

Research article

Open Access

## RIDOM: Comprehensive and public sequence database for identification of *Mycobacterium* species

Dag Harmsen\*<sup>1</sup>, Stefan Dostal<sup>2</sup>, Andreas Roth<sup>3</sup>, Stefan Niemann<sup>4</sup>, Jörg Rothgänger<sup>5</sup>, Michael Sammeth<sup>6</sup>, Jürgen Albert<sup>6</sup>, Matthias Frosch<sup>2</sup> and Elvira Richter<sup>4</sup>

Address: <sup>1</sup>Institut für Hygiene, Universität Münster, Münster, Germany, <sup>2</sup>Institut für Hygiene und Mikrobiologie, Universität Würzburg, Würzburg, Germany, <sup>3</sup>Institut für Mikrobiologie und Immunologie, Lungenklinik Heckeshorn, Berlin, Germany, <sup>4</sup>Nationales Referenzzentrum für Mykobakterien, Forschungszentrum Borstel, Borstel, Germany, <sup>5</sup>Ridom GmbH, Würzburg, Germany and <sup>6</sup>Lehrstuhl für Informatik II, Universität Würzburg, Würzburg, Germany

Email: Dag Harmsen\* - dharmsen@uni-muenster.de; Stefan Dostal - sdostal@web.de; Andreas Roth - mikromau@zedat.fu-berlin.de; Stefan Niemann - sniemann@fz-borstel.de; Jörg Rothgänger - jrothgaenger@ridom.de; Michael Sammeth - micha@sammeth.net; Jürgen Albert - albert@informatik.uni-wuerzburg.de; Matthias Frosch - mfrosch@hygiene.uni-wuerzburg.de; Elvira Richter - erichter@fz-borstel.de

\* Corresponding author

Published: 11 November 2003

Received: 19 August 2003

BMC Infectious Diseases 2003, 3:26

Accepted: 11 November 2003

This article is available from: <http://www.biomedcentral.com/1471-2334/3/26>

© 2003 Harmsen et al; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** Molecular identification of *Mycobacterium* species has two primary advantages when compared to phenotypic identification: rapid turn-around time and improved accuracy. The information content of the 5' end of the 16S ribosomal RNA gene (16S rDNA) is sufficient for identification of most bacterial species. However, reliable sequence-based identification is hampered by many faulty and some missing sequence entries in publicly accessible databases.

**Methods:** In order to establish an improved 16S rDNA sequence database for the identification of clinical and environmental isolates, we sequenced both strands of the 5' end of 16S rDNA (*Escherichia coli* positions 54 to 510) from 199 mycobacterial culture collection isolates. All validly described species (n = 89; up to March 21, 2000) and nearly all published sequevar variants were included. If the 16S rDNA sequences were not discriminatory, the internal transcribed spacer (ITS) region sequences (n = 84) were also determined.

**Results:** Using 5'-16S rDNA sequencing a total of 64 different mycobacterial species (71.9%) could be identified. With the additional input of the ITS sequence, a further 16 species or subspecies could be differentiated. Only *Mycobacterium tuberculosis* complex species, *M. marinum* / *M. ulcerans* and the *M. avium* subspecies could not be differentiated using 5'-16S rDNA or ITS sequencing. A total of 77 culture collection strain sequences, exhibiting an overlap of at least 80% and identical by strain number to the isolates used in this study, were found in the GenBank. Comparing these with our sequences revealed that an average of 4.31 nucleotide differences (SD ± 0.57) were present.

**Conclusions:** The data from this analysis show that it is possible to differentiate most mycobacterial species by sequence analysis of partial 16S rDNA. The high-quality sequences reported here, together with ancillary information (e.g., taxonomic, medical), are available in a public database, which is currently being expanded in the RIDOM project (<http://www.ridom-rdna.de>), for similarity searches.

## Background

Reliable microbial identification using conventional methods often requires several techniques, such as the use of colony morphology, gram staining, determination of nutritional requirements and/or biochemical reactions. Identification of mycobacteria at the species level using conventional biochemical tests is laborious with a long turn-around time, leading to significant delays in diagnosis and ambiguous results occur frequently. Other methods based on lipid analysis, such as high-performance liquid chromatography, thin-layer chromatography and gas-liquid chromatography are used only in a few clinical laboratories [1–3]. Identification using molecular techniques, on the other hand, provides two primary advantages when compared to phenotypic identification: a more rapid turn-around time and improved accuracy in identification [4–6]. With assays based on molecular techniques, the genetic targets vary, as does the method of target characterization. Three targets that have proven useful in identification are the 16S ribosomal RNA (16S rDNA) gene [7–10], the internal transcribed spacer (ITS) region [11] and the *hsp65* gene [12–14]. The main advantage of 16S rRNA gene analysis is that it can be applied in the identification of all bacteria, even those which are dead or are uncultivable [15,16]. The ITS region has a greater discriminatory power than the 16S rDNA, but does not allow the recognition and the reliable phylogenetic placement of species not previously described. The most common method of target characterization is amplification, followed by either probe hybridization, restriction fragment length polymorphism analysis, or sequencing. Although sequence analysis requires more specialized equipment than the other methods, this technology is becoming less expensive and provides the highest level of resolution and portability. Sequencing of the 16S RNA gene is therefore regarded as the most suitable method for identification of mycobacteria in the clinical laboratory setting [5,6,17].

Existing sequence databases and analytical tools (e.g., the National Center for Biotechnology Information [NCBI] GenBank and the Ribosomal Database Project [RDP] [18,19]) are not optimal for accurate identification of clinically relevant microorganisms. The deficiencies in the contents of these databases include the presence of ragged sequence ends (resulting in wrong 'best' matches in similarity searches), faulty sequence entries (due to error-prone sequencing techniques used earlier, e.g., reverse transcriptase sequencing), absence of quality control of sequence entries, noncharacterized entries, outdated nomenclature, and the lack of type strains pertaining to many clinically important microorganisms. Furthermore, search results are not presented in a user-friendly manner. Our ribosomal differentiation of microorganisms (RIDOM) project attempts to overcome these problems [20–22].

Sequencing the entire 16S rDNA is not a practical method for routine identification. However, the information content of the 5' end of the gene is sufficient for specific identification of most *Mycobacterium* species (i.e., only one sequencing run) [7]. Therefore, in order to establish an improved 16S rDNA reference sequence database for the identification of clinical isolates, we sequenced both strands of the 5'-16S rDNA (*Escherichia coli* position 54–510) from 199 mycobacterial isolates. All validly described species (n = 89; up to March 21, 2000) and nearly all published sequevar variants were included. If the 16S rDNA sequences were not discriminatory (i.e., different and unique), the ITS region sequences were also determined. The ultimate goal of this study was to come up with an algorithm for genetic differentiation of all mycobacteria, using insertion element and *gyrB* PCRs in addition to rRNA operon sequencing, when the latter target was not discriminatory enough.

