

RIDOM: Ribosomal Differentiation of Medical Micro-organisms Database

Dag Harmsen^{1,2,*}, Jörg Rothgänger³, Matthias Frosch¹ and Jürgen Albert³

¹Institut für Hygiene und Mikrobiologie, Universität Würzburg, D-97080 Würzburg, Germany, ²CREATOGEN AG, Ulmer Strasse 160a, D-86156 Augsburg, Germany and ³Lehrstuhl für Informatik II, Universität Würzburg, D-97074 Würzburg, Germany

Received August 1, 2001; Revised and Accepted September 10, 2001

ABSTRACT

The ribosomal differentiation of medical micro-organisms (RIDOM) web server, first described by Harmsen *et al.* [Harmsden,D., Rothganger,J., Singer,C., Albert,J. and Frosch,M. (1999) *Lancet*, 353, 291], is an evolving electronic resource designed to provide micro-organism differentiation services for medical identification needs. The diagnostic procedure begins with a specimen partial small subunit ribosomal DNA (16S rDNA) sequence. Resulting from a similarity search, a species or genus name for the specimen in question will be returned. Where the first results are ambiguous or do not define to species level, hints for further molecular, i.e. internal transcribed spacer, and conventional phenotypic differentiation will be offered ('sequential and polyphasic approach'). Additionally, each entry in RIDOM contains detailed medical and taxonomic information linked, context-sensitive, to external World Wide Web services. Nearly all sequences are newly determined and the sequence chromatograms are available for intersubjective quality control. Similarity searches are now also possible by direct submission of trace files (ABI or SCF format). Based on the PHRED/PHRAP software, error probability measures are attached to each predicted nucleotide base and visualised with a new 'Trace Editor'. The RIDOM web site is directly accessible on the World Wide Web at <http://www.ridom.de/>. The email address for questions and comments is webmaster@ridom.de.

INTRODUCTION

One major issue in microbiology is the identification of micro-organisms. Reliable identification by conventional methods often requires several approaches, such as the use of colony morphology, Gram staining, nutritional requirements and/or biochemical reactions. These methods are time consuming and can provide ambiguous results. Molecular identification offers two primary advantages to phenotypic identification: rapid

turn-around time and improved accuracy. With the use of sequence information obtained from the 5' end of the 16S ribosomal RNA gene (16S rDNA), most bacteria—even those dead or uncultivated—can be differentiated (1,2). However, sequence based identification is hampered by some drawbacks in currently accessible databases, e.g. GenBank or RDP (3,4). These problems include the presence of ragged sequence ends (often yielding wrong 'best' matches in similarity searches), faulty sequence entries (mainly due to early error-prone sequencing techniques), non-characterised entries and the use of out-dated nomenclature. Therefore, for the RIDOM project, we are determining non-ragged 5'-16S rDNA and internal transcribed spacer (ITS) sequences of culture collection strains and making the primary data (i.e. the sequence chromatograms) available to users (5). Using the PHRED/PHRAP software (6), we attach error probability measures to each predicted user and database nucleotide base. A 'Trace Editor' shows the sequence traces with error flags and is capable of interactive sequence editing.

USING RIDOM

Upon entering the RIDOM web page at <http://www.ridom.de/>, visitors will find detailed instructions on how to proceed. DIAGNOSTICS links to similarity search analysis of user sequences. The analysis starts with a hypervariable partial 16S rDNA sequence (*Escherichia coli* positions 54–510, i.e. one sequencing run) from the specimen to be identified that is submitted to the query interface. From an interpreted similarity search, a species or genus name is returned. If the first results do not define to species level, hints for further molecular (ITS) and phenotypic identification are given ('sequential and polyphasic approach'). Genus, species descriptions and taxonomic information is generated out of the database according to the search result. Because of the open hypertext-structure of RIDOM, many links to other World Wide Web services are established (<http://www.ridom.de/static/externdbms.html>) thereby augmenting the information content. In the opposite direction GenBank and NCBI Taxonomy link back to RIDOM in the frame of the NCBI LinkOut project.

For quality control reasons, only sequences from strains held in culture collections are included in our reference database. To ensure updated taxonomic entries, bacterial names are checked

*To whom correspondence should be addressed at: RIDOM bioinformatics GmbH, D-97074 Würzburg, Germany. Tel: +49 163 4043854; Email: dharmsen@ridom.de

with the aid of the 'DSMZ Bacterial Nomenclature Up-to-date' database (<http://www.dsmz.de/bactnom/bactname.htm>).

Data contained in the RIDOM database can be accessed in two different modes. SEARCH returns records based on taxon or free text searches. BROWSE permits location of species alphabetically or by phylogenetic placement. Finally, HELP on using the system, tutorials, protocols and primer resources are contained in static HTML files linked to the home page.

To view the sequence chromatograms in the new 'Trace Editor', client computers need a recent version of a standard web browser (Netscape or IE v.4 or higher) and Sun's Java Plug-in 1.2 or higher installed. Main backend components of RIDOM include the PHRED/PHRAP, FASTA and CLUSTAL W programmes that are embedded into Java Servlets (6–8).

FUTURE CHANGES AND ADDITIONS

RIDOM currently contains 235 entries belonging to the *Neisseriaceae*, *Moraxellaceae* (9) and *Mycobacterium* genus (for further mycobacteria specific information and downloads it is referred to <http://www.ridom.de/mycobacteria/>). A constant increase in the number of RIDOM entries is under development and periodic updates are planned. However, we are keen to expand the content of our database at an even faster pace and therefore welcome contributions from additional new partners in the RIDOM project (further information, <http://www.ridom.de/static/partner.html>). Software changes and

additions will be stimulated by user feedback (Email: webmaster@ridom.de).

REFERENCES

1. Fredricks, D.N. and Relman, D.A. (1996) Sequence-based identification of microbial pathogens: a reconsideration of Koch's postulates. *Clin. Microbiol. Rev.*, **9**, 18–33.
2. Harmsen, D., Heesemann, J., Brabletz, T., Kirchner, T. and Müller-Hermelink, H.K. (1994) Heterogeneity among Whipple's-disease-associated bacteria. *Lancet*, **343**, 1288.
3. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A. and Wheeler, D.L. (2000) GenBank. *Nucleic Acids Res.*, **28**, 15–18. Updated article in this issue: *Nucleic Acids Res.* (2002), **30**, 17–20.
4. Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker, C.T., Jr, Saxman, P.R., Farris, R.J., Garrity, G.M., Olsen, G.J., Schmidt, T.M. and Tiedje, J.M. (2001) The RDP-II (Ribosomal Database Project). *Nucleic Acids Res.*, **29**, 173–174.
5. Harmsen, D., Rothgänger, J., Singer, C., Albert, J. and Frosch, M. (1999) Intuitive hypertext-based molecular identification of micro-organisms. *Lancet*, **353**, 291.
6. Ewing, B. and Green, P. (1998) Base-calling of automated sequencer traces using PHRED. II. Error probabilities. *Genome Res.*, **8**, 186–194.
7. Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
8. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
9. Harmsen, D., Singer, C., Rothgänger, J., Tønjum, T., de Hoog, G.S., Shah, H., Albert, J. and Frosch, M. (2001) Diagnostics of *Neisseriaceae* and *Moraxellaceae* by ribosomal DNA sequencing: ribosomal differentiation of medical micro-organisms. *J. Clin. Microbiol.*, **39**, 936–942.