bacterial surface flora of German and French smeared cheeses with respect to their anti-listerial potential. *Int J Food Microbiol* **47**, 89–97.

Carnio, M. C., Höltzel, A., Rudolf, M., Henle, T., Jung, G. & Scherer, S. (2000). The macrocyclic peptide antibiotic Micrococcin P1 is secreted by the food-borne bacterium *Staphylococcus equorum* WS 2733 and inhibits *Listeria monocytogenes* on soft cheese. *Appl Environ Microbiol* **66**, 2378–2384.

Crocker, F. H., Fredrickson, J. K., White, D. C., Ringelberg, D. B. & Balkwill, D. L. (2000). Phylogenetic and physiological diversity of *Arthrobacter* strains isolated from unconsolidated subsurface sediments. *Microbiology* **146**, 1295–1310.

Facius, D., Fartmann, B., Huber, J., Nikoleit, K., Schondelmaier, J. & Steinhauser, S. (1999). *Sequencing Brochure. Instructions for DNA template preparation, primer design and sequencing with the LI-COR DNA Sequencer 4000 and 4200 series*, Version 4. Ebersberg: MWG-BIOTECH.

Felske, A., Vancanneyt, M., Kersters, K. & Akkermans, A. D. L. (1999). Application of temperature-gradient gel electrophoresis in taxonomy of coryneform bacteria. *Int J Syst Bacteriol* **49**, 113–121.

Fox, G. E., Wisotzkey, J. D. & Jurtshuk, P., Jr (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. *Int J Syst Bacteriol* **42**, 166–170.

Funke, G., Stubbs, S., Altwegg, M., Carlotti, A. & Collins, M. D. (1994). *Turicella otitidis* gen. nov., sp. nov., a coryneform bacterium isolated from patients with otitis media. *Int J Syst Bacteriol* **44**, 270–273.

Funke, G., von Graevenitz, A., Clarridge, J. E., III & Bernard, K. A. (1997a). Clinical microbiology of coryneform bacteria. *Clin Microbiol Rev* **10**, 125–159.

Funke, G., Renaud, F. N., Freney, J. & Riegel, P. (1997b). Multicenter evaluation of the updated and extended API (RAPID) Coryne database 2.0. *J Clin Microbiol* **35**, 3122–3126.

Funke, G., Peters, K. & Aravena-Roman, M. (1998). Evaluation of the RapID CB Plus system for identification of coryneform bacteria and *Listeria* spp. *J Clin Microbiol* **36**, 2439–2442.

Goodacre, R., Timmins, É. M., Burton, R., Kaderbhai, N., Woodward, A. M., Kell, D. B. & Rooney, P. J. (1998). Rapid identification of urinary tract infection bacteria using hyperspectral whole-organism fingerprinting and artificial neural networks. *Microbiology* **144**, 1157–1170.

Groth, I., Schumann, P., Rainey, F. A., Martin, K., Schuetze, B. & Augsten, K. (1997). *Bogoriella caseilytica* gen. nov., sp. nov., a new alkaliphilic actinomycete from a soda lake in Africa. *Int J Syst Bacteriol* **47**, 788–794.

Haag, H., Gremlich, H.-U., Bergmann, R. & Sanglier, J.-J. (1996). Characterization and identification of actinomycetes by FT-IR spectroscopy. *J Microbiol Methods* **27**, 157–163.

Helm, D., Labischinski, H. & Naumann, D. (1991). Elaboration of a procedure for identification of bacteria using Fourier-transform IR spectral libraries: a stepwise correlation approach. *J Microbiol Methods* **14**, 127–142.

Holt, C., Hirst, D., Sutherland, A. & MacDonald, F. (1995). Discrimination of species in the genus *Listeria* by Fourier transform infrared spectroscopy and canonical variate analysis. *Appl Environ Microbiol* **61**, 377–378.

Holt, J. G., Krieg, N. R., Sneath, P. H. A., Staley, J. T. & Williams, S. T. (editors) (1994). *Bergey's Manual of Determinative Bacteriology*, 9th edn. Baltimore: Williams & Wilkins.

Irlinger, F., Morvan, A., El Solh, N. & Bergere, J. L. (1997). Taxonomic characterization of coagulase-negative staphylococci in ripening flora from traditional French cheeses. *Syst Appl Microbiol* **20**, 319–328.

Junge, K., Gosink, J. J., Hoppe, H.-G. & Staley, J. T. (1998). *Arthrobacter*, *Brachybacterium* and *Planococcus* isolates identified from antarctic sea ice brine. Description of *Planococcus mcmeekinii*, sp. nov. *Syst Appl Microbiol* **21**, 306–314.