(This study was presented in part at the 101st General Meeting of the American Society for Microbiology, Orlando, Florida, 20 to 24 May 2001.)

## Methods

### Bacterial strains and growth conditions

The strains investigated in this study are listed in Table 1 (see Additional file Table 1). Culture collection isolates, including the type strains, were used in this analysis when available. Most strains were cultivated on Löwenstein-Jensen media at 28°C and 37°C. *Mycobacterium haemophilum* was cultivated on Löwenstein-Jensen media with factor X strips (Becton Dickinson, Heidelberg, Germany), whereas *M. avium* subsp. *paratuberculosis* was cultured on Middlebrook-Cohn 7H10 agar with OADC and mycobactin enrichment. *M. genavense* was grown in broth media (BACTEC 13A medium, Becton Dickinson). For some isolates, only DNA and no culture was available (see Table 1 footnotes, e.g., *M. leprae* or *M. lepraemurium*). All isolates with missing sequence entries in public databases or with sequence discrepancies detected by GenBank BLAST searches were additionally identified using extensive conventional biochemical methods [1]. At least two different culture collection strains from these species were included in this study.

### In vitro amplification and DNA sequencing of the 16S ribosomal RNA genes and its region

A loopful of bacterial cells for extraction of DNA was washed with distilled water and incubated in 200 µl TE buffer (Tris-HCl, 10 mM; EDTA, 1 mM; pH 7.0) for 30 min at 80°C. The DNA was extracted with N-cetyl-N-, N, N-trimethylammoniumbromide (CTAB)/NaCl according to the protocol of van Embden et al. [23]. The final DNA precipitate was suspended in 200 µl TE buffer and stored frozen (-20°C) until PCR was performed. Two microliters

of this suspension (approximately 10 ng of DNA) were used for PCR amplifications. PCR was performed in a total volume of 50  $\mu$ l containing 200  $\mu$ M deoxynucleoside triphosphates (dATP, dCTP, dGTP, and dTTP), 10 pmol of each primer, 5  $\mu$ l of 10-fold concentrated PCR buffer (100 mM Tris-HCl; 500 mM KCl; 15 mM MgCl<sub>2</sub>; pH 8.3), and 1 U of AmpliTaq DNA polymerase (Applied Biosystems, Weiterstadt, Germany). Thermal cycling reactions consisted of an initial denaturation (80°C, 5 min) followed by 28 cycles of denaturation (94°C, 45 s), annealing (53°C for both 16S rDNA- and ITS-PCR, 1 min), and extension (72°C, 90 s), with a single final extension (72°C, 10 min). The broad-range primers 16S-27f (5'-AGA GTT TGA TCM TGG CTC AG -3') and 16S-907r (5'-CCG TCA ATT CMT TTR AGT TT -3') were used for 16S ribosomal DNA PCR. The universal primers 16S-1511f (5'-AAG TCG TAA CAA GGT ARC CG -3') and 23S-23r (5'-TCG CCA AGG CAT CCA CC -3') were used for amplification of the ITS region. Identical or near-identical primer binding sites have already been described by Lane [24]. Reactions took place in a dedicated automated DNA thermal cycler (GeneAmp 2400, Applied Biosystems). In order to control for the presence of contaminating nucleic acids, controls containing water in place of template DNA, were run in parallel in each run. The amplicons were sequenced using the BigDye Terminator V2.0 Ready Reaction Cycle Sequencing Kit (Applied Biosystems). The sequencing reaction required 2  $\mu$ l of Premix, 5 pmol of sequencing primer and 0.2  $\mu$ g of the PCR product template in a total volume of 10  $\mu$ l. For sequencing 16S rDNA either the primer 16S-27f or 16S-519r (5'-GWA TTA CCG CGG CKG CTG -3') were used both with annealing temperature of 53°C. For sequencing ITS either the primer 16S-1511f or 23S-23r was employed, with an annealing temperatures of 55°C and 51°C, respectively. All sequencing reactions were performed using the GeneAmp 2400 system with 25 cycles of denaturation (96°C, 10 s), annealing (temperature depending on the sequencing primer used, 5 s), and extension (60°C, 4 min). The sequencing products were purified using the recommended Centri-Sep Spin Columns (Princeton Separations, Adelphia, NJ), followed by preparation for running onto the ABI Prism 377 or 310 Genetic Analyzer, in accordance with the instructions of the manufacturer (Applied Biosystems). The nucleotide sequences for both DNA strands were determined. Ambiguities were resequenced and at least 98% percent of the complete double-stranded sequences of the 16S rDNA and ITS targets were obtained.

#### Subcloning of PCR products

*M. celatum* isolates exhibit 16S rDNA interoperon variability. Furthermore, several fast growing mycobacteria contain ITS operons which differ in length and/or base composition (Table 1, explanatory footnotes). Direct

sequencing of PCR products of these isolates was therefore not possible. PCR products of these strains were separated on an agarose gel and the first band, larger than 200 bp, was cut and cleaned with the Jetsorb Gel Extraction kit (Genomed, Bad Oeynhausen, Germany). The cleaned DNA was subcloned in a plasmid vector with the TOPO TA Cloning kit (Invitrogen, Carlsbad, CA) according to the recommendations of the manufacturer. Transformed *Escherichia coli* strains were cultured and crude DNA extractions were performed by heating and centrifugation. PCRs with M13 primers were run with an aliquot of the supernatant and the PCR products of three subclones each were sequenced as stated above.

#### Analysis of the ribosomal DNA sequences and statistical analysis

The sequencing output from the ABI Prism Genetic Analysers was analysed using the Sequence Navigator version 1.0.1 computer software (Applied Biosystems). The region from base positions 54 to 510 (corresponding to *E. coli* 16S rDNA positions) for the 16S rDNA and the complete ITS were further analysed. Sequences from primer regions were not included in this analysis. The MegAlign (version 3.11) component of the Lasergene program (DNASTAR Inc., Madison, WI) was used for multiple alignment and phylogenetic analysis. Multiple sequence alignments were determined using the CLUSTAL W algorithm. Spearman's rank correlation test, a non-parametric measure of the degree of association between two numerical values, was used to access the correlation between the means of base differences (i.e., differences between GenBank and RIDOM sequence data) stratified by years and the GenBank submission date. The StatView version 5.0 statistical software package was used to calculate the Spearman's rank correlation ( $r_s$ ) and the significance of association (SAS Institute Inc., Cary, NC).

