

A New Web Server for the Rapid Identification of Microorganisms

Olivier Croce^{1*}, François Chevenet² and Richard Christen¹

¹University of Nice Sophia-Antipolis and CNRS UMR 6543. Institute of Signaling, Developmental Biology and Cancer. Virtual Biology Lab. Parc Valrose. F06108 Nice, France

²Génétique et Evolution des Maladies Infectieuses, IRD / CNRS UMR 2724 - IRD, 911 avenue Agropolis, B.P. 64501, F34394 Montpellier Cedex 5, France

Abstract

Small subunit ribosomal RNA (SSU rRNA) gene sequences are now widely used to identify microbial organisms. However, the public databases now comprise mostly rRNA gene sequences resulting from PCR and cloning of DNA isolated from environmental samples. Most of these sequences are poorly annotated and do not bear proper taxonomic assignments. As a result, this flood of sequences makes routine identifications often very tedious.

We propose a new web server for fast and reliable identifications of microbial isolates. It allows retrieving related sequences and their taxonomies, in order to proceed to phylogenetic analyses. This web tool is based on specific databases of cultured species for Bacteria, Archaea and Protozoa. In addition, various data helping the identification procedure can be downloaded. The server also includes online software to automatically display annotated phylogenetic trees.

Keywords: Identification of microorganisms; RNA sequences; BLAST; Web server; Phylogenetic tree

Introduction

Identifications of microbial organisms are now usually done by comparing their SSU rRNA gene sequences to those of known organisms (Stackebrandt and Goebel, 1994). The usual application is to study the composition of the microbial community within a given environmental or clinical sample. SSU rRNA gene sequences are thus obtained (McCabe et al., 1999), either after cloning the PCR products and random sequencing a set of clones (Amann et al., 1995; Hugenholz and Pace, 1996) or by pyrosequencing (Jonasson et al., 2002; Roesch et al., 2007; Huse et al., 2008; Christen, 2008). The questions are to find out if these sequences are related to other sequences already found in environmental samples, and/or related to well known cultured microorganisms and eventually a type strain (Albuquerque et al., 2009).

The general process starts with a similarity search of the new sequence (s) against the public databases usually using BLAST (Altschul et al., 1990), then align (s) to the most similar sequences and finally do a classification or a phylogenetic analysis in order to proceed to identifications. Online BLAST servers are a common choice to quickly identify related sequences, but the databases they rely on are now filled with partial sequences from environmental samples that are often poorly annotated. For example, the NCBI BLAST server (www.ncbi.nlm.nih.gov/BLAST/) and the EBI server (www.ebi.ac.uk/blast2/) allow the exclusion of environmental sequences, but these databases still contain many inaccurate descriptions. A similar database restricted to 16S rRNA sequences exists at DDBJ (<http://blast.ddbj.nig.ac.jp/top-e.html>). As a result, any BLAST query returns many (sometimes only) poorly described sequences (Lin et al., 2008; Clarridge, 2004). Thus, a tedious manual analysis of the BLAST results is necessary to identify closely related well described species. Some other tools such as Etaxon (Chun et al., 2007) are more specific. Etaxon is associated with a hand-curated database of 16S rRNA gene sequences for bacterial type strains. It allows users to perform similarity-based searches, multiple sequences alignment and various phylogenetic analyses. The database of Etaxon is extremely useful for characterization of a new species before its publication, but being

restricted to bacterial type strains, it does not allow to identify well deposited cultured species that have not been validated as type strains. Also, it does not allow for identifications of protozoa and archaea. Finally, it does not provide taxonomic informations required to construct a fully annotated phylogenetic tree.

Our server allows BLAST searches on cultured species, with restrictions on sequence's length as well as for example using only two sequences per species. Moreover, we made an online tool named "Blast2Tree" (PHP / Javascript / MySQL). The main goal of Blast2Tree is to derive a fully annotated phylogenetic tree in a very simple way from the BLAST hits or the results obtained from Etaxon. Annotations are provided from a local database and through a pipeline that uses ScripTree (Chevenet et al., 2010). ScripTree is a tool for scripting phylogenetic graphics. It allows the management of multiple trees and usual kinds of annotations. It can be used either as a stand-alone package or included in a pipeline and linked to a HTTP server. Also, our online tool Blast2Tree is able to download every sequence from a clade as well as the related annotations to be used by software such as TreeDyn (Chevenet et al., 2006).