Kämpfer, P. & Kroppenstedt, R. M. (1996). Numerical analysis of fatty acid patterns of coryneform bacteria and related taxa. *Can J Microbiol* **42**, 989–1005.

Kämpfer, P., Seiler, H. & Dott, W. (1993). Numerical classification of coryneform bacteria and related taxa. *J Gen Appl Microbiol* **39**, 135–214.

Kämpfer, P., Rainey, F. A., Andersson, M. A., Nurmiaho Lassila, E.-L., Ulrych, U., Busse, H.-J., Weiss, N., Mikkola, R. & Salkinoja-Salonen, M. (2000). *Frigoribacterium faeni* gen. nov., sp. nov., a novel psychrophilic genus of the family *Microbacteriaceae*. *Int J Syst Evol Microbiol* **50**, 355–363.

Klatte, S. (1995). *Polyphasische automatisierte Identifizierung der Gattungen Corynebacterium Lehmann & Neumann 1896 und Rhodococcus Zopf 1891*. PhD thesis, University of Osnabrück.

Kollöffel, B., Burri, S., Meile, L. & Teuber, M. (1997). Development of 16S-rRNA oligonucleotide probes for *Brevibacterium*, *Micrococcus*/*Arthrobacter* and *Microbacterium*/*Aureobacterium* used in dairy starter cultures. *Syst Appl Microbiol* **20**, 409–417.

Kollöffel, B., Meile, L. & Teuber, M. (1999). Analysis of brevibacteria on the surface of Gruyère cheese detected by in situ hybridization and by colony hybridization. *Lett Appl Microbiol* **29**, 317–322.

Kümmerle, M., Scherer, S. & Seiler, H. (1998). Rapid and reliable identification of fermentative yeasts by Fourier-transform infrared spectroscopy. *Appl Environ Microbiol* **64**, 2207–2214.

Lechner, S., Mayr, R., Francis, K. P., Prüß, B. M., Kaplan, T., Wießner-Gunkel, E., Stewart, G. S. A. B. & Scherer, S. (1998). *Bacillus weihenstephanensis* sp. nov., is a new psychrotolerant species of the *Bacillus cereus* group. *Int J Syst Bacteriol* **48**, 1373–1382.

Lednická, D., Mergaert, J., Cnockaert, M. C. & Swings, J. (2000). Isolation and identification of cellulolytic bacteria involved in the degradation of natural cellulosic fibres. *Syst Appl Microbiol* **23**, 292–299.

Lindenmann, K., von Graevenitz, A. & Funke, G. (1995). Evaluation of the Biolog system for the identification of asporogenous, aerobic gram-positive rods. *Med Microbiol Lett* **4**, 287–296.

Ludwig, W., Strunk, O., Klugbauer, S., Klugbauer, N., Weizenegger, M., Neumaier, J. & Bachleitner, M. (1998). Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* **19**, 554–568.

Maidak, B. L., Cole, J. R., Lilburn, T. G. & 9 other authors (2000). The RDP (Ribosomal Database Project) continues. *Nucleic Acids Res* **28**, 173–174.

Maszenan, A. M., Seviour, R. J., Patel, B. K. C., Schumann, P., Burghardt, J., Tokiwa, Y. & Stratton, H. M. (2000). Three isolates of novel polyphosphate-accumulating Gram-positive cocci, obtained from activated sludge, belong to a new genus, *Tetrasphaera* gen. nov., and description of two new species, *Tetrasphaera japonica* sp. nov., and *Tetrasphaera australiensis* sp. nov. *Int J Syst Evol Microbiol* **50**, 593–603.

Naumann, D., Helm, D. & Schultz, C. (1994). Characterization and identification of micro-organisms by FT-IR spectroscopy and FT-IR microscopy. In *Bacterial Diversity and Systematics*, pp. 67–85. Edited by F. G. Priest, A. Ramos Cormenzana & B. J. Tindall. New York: Plenum.

Pascual, C., Lawson, P. A., Farrow, J. A. E., Navarro Gimenez, M. & Collins, M. D. (1995). Phylogenetic analysis of the genus *Corynebacterium* based on 16S rRNA gene sequences. *Int J Syst Bacteriol* **45**, 724–728.

Rainey, F., Weiss, N., Prauser, H. & Stackebrandt, E. (1994). Further evidence for the phylogenetic coherence of actinomycetes with group B-peptidoglycan and evidence for the phylogenetic intermixing of the genera *Microbacterium* and *Aureobacterium* as determined by 16S rDNA analysis. *FEMS Microbiol Lett* **118**, 135–140.