#### RIDOM implementation

We have recently changed the uniform resource locator (URL) of our RIDOM service (from <http://www.ridom.de> to <http://www.ridom-rdna.de>) and substantially improved the implementation. Main backend components of RIDOM include the PHRED/PHRAP, FASTA and CLUSTAL W programs that are embedded into Java Servlets [25–27]. In order to view the sequence chromatograms in the new "Trace Editor", client computers need to have a recent version of a standard WWW browser (Netscape or IE version 4 or higher) and Sun's Java Plug-in (1.2 or higher) installed.

#### Results

A total of 199 partial 16S rDNA (corresponding to *E. coli* positions 54 to 510) and 84 complete ITS sequences from mycobacterial culture isolates were newly determined, for the purpose of building up a high-quality reference

sequence database. All validly described species and subspecies ( $n = 89$ ; up to March 21, 2000) were included. In this study a valid publication of a new name or new nomenclature combination refers to publications appearing in the International Journal of Systematic Bacteriology (IJSB) / International Journal of Systematic and Evolutionary Microbiology (IJSEM, from January 2000), either as an original article or in the Validation Lists regularly appearing in this journal. The Validation Lists constitute valid publication of new names and new combinations that meet validation criteria and which have been previously published in journals other than IJSB and IJSEM. Names not considered validly published should no longer be used or should be used in quotation marks (e.g. "Mycobacterium album") to indicate that the name has not been validly published. One hundred sixty of the 199 isolates sequenced were obtained from culture collections. The remaining 39 strains were obtained for sequencing from private collections (Table 1, footnotes). Additionally, fifteen 16S rDNA and 19 ITS GenBank entries were included in the subsequent analysis (Tables 1, footnotes). The 16S rDNA from many *Mycobacterium* species had been previously published. In contrast, the ITS sequence was generated from several species for the first time.

#### **Differentiation of *Mycobacterium* species based on rRNA operon sequencing**

A 16S rDNA phylogenetic tree was created that included one representative strain of each sequence variant. Species having identical sequences are shown in bold (Figure 1). According to 5'-16S rDNA sequencing, 64 different mycobacterial species (71.9%) could be identified. With the additional input of the ITS sequence, a further 16 species or subspecies could be resolved. The groups that shared identical partial 16S rDNA and which could be discriminated with the aid of their ITS sequences were: (i) *Mycobacterium abscessus* and *M. chelonae* sequevar I; (ii) *M. gastri* and *M. kansasii* sequevars I & IV; (iii) *M. fortuitum* 3<sup>rd</sup> biovariant (sorbitol +, sequevar II), *M. farcinogenes* and *M. senegalense*; (iv) *M. fortuitum* 3<sup>rd</sup> biovariant (sorbitol -, sequevar III) and *M. porcinum*; (v) *M. fortuitum* subsp. *acetamidolyticum* and *M. fortuitum* subsp. *fortuitum* sequevar I; (vi) *M. peregrinum* and *M. septicum*; (vii) *M. murale* and *M. tokaiense*; and (viii) *M. flavescens* sequevar II and *M. novocastrense*. Only the four *Mycobacterium tuberculosis* complex species, *M. marinum* / *M. ulcerans* (Mul A) and the three *M. avium* subspecies could not be differentiated using 5'-16S rDNA or ITS sequencing.

#### **16S rRNA gene variability of *Mycobacterium* species**

Intraspecies rRNA gene heterogeneity was encountered in the case of some mycobacterial species. Sequevar (sqv.) designations were then chosen according to the nomenclature of Frothingham and Wilson [28]. ITS variants were

labelled with a species name acronym and an Arabic capital letter (e.g., Mka A for *M. kansasii* ITS sequevar variant A). 16S rDNA sequevars were designated with Roman numerals (e.g., *M. chelonae* sqv. I for *M. chelonae* 5'-16S rDNA variant I). The 16S rDNA sequevar designations for *M. kansasii* are somewhat inconsistent (i.e., *M. kansasii* sqv. I and sqv. IV as well as *M. kansasii* sqv. III and IV-2 have identical 5'-16S rDNA sequences) because the sequence variants were initially determined by *hsp65* analysis [11,13,29]. The following 16S rDNA sequence variants were observed: (i) *M. avium* sqv. I-II, (ii) *M. chelonae* sqv. I-II, (iii) *M. flavescens* sqv. I-II, (iv) *M. fortuitum* sqv. I-V, (v) *M. gordonae* sqv. I-V, (vi) *M. intracellulare* sqv. I-V, (vii) *M. kansasii* sqv. I & IV, II, III & VI-2, V, VI-1 and VI-3, (viii) *M. lentiflavum* sqv. I-II, (ix) *M. parafortuitum* sqv. I-II, (x) *M. simiae* sqv. I-II, (xi) *M. terrae* sqv. I-III, and (xii) *M. xenopi* sqv. I-III.

#### **ITS micro-heterogeneity of the genus *Mycobacterium***

A total of 84 complete ITS sequences from mycobacterial culture isolates were newly determined. We were not able to obtain isolates of all published ITS sequence variants. Therefore, for some *M. avium*, *M. intracellulare*, *M. kansasii* and *M. simiae* variants our ITS analysis relied on a few recently submitted GenBank entries (Table 1, explanatory footnote). Furthermore, the ITS region was not studied in the same detail as the 5'-16S rDNA. Nevertheless several new sequence variants were observed. The following is a detailed listing of the results: (i) *M. avium* Mav A-E, (ii) *M. chelonae* Mche A-C, (iii) *M. flavescens* Mfla A-B, (iv) *M. fortuitum* Mfor A-D, (v) *M. gordonae* Mgo A-E, (vi) *M. intracellulare* MAC A-F, MAC H-L and Min A-D, (vii) *M. kansasii* Mka A-F, (viii) *M. peregrinum* Mpe A-C, (ix) *M. phlei* Mphle-A-B, (x) *M. scrofulaceum* Mscro A-B, (xi) *M. simiae* Msi A-E, (xii) *M. ulcerans* Mul A-B, and (xiii) *M. xenopi* Mxe A-C.