Materials and Methods

Finding similar rRNA sequences using a specific BLAST

The database server described here contains SSU rRNA sequences extracted from the EMBL database. Each entry is parsed with a Python script to check if the species definitions line reveals or not that a proper latin species name has been used. Therefore, descriptions

***Corresponding author:** Dr. Olivier Croce, University of Nice Sophia-Antipolis, Institute of Signaling, Developmental Biology and Cancer. Virtual Biology Lab, Faculté des Sciences, Parc Valrose, 06108 Nice, France, Tel: +33 492 07 6947; E-mail: croce@unice.fr

Received June 08, 2010; **Accepted** June 29, 2010; **Published** June 29, 2010

Citation: Croce O, Chevenet F, Christen R (2010) A New Web Server for the Rapid Identification of Microorganisms. J Microbial Biochem Technol 2: 084-088. doi:10.4172/1948-5948.1000029

Copyright: © 2010 Croce O, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



such as 'uncultured', 'env', 'Genus sp.', 'genomosp.' are automatically excluded. If necessary SSU rRNA subsequences are extracted from longer genomic fragment. The NCBI taxonomic description is also fully extracted, checked and associated to each sequence. Finally, we use the "List of Prokaryotic names with Standing in Nomenclature" at <http://www.bacterio.net> (Euzéby, 2008) to identify all the bacterial type strains. The data is stored into a local relational database (MySQL). Several databases for BLAST are formatted with the previous data, for Bacteria, Archaea and Protozoa divisions. For each division we propose sub-databases containing sequences of minimal lengths (500, 800, 1000, 1200 nt), since longer sequences are more appropriated to appreciate deeper branching. The databases often contain many sequences of the same species. Therefore, we built sub-databases including only 2 sequences per species, allowing, in each case, to produce a workable phylogenetic tree with proper outgroups. These two sequences are selected as being among the longest of each species.

Two different web interfaces are proposed: the NCBI BLAST default interface (Ye et al., 2006; McGinnis and Madden, 2004) and the ViroBLAST interface (Deng et al., 2007) which extends the utility of BLAST to query against multiple sequence databases and user sequence datasets. It also offers a friendly output to easily parse and navigate BLAST results. Due to the restricted amount of sequences in the database, a BLAST performs only in a few seconds (we have limited the max BLAST hits number at 500).

Recovering information and display using blast2tree

Recovering an informative taxonomy of a sequence is often very difficult using conventional BLAST servers. The definition line of the fasta format is too long and then often truncated by phylogeny programs such as Phylip (Felsenstein, 1989) or Clustal (Larkin et al., 2007). Moreover, this line contains no information concerning the taxonomic assignment. The online tool Blast2Tree intends to solve these problems. It helps to perform an easy retrieval of sequences and associated information. A simple copy/paste of any number of lines from the BLAST hits allows to get the corresponding SSU rRNA sequences (under two different formats) as well as the associated taxonomy. This can be done several times from different parts of a same BLAST result or from different queries. The final list can be

hand-edited and the choice can be visualized as a table with complete taxonomic assignments displayed. A phylogenetic tree can be drawn on screen and exported before or after data retrieval. This interface also allows to copy/paste data from Eztaxon.

Downloads of the following files are possible:

1. The exact SSU rRNA sequences (FASTA format), even when such sequences are embedded into a larger sequence (such as complete genomes). It is possible either to download sequences as selected from the BLAST result, or to download a clade of sequences (such as every sequence in a genus or a given species).
2. The annotations (tlf format). These annotations comprise the complete taxonomy description, the name of the strain if available and the notification of type strain. The tlf format is a simple ASCII file that can be used by softwares such as TreeDyn (<http://www.treedyn.org>) or ScripTree (<http://atgc.lirmm.fr/scriptree/>).
3. The full taxonomy (HTML format) is sorted under similar terms. The taxonomical terms can be retrieved from the pasted BLAST hits or from keywords entered into the form.
4. A phylogenetic tree (Newick format). Blast2Tree is able to automatically build a phylogenetic tree based on MAFFT software (Katoh et al., 2009). MAFFT is a program for multiple sequence alignment. It offers various multiple alignment strategies and we integrated the fastest one (although the less accurate) which is a simple progressive method like Clustal. The detailed algorithms are described by Katoh et al. (2009) (Katoh et al., 2002). The automatic tree building is also very fast and takes less than 40 seconds for 500 leaves, 7 seconds for 100 leaves, 3 seconds for 10 sequences. The user can also upload his own tree.
5. An image of a phylogenetic tree (jpeg, ps or tiff formats). Blast2Tree integrates the scripting language "ScripTree" to generate high-quality images of phylogenetic trees. A simple click is enough to display a downloadable image. This image includes the sequence to identify, the set of sequences selected from the BLAST results and the related annotations.