Remel (2000). RapID CB Plus System. Lenexa, KS: Remel.

Ruimy, R., Riegel, P., Boiron, P., Monteil, H. & Christen, R. (1995). Phylogeny of the genus *Corynebacterium* deduced from analyses of small-subunit ribosomal DNA sequences. *Int J Syst Bacteriol* **45**, 740–746.

Sanders, J. E. & Fryer, J. L. (1980). *Renibacterium salmoninarum* gen. nov., sp. nov., the causative agent of bacterial kidney disease in salmonid fishes. *Int J Syst Bacteriol* **30**, 496–502.

Schubert, K., Ludwig, W., Springer, N., Kroppenstedt, R. M., Accolas, J.-P. & Fiedler, F. (1996). Two coryneform bacteria isolated from the surface of French Gruyère and Beaufort cheeses are new species of the genus *Brachybacterium*: *Brachybacterium alimentarium* sp. nov. and *Brachybacterium tyrofermentans* sp. nov. *Int J Syst Bacteriol* **46**, 81–87.

Seiler, H. (1986). Identification of cheese-smear coryneform bacteria. *J Dairy Res* **53**, 439–449.

Stackebrandt, E. & Goebel, B. M. (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**, 846–849.

Stackebrandt, E. & Prauser, H. (1994). The family *Cellulomonadaceae*. In *The Prokaryotes*, pp. 1323–1345. Edited by A. Balows, H. G. Trüper, M. Dworkin, W. Harder & K. H. Schleifer. New York: Springer.

Stackebrandt, E. & Schumann, P. (2000). Description of *Bogoriellaceae* fam. nov., *Dermacoccaceae* fam. nov., *Rarobacteraceae* fam. nov. and *Sanguibacteraceae* fam. nov. and emendation of some families of the suborder *Micrococcineae*. *Int J Syst Evol Microbiol* **50**, 1279–1285.

Stackebrandt, E., Seiler, H. & Schleifer, K. H. (1982). Union of the genera *Cellulomonas* Bergey *et al*. and *Oerskovia* Prauser *et al*. in a redefined genus *Cellulomonas*. *Zentbl Bakteriol Hyg I Abt Orig* **C3**, 401–409.

Stackebrandt, E., Rainey, F. A. & Ward-Rainey, N. L. (1997). Proposal for a new hierarchic classification system, *Actinobacteria* classis nov. *Int J Syst Bacteriol* **47**, 479–491.

von Stetten, F., Francis, K. P., Lechner, S., Neuhaus, K. & Scherer, S. (1998). Rapid discrimination of psychrotolerant and mesophilic strains of the *Bacillus cereus* group by PCR targeting of 16S rDNA. *J Microbiol Methods* **34**, 99–106.

Suzuki, K., Sasaki, J., Uramoto, M., Nakase, T. & Komagata, K. (1997). *Cryobacterium psychrophilum* gen. nov., sp. nov., nom. rev., comb. nov., an obligately psychrophilic actinomycete to accommodate "*Curtobacterium psychrophilum*" Inoue and Komagata 1976. *Int J Syst Bacteriol* **47**, 474–478.

Takeuchi, M. & Hatano, K. (1998). Union of the genera *Microbacterium* Orla-Jensen and *Aureobacterium* Collins *et al*. in a redefined genus *Microbacterium*. *Int J Syst Bacteriol* **48**, 739–747.

Timmins, É. M., Howell, S. A., Alsberg, B. K., Noble, W. C. & Goodacre, R. (1998). Rapid differentiation of closely related *Candida* species and strains by pyrolysis-mass spectrometry and Fourier transform-infrared spectroscopy. *J Clin Microbiol* **36**, 367–374.

Tindall, B. J., Brambilla, E., Steffen, M., Neumann, R., Pukall, R., Kroppenstedt, R. M. & Stackebrandt, E. (2000). Cultivatable microbial diversity: gnawing at the Gordian knot. *Environ Microbiol* **2**, 310–318.

Valdés-Stauber, N., Scherer, S. & Seiler, H. (1997). Identification of yeasts and coryneform bacteria from the surface microflora of brick cheeses. *Int J Food Microbiol* **34**, 115–129.

Yamamoto, N., Sato, S., Saito, K., Hasuo, T., Tadenuma, M., Suzuki, K., Tamaoka, J. & Komagata, K. (1988). *Rarobacter faecitabidus* gen. nov., sp. nov., a yeast-lysing coryneform bacterium. *Int J Syst Bacteriol* **38**, 7–11.