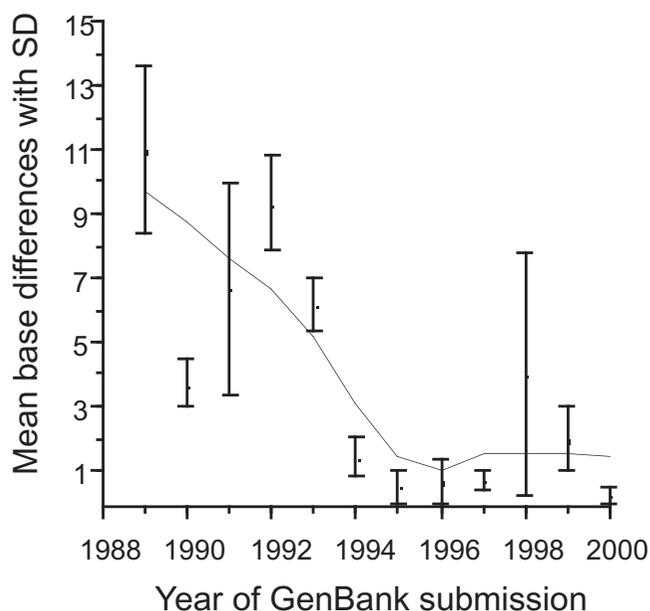
#### **Comparison of RIDOM and GenBank mycobacterial 16S rDNA sequences**

Performing a similarity search with RIDOM sequences against GenBank, we found sequences of 77 identical culture collection strains with a minimum overlap of 80%. Comparing these entries in detail with our sequences, we found an average of 4.31 nucleotide differences (SD  $\pm$  0.57). Using the Spearman's rank correlation test a significant negative correlation between the means of base differences stratified by years, and the submission date was also found (with  $r_s$  -0.56 and  $p < 0.0001$ ; Figure 2). Furthermore, seven out of the 160 sequenced culture collection strains turned out to be "wrong" (4.4%), i.e., differed excessively from published sequence and phenotypic data. These isolates were omitted from further analysis (Table 1, footnotes).



**Figure 1**

5'-16S rDNA based phylogenetic tree of the genus *Mycobacterium*, including one representative strain of each sequence variant (sequevar = sqv.). T designates the type strain of this species. Multiple sequence alignments were determined using the CLUSTAL W algorithm in the MegAlign component of the Lasergene program. The tree was rooted using *Corynebacterium pseudodiphtheriticum* as the outgroup sequence. Sequences were all determined in our laboratory unless indicated by GenBank accession number. Species having identical sequences are shown in bold. The numbers on the abscissa represent the percent distance between different isolates.



**Figure 2**  
5'-16S rDNA comparison of GenBank submissions (n = 77, 1989–2000) and RIDOM sequences with a minimum overlap of 80%. On the ordinate are the means of base differences and their standard deviation, stratified by years with respect to the GenBank submission date (abscissa) shown.

## Discussion

Differentiation of *Mycobacterium* species has traditionally relied upon biochemical test profiles of pure cultures, methodologies that require skilled microbiology technicians and time periods of 2 to 6 weeks before results can be reported. For this reason alone, molecular identification of mycobacteria is likely to become the standard employed. The rRNA gene is an attractive target for the genotypic identification as it contains information suitable for the identification of mycobacteria at the species level as well as for the rapid recognition of previously undescribed species [16,30]. Commercial probe assays targeted against the rRNA operon are already available, but these assays only test for one or a few species at a time [31,32]. Until now, molecular approaches, which can be applied universally for the identification of *Mycobacterium* isolates, are hampered because of the many faulty and sometimes missing sequence entries in publicly accessible databases. Assuming that we have produced correct sequences – all of our tested sequences were confirmed to be 100% similar to the independently determined sequences of Turenne et al. [33]–, this is clearly shown by a comparison of our newly determined sequences with the GenBank sequence entries from identical culture collection strains previously deposited (Figure 2). An error-

rate of approximately 1% is more than can be tolerated in medical species identification and may lead to wrong or confusing results. On the other hand, there has been a marked improvement in sequence quality since 1994. This is most probably due to changes in sequencing techniques (*Taq*-cycle sequencing and automated sequencers). However, more than 57% of all mycobacterial sequences examined were deposited before this date.

Marked microheterogeneity within species, sometimes hindering a straightforward species differentiation, was observed during this study. Intraspecies rRNA gene variability in mycobacteria has already been independently reported for several species; i.e., (i) *M. avium-intracellulare* [11,28,34–37], (ii) *M. fortuitum* complex [38], (iii) *M. goodii* [39], (iv) *M. kansasii* [29,40,41], (v) *M. lentiflavum* [42], (vi) *M. scrofulaceum* [11], (vii) *M. simiae* [43], (viii) *M. terrae* [44], (ix) *M. ulcerans* [45], and (x) *M. xenopi* [43]. However, the present study constitutes the most complete analysis with respect to mycobacterial microheterogeneity. A document with the multiple alignments of the partial 16S rDNA and ITS sequence variants has been deposited to serve as a reference in the future (see Additional file Figure 3). Where known, the cross-references of the 16S rDNA and ITS sequences are also stated in the document.

Neither 16S rDNA nor ITS sequencing could differentiate all mycobacterial species. Of course it is possible to discriminate indistinguishable species by key-phenotypic reactions; for example the closely related species *M. kansasii* sqv. I & IV and *M. gastri* have an identical 5'-16S rDNA sequence, yet the simple addition of a pigmentation criterion results in a specific test for these two taxa. Nevertheless, we tried to incorporate other molecular targets, which have quite recently become partially available, in a molecular and universal mycobacterial identification scheme. This algorithm for genetic differentiation of all mycobacteria employs insertion element and *gyrB* PCR in addition to rRNA operon sequencing. This differentiation scheme has been also deposited (see Additional file Figure 4). This file includes primer sequences, experimentally verified *gyrB*- and IS-PCR conditions as well as references to the various methods [11,46–52]. Acid-fast bacteria are grouped according to this algorithm by a *M. tuberculosis* complex-specific *gyrB* PCR in either the *M. tuberculosis* complex or non tuberculous mycobacteria (NTM) group [46]. All members of the *M. tuberculosis* complex are further characterised by specific *gyrB* PCRs or by restriction analysis of the initial *M. tuberculosis* complex *gyrB* PCR product [46,47]. With the only exception of *M. tuberculosis* and *M. africanum* subtype II, it is thus possible to identify all *M. tuberculosis* complex members from each other [47]. To differentiate these indistinguishable species, phenotypic characteristics must be still relied upon (e.g., *M.*

*tuberculosis* shows an eugonic and aerophilic growth on Lebek medium, whereas *M. africanum* subtype II grows dysgonic and microaerophilic, [53]). Furthermore, if desired, virulent *M. bovis* isolates can be distinguished from the BCG *M. bovis* isolates with the help of a RD1 multiplex PCR [48]. NTM isolates are 5'-16S rDNA sequenced, which, in most cases, results in the unequivocal identification of a known or sometimes unknown mycobacterial species. Further molecular characterisation will be needed in only a few cases. To distinguish *M. marinum* from *M. ulcerans*, a PCR targeting the insertion element (IS) 2404 can be used [49]. Similarly, to differentiate the *M. avium* subspecies from each other IS 900 and IS 902 PCRs can be employed [50,51]. However most serovar 2 (porcine origin) and some serovar 1 and 3 *M. avium* subsp. *avium* strains isolated from animals are "wrongly" IS 902 positive [52]. The remaining pairs of species, including the clinically important *M. chelonae* sqv. I / *M. abscessus* and *M. gastri* / *M. kansasii* sqv. I & IV, can be distinguished by an ITS PCR followed by either a sequence determination or by a restriction endonuclease (RE) assay of the PCR product [11].