Results and Discussion

We have compared the results of the common BLAST servers (NCBI

>New.sequence

```
AACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGGCTGTGCCCTTGAAGCCT
GGCTTCCGGAGCTAACCGGTTAAGTCGACCGCCTGGGGAGTACGGCCGCAAGGTTAAAACCTCAAATGAAT
TGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTTAATTCGATGCAACGCGAAGAACCCTTACCTACTC
TTGACATCCAGAGAACTTAGCAGAGATGCTTTGGTGCCTTCGGGAACCTCTGAGACAGGTGCTGCATGGCT
GTCGTCAGCTCGTGTGTGAAATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATCCTTTGTGTCGA
GCGGTTCCGGCCGGAACTCATAGGAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAAGTC
ATCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATGGCGTATACAAAGAGAAGCGACCTCGCGAG
AGCAAGCGGACCTCATAAAGTACGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAA
TCGCTAGTAATCGTAGATCAGAATGCTACGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCAAG
CCATGGGAGTGGGTTGCAGAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCACCTTTGTGATTCA
TGACTGGGGTGAAGTCGTAACAAGGTAGCCGTA
```

Table 1: Sequence [EMBL: GU084214] described as an uncultured bacteria. This sequence is taken as an example of an "unknown sequence" to highlight the differences between the results of the usual online BLASTs (EBI, NCBI, DDBJ) and our BLAST server.



BLAST, EBI WuBlast2, DDBJ BLAST) and our web server. We took a given sequence as example, which is described as uncultured bacteria in the related NCBI entry (Table 1). The objective was to determine which known species are most related to the input sequence. For each server and similar options, the results are very different:

1. Most of the 100 similar sequences returned by the NCBI BLAST server are annotated as "Uncultured...". Recently, the NCBI BLAST added a feature to exclude the environmental sample sequences. If this option is enabled, the results become better, but are not suitable because of the amount of fuzzy descriptions (e.g. many descriptions with "Enterobacter sp.").
2. WU-Blast2 on EBI server with the database "embl release" gives results similar to NCBI. Improved results can be obtained by selecting the database "embl standard prokaryotes", the most frequent sequences retrieved being "Enterobacter sp" blended with various species. Although sequences from the environment are not included in this database, many sequences are still poorly described.
3. BLAST on DDBJ allows to choose a database of 16S rRNA sequences of prokaryotes, which behaves very similar to the "embl standard prokaryotes" database at EBI.
4. Concerning our server, we selected the database containing bacterial 16 rRNA and only 2 sequences for each species. The first results clearly show that the closest species are from the genera *Kluyvera* and *Enterobacter* (Table 2).

This example highlights that none of the common public BLAST servers are able to easily retrieve either a set of sequences with an informative and complete taxonomy or subsequences embedded in a larger genomic fragment. Only the use of a dedicated database