The logic incorporated in the commercially available, MicroSeq 500 16S rDNA bacterial identification system (Applied Biosystems, Forster City, CA) is similar to that in the RIDOM system since it uses newly-determined, non-ragged 16S rDNA sequences from ATCC culture collection isolates as a reference database. However, there are some fundamental differences between Microseq and RIDOM. The most notable of these is that the RIDOM system is publicly accessible and, because of its open hypertext structure, allows the incorporation of other useful Internet sources. Furthermore, the RIDOM approach is far-reaching in that it not only tries to include sequences and species names in its database, but also additional information related to taxonomy and disease. The MicroSeq database has only one entry per species (i.e., in most cases the type strain sequence) and currently contains only 63 unique sequences (software version 1.36), whereas RIDOM (version 1.0) incorporates 123 unique, newly-determined 5'-16S rDNA mycobacterial sequences. The superiority of MicroSeq in comparison to approaches based on phenotypic identification of *Mycobacterium* isolates has been demonstrated [6]. Because RIDOM addresses intraspecies variation, a procedure which is totally absent in the commercial MicroSeq database, the performance of RIDOM in differentiating mycobacteria is superior to that of MicroSeq. This was recently shown [54]. Detailed descriptions of MicroSeq and RIDOM have been published [20,21,55]. RIDOM, being one of the first solely diagnostic-orientated genetic public databases, was also recently included in the database issue of the Nucleic Acids Research journal [22].

Evaluations of the quality and accuracy of results obtained using the two specialized databases (RIDOM version 1.0 and MicroSeq 500 v. 1.36) and the more general GenBank and RDP-II databases have recently been published [33,54]. Newly determined 5'-16S rDNA sequences from ATCC *Mycobacterium* type strains ( $n = 79$ ) and from clinical isolates ( $n = 94$ ) were analyzed. All of the type strain sequences analyzed using RIDOM were correct with 100% similarity. MicroSeq does not include sequences of all established species. Those for *M. lentiflavum* for example, and many of the most recently described species are not available and consequently some type strains were misidentified. In contrast, only 23% and 25% of species had a perfect match with sequences from GenBank and RDP-II, respectively. A high percentage, 39% and 34% of the type strain sequences were not given top scores against GenBank and RDP-II, respectively. Therefore, these strains would not have been identified correctly [33]. Commenting on their results, Turenne et al. state that: "The high proportion of misleading results obtained from public databases is not surprising, since submissions are not peer reviewed. Similarity searches can result in the user not obtaining a true identification of an organism, even when the organism sequence is present in the database [33]." On querying the different databases with the 94 clinical isolate sequences, RIDOM gave a perfect match in 92.5%, whereas MicroSeq yielded this result in only 69.2% of cases. Only 4.3% of the RIDOM results had a similarity of 99% or below, which we regard as a threshold for the criteria of a "distinct species". MicroSeq failed to surpass this threshold in 17.0% of cases [54]. Cloud et al., although expressing concern about the costs of sequencing, argue in their study that the sequencing technique in combination with a high-quality database is an excellent tool for species identification of mycobacteria, which reduces turnaround time and makes repeat analysis and confirmation of questionable results with biochemicals unnecessary [54].

The availability of a comprehensive, high-quality sequence database enabled us to systematically examine named, but not authenticated, GenBank entries and sequences from culture collection strains. Table 2 (see Additional file Table 2) lists the similarity search best matches for these sequences together with our conclusions regarding the justification for considering a previously non-valid named species as a new species or whether the non-valid name refers to just a synonym. Out of 29 analyzed entries, 9 showed a 100% similarity match with a valid species. Therefore the names of these isolates are most probably synonyms of already known species. One entry, "*M. album*", does not appear to be a *Mycobacterium* at all. When the reporting criteria, established by Patel et al. [6], were applied to the remaining entries, seven ("*M. acapulcensis*", *M. doricum*, "*M. fluoroan-*

thenivorans", *M. immunogenum*, *M. kubicae*, "M. monacense", and "M. petroleophilum") showed a best match with a validly described species below 98.2%. The chance is therefore high that these are not just genospecies but indeed new species. It is difficult to predict the correct status of the remaining 12 entries. However, nine of these isolates had a best match equal or above 99.0% and are therefore most probably not new species. It needs to be mentioned that our evaluations are based solely upon the percentage similarity of the partial 16S rDNA with that of a known species. Other methods, in addition to rDNA sequencing (e.g., DNA-DNA hybridization and phenotypic tests), need to be employed before confirming or rejecting new species [56]. Therefore, while the RIDOM database is quite complete, one should not accept it as the sole definitive authority for establishing mycobacterial species. The pitfalls in the 16S rDNA "similarity-only" approach is illustrated in the case of *M. elephantis* (99.3% best match with *M. pulveris*). This recently and validly described species would not have been regarded as a new species using the above stated criteria [57]. It is interesting to note that *M. elephantis* was established mainly because of its unique, nearly complete, 16S rDNA sequence. Up to now, however, GenBank does not contain even a partial 16S rDNA sequence of *M. pulveris*, a species described as early as 1983 and which is apparently most closely related to *M. elephantis*.

### Conclusions

The data from this analysis of all validly published mycobacterial species, in conjunction with the previously published evaluations of our database [33,54], show that it is possible to differentiate most mycobacterial species by sequence analysis of partial 16S rDNA. A database should be exhaustive [58] and, should include more than just one representative strain of each species because of the marked intraspecies variability. A molecular diagnosis system must involve multiple molecular targets, since not all *Mycobacterium* species can be differentiated using 5'-16S rDNA sequencing alone. The cost burden for the sequencing method is constantly dropping. Under certain conditions, it may be already less expensive than conventional methods [59]. Therefore, the sequencing technique should be considered for routine application, not only for reference laboratories. For this purpose, a high-quality database needed for such an implementation is available under the URL <http://www.ridom-rdna.de>. Users can submit a sequence and conduct a similarity search for mycobacterial identification purposes against the RIDOM reference database. Furthermore, because of the open hypertext structure of RIDOM, many links to other World Wide Web services are established, thereby augmenting the information content. Links in the opposite direction are also possible since GenBank and NCBI Taxonomy, for

example, link back to RIDOM in the frame of the NCBI LinkOut project [60].

### Competing interests

DH is cofounder and one of the managing directors of a bioinformatics start-up company called Ridom GmbH (Würzburg, Germany).

### Authors' contributions

DH is the principal investigator of the RIDOM project. He designed the study and drafted primarily the manuscript. SD has performed most experimental work of the study (if not stated explicitly differently following) and compiled mycobacterial species descriptions. AR has done the ITS restriction analyses and helped drafting the manuscript. SN helped in *gyrB* analysis of the mycobacteria. JR and MS were responsible under guidance of JA for the implementation of the RIDOM web service. MF participated in drafting the manuscript. ER phenotypically characterised the mycobacterial isolates and also helped drafting the manuscript. All authors read and approved the final manuscript.

### Additional material

#### Additional File Table 1

5'-16S rDNA (16S) and internal transcribed spacer (ITS) sequenced *Mycobacterium* species included in the RIDOM database (Tab\_1.pdf, with references [11] and [61]).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2334-3-26-S1.pdf>]

#### Additional File Figure 3

5'-16S rDNA and ITS multiple alignments of all observed mycobacterial species sequenced (Fig\_3.pdf).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2334-3-26-S3.pdf>]

#### Additional File Figure 4

Algorithm for genetic differentiation of all mycobacteria, including primer sequences, experimentally verified *gyrB*- and IS-PCR conditions as well as references to the various methods (Fig\_4.pdf).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2334-3-26-S4.pdf>]

#### Additional File Table 2

5'-16S rDNA and ITS RIDOM or GenBank similarity search best matches with non-valid named mycobacterial GenBank entries or newly determined sequences from non-valid named culture collection strains (Tab\_2.pdf, with references [57], [62], [63] and [64]).

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2334-3-26-S2.pdf>]

## Acknowledgments

We are very grateful to Monika Bergmann, Susanne Ebner, Angelika Hansen, Marion Patzke-Öchsner and Inge Ray from the Institute for Hygiene and Microbiology, Würzburg for their excellent technical assistance. We also thank scientists all over the world who provided us isolates from their collections for inclusion into our database available. In particular, we would like to thank the following persons: Gisela Bretzel, Armauer Hansen Institut, Würzburg, Germany; Geoff de Lisle, AgResearch, Wallaceville Animal Research Centre, Upper Hutt, New Zealand; Siobhan Hughes, Veterinary Sciences Division, Department of Agriculture for Northern Ireland, Belfast, Northern Ireland; Yoshiko Kashiwabara, Leprosy Research Center, Tokyo, Japan; Francoise Portael, Institute of Tropical Medicine, Department of Microbiology, Antwerpen, Belgium; Udo Reischl, Institut für Mikrobiologie Universität Regensburg, Germany; Pirjo Torckko, National Public Health Institute, Kuopio, Finland; Christine Turenne, National Reference Center for Mycobacteriology, Winnipeg, Canada; and Veronique Vincent, Laboratoire de Reference des Mycobacteries, Institut Pasteur, Paris, France.