Sequences producing significant alignments	Score	E-value
AM933754 <i>Kluyvera cryocrescens</i>	1405	0.0
AM778415 <i>Enterobacter cloacae</i>	1398	0.0
AJ251468 <i>Enterobacter aerogenes</i>	1398	0.0
AM933757 <i>Kluyvera cryocrescens</i>	1394	0.0
AJ854062.16S-RRNA <i>Serratia ureilytica</i>	1392	0.0
AJ853891 <i>Enterobacter ludwigii</i>	1390	0.0
Z96078 <i>Enterobacter cancerogenus</i>	1382	0.0
DQ202394 <i>Enterobacter cloacae</i>	1382	0.0
AY825036 <i>Enterobacter aerogenes</i>	1382	0.0
AF025373 <i>Citrobacter werkmanii</i>	1382	0.0
AF025369 <i>Citrobacter murlinae</i>	1382	0.0
AF025368 <i>Citrobacter braakii</i>	1382	0.0
FJ662869 <i>Serratia nematodiphila</i>	1372	0.0
FJ462700 <i>Leclercia adecarboxylata</i>	1372	0.0
X93216 <i>Raoultella planticola</i>	1370	0.0
AF181574 <i>Raoultella planticola</i>	1370	0.0
EF175735 <i>Enterobacter ludwigii</i>	1370	0.0
DQ229104 <i>Citrobacter tnt5</i>	1368	0.0
FJ853424 <i>Serratia marcescens</i>	1366	0.0
AB364958 <i>Raoultella ornithinolytica</i>	1366	0.0
AJ251467 <i>Raoultella ornithinolytica</i>	1366	0.0
Z96077 <i>Enterobacter nimipressuralis</i>	1366	0.0
DQ294288 <i>Citrobacter freundii</i>	1366	0.0
DQ229103 <i>Citrobacter tnt4</i>	1366	0.0
EU221358 <i>Enterobacter asburiae</i>	1362	0.0
AF493976 <i>Cedecea davisae</i>	1362	0.0
EU914257 <i>Serratia nematodiphila</i>	1360	0.0
AM062693 <i>Enterobacter amnigenus</i>	1360	0.0
EF688006 <i>Pantoea punctata</i>	1358	0.0
AJ627202 <i>Kluyvera ascorbata</i>	1358	0.0
AJ627201 <i>Kluyvera ascorbata</i>	1358	0.0
AY567708 <i>Candidatus Cuticobacterium kirbyi</i>	1356	0.0

Table 2: Example of BLAST output from <http://bioinfo.unice.fr/blast>. The BLASTed sequence is [EMBL: GU084214] (the target sequence to identify) and the database used is "2 Seq/Bacteria". Only the first 33 most similar sequences are displayed.

returns meaningful BLAST results that are required for an easy identification of microbial rRNA sequences.

Figure 1 shows the phylogenetic tree built with Blast2Tree. This tree includes the sequences retrieved from our server and the sequence to identify. The default view displays the tree with the name of each species as leaves, the strain (with "T" as exponent, if this strain is a type strain) and the accession number. In this figure, the sequence to identify (named "New.sequence") is clearly related to the genus *Kluyvera* and the figure is almost ready for publication. The user can also refine the analysis by adding for example all sequences of *Kluyvera*.

Besides these common BLAST servers, some projects based on specific rRNA databases included tools which intend to help in determining the sequences of prokaryotes.

1. The Greengenes web application (DeSantis et al., 2006) gives access to a database of 16S rRNA aligned sequences (<http://greengenes.lbl.gov>). It allows to export, slice, browse and compare the sequences, and to search probes through many graphical tools. The data and tools presented by Greengenes aim to assist the researcher in choosing phylogenetically specific probes, interpreting microarray results, and aligning/annotating novel sequences. These tools also propose to compare a given sequence by alignment against the Greengenes database using a specific BLAST or a tool called "Simrank" based on shared 7-mers. They display an interactive table of BLAST or Simrank hits arranged by taxonomy.
2. The Ribosomal Database Project (RDP) (Cole et al., 2009) provides ribosome related databases and online tools for data analysis to the scientific community (<http://rdp.cme.msu.edu>). As of May 2010 (release 10.20), RDP maintains 1,237,963 aligned and annotated 16S rRNA sequences of Archaea and Bacteria, including sequences from cultured organisms and sequences obtained from environmental samples. Among the available services, RDB provides a classifier that assigns 16S rRNA gene sequences to the new phylogenetically consistent higher-order bacterial taxonomy using a naïve Bayesian classifier (Wang et al., 2007). This service allows to select a set of sequences with some relevant criteria such as sequences length or sequences from isolates only. Then, it is possible to build a crude online tree which is downloadable into the Newick format.
3. Silva is an online resource that provides comprehensive, quality checked and regularly updated databases of aligned 16S, 18S, 23S, 28S ribosomal RNA sequences for Bacteria, Archaea and Eukarya (Pruesse et al., 2007) (<http://www.arb-silva.de>). The project "The All-Species Living Tree" is associated with the Silva databases. This project aims to reconstruct a single 16S rRNA tree harboring all sequenced type strains of the hitherto classified species of Archaea and Bacteria (Yarza et al., 2008). Also, Silva proposes to do phylogenetic classification linked to the ARB software (Ludwig et al., 2004) (<http://www.arb-home.de>). An online form helps to align the sequence to identify with the closest sequences included into the database of Silva. A file is created (fasta+metadata or ARB file format) and can be used with the ARB software for further phylogenetic studies and for the visualization of the tree. The ARB software package has to be installed on the system and is not available for all operating systems (i.e. Windows). Note that an independent BLAST server exists which uses a part of Silva data (<http://www.sepsitest-blast.de>). Sepsitest has similarities with our Blast, but the database includes only sequences of type strains.