## References

- Murray PR, Baron EJ, Pfaller MA, Tenover FC and Tenover RH: *Manual of clinical microbiology* 7th edition. Edited by: Murray PR, Baron EJ, Pfaller MA, Tenover FC and Tenover RH. Washington, D.C., American Society for Microbiology; 1999.
- Butler WR, Jost K.C., Jr and Kilburn JO: **Identification of mycobacteria by high-performance liquid chromatography.** *J Clin Microbiol* 1991, **29**:2468-2472.
- Luquin M, Ausina V, Lopez Calahorra F, Belda F, Garcia Barcelo M., Celma C and Prats G: **Evaluation of practical chromatographic procedures for identification of clinical isolates of mycobacteria.** *J Clin Microbiol* 1991, **29**:120-130.
- Persing DH, Smith TF, Tenover FC and White TJ: *Diagnostic molecular microbiology: principles and applications* Edited by: Persing DH, Smith TF, Tenover FC and White TJ. Washington, D.C., American Society for Microbiology; 1993.
- Springer B, Stockman L, Teschner K, Roberts GD and Böttger EC: **Two-laboratory collaborative study on identification of mycobacteria: molecular versus phenotypic methods.** *J Clin Microbiol* 1996, **34**:296-303.
- Patel JB, Leonard DG, Pan X, Musser JM, Berman RE and Nachamkin I: **Sequence-based identification of Mycobacterium species using the MicroSeq 500 16S rDNA bacterial identification system.** *J Clin Microbiol* 2000, **38**:246-251.
- Kirschner P, Springer B, Vogel U, Meier A, Wrede A, Kiekenbeck M, Bange FC and Böttger EC: **Genotypic identification of mycobacteria by nucleic acid sequence determination: report of a 2-year experience in a clinical laboratory.** *J Clin Microbiol* 1993, **31**:2882-2889.
- Stahl DA and Urbance JW: **The division between fast- and slow-growing species corresponds to natural relationships among the mycobacteria.** *J Bacteriol* 1990, **172**:116-124.
- Rogall T, Wolters J, Flohr T and Böttger EC: **Towards a phylogeny and definition of species at the molecular level within the genus Mycobacterium.** *Int J Syst Bacteriol* 1990, **40**:323-330.
- Pitulle C, Dorsch M, Kazda J, Wolters J and Stackebrandt E: **Phylogeny of rapidly growing members of the genus Mycobacterium.** *Int J Syst Bacteriol* 1992, **42**:337-343.
- Roth A, Reischl U, Streubel A, Naumann L, Kroppenstedt RM, Habicht M, Fischer M and Mauch H: **Novel diagnostic algorithm for identification of mycobacteria using genus-specific amplification of the 16S-23S rRNA gene spacer and restriction endonucleases.** *J Clin Microbiol* 2000, **38**:1094-1104.
- Telenti A, Marchesi F, Balz M, Bally F, Böttger EC and Bodmer T: **Rapid identification of mycobacteria to the species level by polymerase chain reaction and restriction enzyme analysis.** *J Clin Microbiol* 1993, **31**:175-178.
- Devallois A, Goh KS and Rastogi N: **Rapid identification of mycobacteria to species level by PCR-restriction fragment length polymorphism analysis of the hsp65 gene and proposition of an algorithm to differentiate 34 mycobacterial species.** *J Clin Microbiol* 1997, **35**:2969-2973.
- Ringuet H, Akoua-Koffi C, Honore S, Varnerot A, Vincent V, Berche P, Gaillard JL and Pierre-Audigier C: **hsp65 sequencing for identification of rapidly growing mycobacteria.** *J Clin Microbiol* 1999, **37**:852-857.
- Woese CR: **Bacterial evolution.** *Microbiol Rev* 1987, **51**:221-271.
- Harmsen D, Heesemann J, Brabletz T, Kirchner T and Müller-Hermelink HK: **Heterogeneity among Whipple's-disease-associated bacteria.** *Lancet* 1994, **343**:1288..
- Holberg-Petersen M, Steinbakk M, Figenschau KJ, Jantzen E, Eng J and Melby KK: **Identification of clinical isolates of Mycobacterium spp. by sequence analysis of the 16S ribosomal RNA gene. Experience from a clinical laboratory.** *APMIS* 1999, **107**:231-239.
- Maidak BL, Cole JR, Lilburn TG, Parker C.T., Jr, Saxman PR, Farris RJ, Garrity GM, Olsen GJ, Schmidt TM and Tiedje JM: **The RDP-II (ribosomal database project).** *Nucleic Acids Res* 2001, **29**:173-174.
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA and Wheeler DL: **GenBank.** *Nucleic Acids Res* 2000, **28**:15-18.
- Harmsen D, Rothgänger J, Singer C, Albert J and Frosch M: **Intuitive hypertext-based molecular identification of micro-organisms.** *Lancet* 1999, **353**:291..
- Harmsen D, Singer C, Rothgänger J, Tønjum T, de Hoog GS, Shah H, Albert J and Frosch M: **Diagnostics of Neisseriaceae and Moraxellaceae by ribosomal DNA sequencing: ribosomal differentiation of medical microorganisms.** *J Clin Microbiol* 2001, **39**:936-942.
- Harmsen D, Rothgänger J, Frosch M and Albert J: **RIDOM: Ribosomal Differentiation of Medical Micro-organisms Database.** *Nucleic Acids Res* 2002, **30**:416-417.
- van Embden JD, Cave MD, Crawford JT, Dale JW, Eisenach KD, Giquel B, Hermans P, Martin C, McAdam R, Shinnick TM and Small PM: **Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology.** *J Clin Microbiol* 1993, **31**:406-409.
- Lane DJ: **16S/23S rRNA sequencing.** *Nucleic acid techniques in bacterial systematics* Edited by: Stackebrandt E and Goodfellow M. Chichester, United Kingdom, John Wiley & Sons Ltd.; 1991:115-175.
- Ewing B and Green P: **Base-calling of automated sequencer traces using phred. II. Error probabilities.** *Genome Res* 1998, **8**:186-194.
- Pearson WR and Lipman DJ: **Improved tools for biological sequence comparison.** *Proc Natl Acad Sci USA* 1988, **85**:2444-2448.
- Thompson JD, Higgins DG and Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
- Frothingham R and Wilson KH: **Sequence-based differentiation of strains in the Mycobacterium avium complex.** *J Bacteriol* 1993, **175**:2818-2825.
- Picardeau M, Prod'Hom G, Raskine L, LePennec MP and Vincent V: **Genotypic characterization of five subspecies of Mycobacterium kansasii.** *J Clin Microbiol* 1997, **35**:25-32.
- Olsen GJ, Woese CR and Overbeek R: **The winds of (evolutionary) change: breathing new life into microbiology.** *J Bacteriol* 1994, **176**:1-6.
- Peterson EM, Lu R, Floyd C, Nakasone A, Friedly G and de la Maza LM: **Direct identification of Mycobacterium tuberculosis, Mycobacterium avium, and Mycobacterium intracellulare from amplified primary cultures in BACTEC media using DNA probes.** *J Clin Microbiol* 1989, **27**:1543-1547.
- Tortoli E, Nanetti A, Piersimoni C, Cichero P, Farina C, Mucignat G, Scarparo C, Bartolini L, Valentini R, Nista D, Gesu G, Tosi CP, Crovatto M and Brusarosco G: **Performance assessment of new multiplex probe assay for identification of mycobacteria.** *J Clin Microbiol* 2001, **39**:1079-1084.
- Turenne CY, Tschetter L, Wolfe J and Kabani A: **Necessity of quality-controlled 16S rRNA gene sequence databases: identifying nontuberculous Mycobacterium species.** *J Clin Microbiol* 2001, **39**:3637-3648.
- Frothingham R and Wilson KH: **Molecular phylogeny of the Mycobacterium avium complex demonstrates clinically meaningful divisions.** *J Infect Dis* 1994, **169**:305-312.
- Bödinghaus B, Wolters J, Heikens W and Böttger EC: **Phylogenetic analysis and identification of different serovars of**

- Mycobacterium intracellulare** at the molecular level. *FEMS Microbiol Lett* 1990, **70**:197-203.
36. De Smet KA, Hellyer TJ, Khan AW, Brown IN and Ivanyi J: **Genetic and serovar typing of clinical isolates of the *Mycobacterium avium-intracellulare* complex.** *Tuber Lung Dis* 1996, **77**:71-76.
  37. De Smet KA, Brown IN, Yates M and Ivanyi J: **Ribosomal internal transcribed spacer sequences are identical among *Mycobacterium avium-intracellulare* complex isolates from AIDS patients, but vary among isolates from elderly pulmonary disease patients.** *Microbiology* 1995, **141**:2739-2747.
  38. Springer B, Böttger EC, Kirschner P and Wallace RJ: **Phylogeny of the *Mycobacterium chelonae*-like organism based on partial sequencing of the 16S rRNA gene and proposal of *Mycobacterium mucogenicum* sp. nov.** *Int J Syst Bacteriol* 1995, **45**:262-267.
  39. Kirschner P and Böttger EC: **Microheterogeneity within rRNA of *Mycobacterium goodnae*.** *J Clin Microbiol* 1992, **30**:1049-1050.
  40. Richter E, Niemann S, Rüsck-Gerdes S and Hoffner S: **Identification of *Mycobacterium kansasii* by using a DNA probe (Accu-Probe) and molecular techniques.** *J Clin Microbiol* 1999, **37**:964-970.
  41. Alcaide F, Richter I, Bernasconi C, Springer B, Hagenau C, Schulze-Röbbecke R, Tortoli E, Martin R, Böttger EC and Telenti A: **Heterogeneity and clonality among isolates of *Mycobacterium kansasii*: implications for epidemiological and pathogenicity studies.** *J Clin Microbiol* 1997, **35**:1959-1964.
  42. Springer B, Wu W-K, Bodmer T, Haase G, Pfyffer GE, Kroppenstedt RM, Schröder K-H, Emler S, Kilburn JO, Kirschner P, Telenti A, Coyle MB and Böttger EC: **Isolation and characterization of a unique group of slowly growing mycobacteria: description of *Mycobacterium lentiflavum* sp. nov.** *J Clin Microbiol* 1996, **34**:1100-1107.
  43. Roth A, Fischer M, Hamid ME, Michalke S, Ludwig W and Mauch H: **Differentiation of phylogenetically related slowly growing mycobacteria based on 16S-23S rRNA gene internal transcribed spacer sequences.** *J Clin Microbiol* 1998, **36**:139-147.
  44. Torkko P, Suutari M, Suomalainen S, Paulin L, Larsson L and Katila ML: **Separation among species of *Mycobacterium terrae* complex by lipid analyses: comparison with biochemical tests and 16S rRNA sequencing.** *J Clin Microbiol* 1998, **36**:499-505.
  45. Portaels F, Fonteyne P-A, de Beenhouwer H, de Rijk P, Guedenon A, Hayman J and Meyers WM: **Variability in 3' end of 16S rRNA sequence of *Mycobacterium ulcerans* is related to geographic origin of isolates.** *J Clin Microbiol* 1996, **34**:962-965.
  46. Kasai H, Ezaki T and Harayama S: **Differentiation of phylogenetically related slowly growing mycobacteria by their *gyrB* sequences.** *J Clin Microbiol* 2000, **38**:301-308.
  47. Niemann S, Harmsen D, Rüsck-Gerdes S and Richter E: **Differentiation of clinical *Mycobacterium tuberculosis* complex isolates by *gyrB* DNA sequence polymorphism analysis.** *J Clin Microbiol* 2000, **38**:3231-3234.
  48. Talbot EA, Williams DL and Frothingham R: **PCR identification of *Mycobacterium bovis* BCG.** *J Clin Microbiol* 1997, **35**:566-569.
  49. Steinar T, Ross BC, Davies JK, Marino L, Robins-Browne RM, Oppedisano F, Sievers A and Johnson PD: **Identification and characterization of IS2404 and IS2606: two distinct repeated sequences for detection of *Mycobacterium ulcerans* by PCR.** *J Clin Microbiol* 1999, **37**:1018-1023.
  50. Moss MT, Sanderson JD, Tizard MLV, Hermon-Taylor J, El-Zaatari FAK, Markesich DC and Graham DY: **Polymerase chain reaction detection of *Mycobacterium paratuberculosis* and *Mycobacterium avium* subsp. *silvaticum* in long term cultures from Crohn's disease and control tissues.** *Gut* 1992, **33**:1209-1213.
  51. Sanderson JD, Moss MT, Tizard ML and Hermon-Taylor J: ***Mycobacterium paratuberculosis* DNA in Crohn's disease tissue.** *Gut* 1992, **33**:890-896.
  52. Ahrens P, Giese SB, Klausen J and Inglis NF: **Two markers, IS901-IS902 and p40, identified by PCR and by using monoclonal antibodies in *Mycobacterium avium* strains.** *J Clin Microbiol* 1995, **33**:1049-1053.
  53. Niemann S, Richter E and Rüsck-Gerdes S: **Differentiation among members of the *Mycobacterium tuberculosis* complex by molecular and biochemical features: evidence for two pyrazinamide-susceptible subtypes of *M. bovis*.** *J Clin Microbiol* 2000, **38**:152-157.
  54. Cloud JL, Neal H, Rosenberry R, Turenne CY, Jama M, Hillyard DR and Carroll KC: **Identification of *Mycobacterium* spp. by using a commercial 16S ribosomal DNA sequencing kit and additional sequencing libraries.** *J Clin Microbiol* 2002, **40**:400-406.
  55. Tang YW, Ellis NM, Hopkins MK, Smith DH, Dodge DE and Persing DH: **Comparison of phenotypic and genotypic techniques for identification of unusual aerobic pathogenic gram-negative bacilli.** *J Clin Microbiol* 1998, **36**:3674-3679.
  56. Stackebrandt E and Goebel BM: **Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology.** *Int J Syst Bacteriol* 1994, **44**:846-849.
  57. Shojaei H, Magee JG, Freeman R, Yates M, Horadagoda NU and Goodfellow M: ***Mycobacterium elephantis* sp. nov., a rapidly growing non-chromogenic *Mycobacterium* isolated from an elephant.** *Int J Syst Evol Microbiol* 2000, **50**:1817-1820.
  58. Clayton RA, Sutton G, Hinkle P.S., Jr, Bult C and Fields C: **Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa.** *Int J Syst Bacteriol* 1995, **45**:595-599.
  59. Cook VJ, Turenne CY, Wolfe J, Pauls R and Kabani A: **Conventional methods versus 16S ribosomal DNA sequencing for identification of nontuberculous mycobacteria: cost analysis.** *J Clin Microbiol* 2003, **41**:1010-1015.
  60. Wheeler DL, Church DM, Lash AE, Leipe DD, Madden TL, Pontius JU, Schuler GD, Schriml LM, Tatusova TA, Wagner L and Rapp BA: **Database resources of the National Center for Biotechnology Information.** *Nucleic Acids Res* 2001, **29**:11-16.
  61. Roth A, Reischl U, Schonfeld N, Naumann L, Emler S, Fischer M, Mauch H, Loddenkemper R and Kroppenstedt RM: ***Mycobacterium heckeshornense* sp. nov., a new pathogenic slowly growing *Mycobacterium* sp. causing cavitory lung disease in an immunocompetent patient.** *J Clin Microbiol* 2000, **38**:4102-4107.
  62. Tortoli E, Piersimoni C, Kroppenstedt RM, Montoya-Burgos JI, Reischl U, Giacometti A and Emler S: ***Mycobacterium doricum* sp. nov.** *Int J Syst Evol Microbiol* 2001, **51**:2007-2012.
  63. Wilson RW, Steingrube VA, Böttger EC, Springer B, Brown-Elliott BA, Vincent V, Jost K.C., Jr, Zhang Y, Garcia MJ, Chiu SH, Onyi GO, Rossmoore H, Nash DR and Wallace RJ: ***Mycobacterium immunogenium* sp. nov., a novel species related to *Mycobacterium mabscessus* and associated with clinical disease, pseudo-outbreaks and contaminated metalworking fluids: an international cooperative study on mycobacterial taxonomy.** *Int J Syst Evol Microbiol* 2001, **51**:1751-1764.
  64. Floyd MM, Gross WM, Bonato DA, Silcox VA, Smithwick RW, Metchock B, Crawford JT and Butler WR: ***Mycobacterium kubicae* sp. nov., a slowly growing, scotochromogenic *Mycobacterium*.** *Int J Syst Evol Microbiol* 2000, **50**:1811-1816.

## Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2334/3/26/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