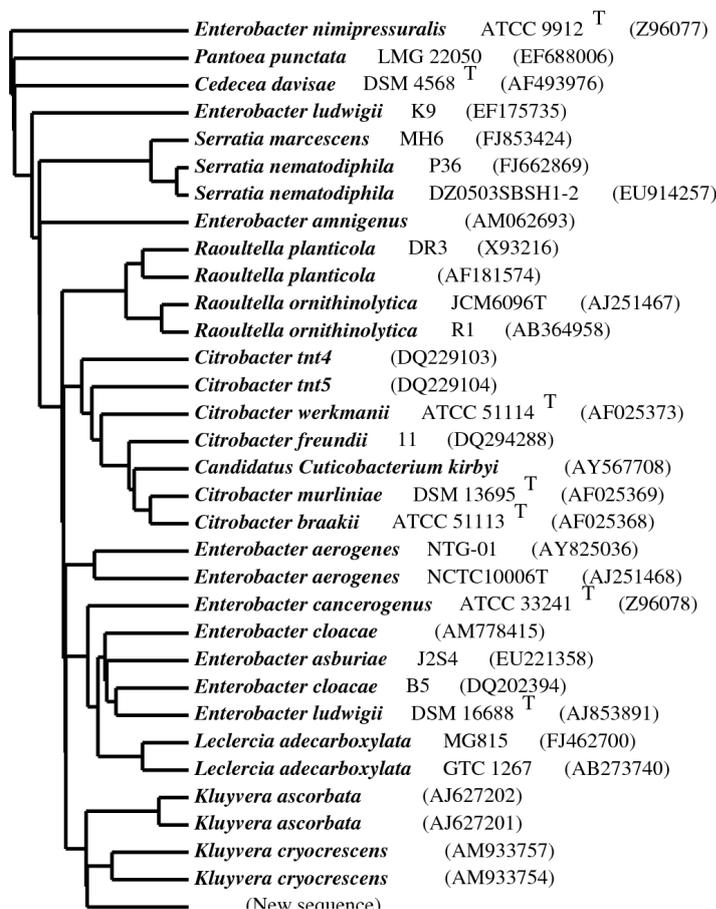


Figure 1: Phylogenetic tree built by Blast2Tree. It contains the sequence to identify (in this example we use [EMBL: GU084214] which is named "New.sequence" here) and the closed sequences from our server.

However, we also added the possibility to paste the output BLAST of SepstiTest to Blast2Tree.

The final phylogenetic tree may not include sequences that are not clearly named at the species level. Depending on the database provided by these described tools or depending of the ability of the user, it could be very tedious to avoid sequences annotated with some flooding terms such as 'Candidatus', 'unclassified' or 'environmental samples'. Moreover, the user can spend much time to obtain representative annotations in order to have a "ready to print" tree for publication. Finally, it is not easy to obtain a quick overview of a relevant set of closed sequences related to an unknown sequence to identify. Our web server is a tool intended to fill this gap.

Conclusions

Quality of analysis and display

The MAFFT algorithm used to build the trees is very fast but gives an average quality. No bootstrap confidence estimates or other statistical analysis are displayed. To perform a more robust tree we recommend to get sequences from Blast2tree and to build the tree by using classical phylogenetic methods. This procedure is longer but should provide better results. It is then possible to re-import the tree (in Newick format) into Blast2Tree. Concerning the final tree image, Blast2tree performs automatic procedures with the scripting facilities

of "ScripTree" (www.scripree.org). Some options allow to select the annotations to add onto the image. Also, it is possible to export the generated files from Blast2Tree to ScripTree software in order to obtain a more customized tree output. Another way to modify the tree consists in using the stand-alone software TreeDyn which is an advanced tool for the graphical management of trees.

Advantages of the system

The BLAST on main public servers are not appropriate for an accurate identification of prokaryotes, because it is not easy to find taxonomically meaningful relatives to a query sequence. The main advantages using a specific server are (i) to get clear and informative results (every hit will be from a well described species); (ii) a fast BLAST computing due to a low amount of data (the databases of sequences are much smaller than public databases); (iii) the possibility to export the true SSU rRNA sequences and taxonomic descriptions which facilitates the phylogenetic analysis and helps the user to integrate new sequences into a meaningful phylogenetic tree.

Existing tools based on specific rRNA databases offer many advantages, but they are not fully suitable for a fast and easy identification of an unknown prokaryote sequence that have to be included into an annotated tree. However, the high quality of the databases such as those provided by Silva or Greengenes could be used in the future in addition to the database used by our server.



Finally, the online tool Blast2Tree provides a fast and useful way to see how a new sequence is positioned within a tree of relevant closed sequences.

Availability and requirements

The server is freely accessible over the Internet at <http://bioinfo.unice.fr/blast/>. An online help is also available on <http://bioinfo.unice.fr/blast/blast2tree/help/>, it contains tutorials and examples of use. This server for microbial identification has been running for about one year and is used within the framework of a European project. Server and database are available for local installation, interested individuals are invited to contact the authors for more information.

Acknowledgements

This work was supported by funds from the European Commission for the HEALTHY WATER project (FOOD-CT-2006-036306) to R. Christen. The authors are solely responsible for the content of this publication. It does not represent the opinion of the European Commission. The European Commission is not responsible for any use that might be made of data appearing therein.

References

- Albuquerque P, Mendes MV, Santos CL, Moradas-Ferreira P, Tavares F (2009) DNA signature-based approaches for bacterial detection and identification. *Sci Total Environ* 407: 3641-3651.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
- Amann RI, Ludwig W, Schleifer KH (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59: 143-169.
- Chevenet F, Brun C, Bañuls AL, Jacq B, Christen R (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics* 7: 439.
- Chevenet F, Croce O, Hebrard M, Christen R, Berry V (2010) ScripTree: scripting phylogenetic graphics. *Bioinformatics* 26: 1125-1126.
- Christen R (2008) Identifications of pathogens - a bioinformatic point of view. *Curr Opin Biotechnol* 19: 266-273.
- Chun J, Lee JH, Jung Y, Kim M, Kim S, et al. (2007) EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int J Syst Evol. Microbiol* 57: 2259-2261.
- Clarridge JE 3rd (2004) Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin Microbiol Rev* 17: 840-862.
- Cole JR, Wang Q, Cardenas E, Fish J, Chai B, et al. (2009) The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37: D141-145.
- Deng W, Nickle DC, Learn GH, Maust B, Mullins JI (2007) ViroBLAST: a stand-alone BLAST web server for flexible queries of multiple databases and user's datasets. *Bioinformatics* 23: 2334-2336.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, et al. (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72: 5069-5072.
- Euzéby JP (2008) List of prokaryotic names with standing in nomenclature. Available at <http://www.bacterio.cict.fr/>
- Felsenstein J (1989) PHYLIP - Phylogeny Inference Package. *Cladistics* 5: 164-166.
- Hugenholtz P, Pace NR (1996) Identifying microbial diversity in the natural environment: a molecular phylogenetic approach. *Trends Biotechnol* 14: 190-197.
- Huse SM, Dethlefsen L, Huber JA, Welch DM, Relman DA, et al. (2008) Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. *PLoS Genet* 4: e1000255.
- Jonasson J, Olofsson M, Monstein HJ (2002) Classification, identification and subtyping of bacteria based on pyrosequencing and signature matching of 16S rDNA fragments. *APMIS* 110: 263-272.
- Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol* 537: 39-64.
- Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30: 3059-3066.
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.
- Lin YH, Chang BC, Chiang PW, Tang S (2008) Questionable 16S ribosomal RNA gene annotations are frequent in completed microbial genomes. *Gene* 416: 44-47.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* 32: 1363-1371.
- McCabe KM, Zhang YH, Huang BL, Wagar EA, McCabe ER (1999) Bacterial species identification after DNA amplification with a universal primer pair. *Mol Genet Metab* 66: 205-211.
- McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res* 32: W20-25.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, et al. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35: 7188-7196.
- Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AK, et al. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J* 1: 283-290.
- Stackebrandt E, Goebel BM (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.
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73: 5261-5267.
- Yarza P, Richter M, Peplies J, Euzéby J, Amann R, et al. (2008) The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. *Syst Appl Microbiol* 31: 241-250.
- Ye J, McGinnis S, Madden TL (2006) BLAST: improvements for better sequence analysis. *Nucleic Acids Res* 34: W6-9.

