Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Release LTPs104 of the All-Species Living Tree

Raúl Munoz [a,1], Pablo Yarza [a,*,1], Wolfgang Ludwig [b], Jean Euzéby [c], Rudolf Amann [d],
Karl-Heinz Schleifer [b], Frank Oliver Glöckner [d,e], Ramon Rosselló-Móra [a]

[a] Marine Microbiology Group, Department of Ecology and Marine Resources, Institut Mediterrani d'Estudis Avançats (CSIC-UIB), C/Miquel Marqués 21, E-07190 Esporles, Illes Balears, Spain
[b] Lehrstuhl für Mikrobiologie, Technische Universität München, D-85350 Freising, Germany
[c] Société de Bactériologie Systématique et Vétérinaire (SBSV) & École Nationale Vétérinaire de Toulouse (ENVT), F-31076 Tolouse cedex 03, France
[d] Max Planck Institute for Marine Microbiology, D-28359 Bremen, Germany
[e] Jacobs University Bremen, D-28759 Bremen, Germany

The Living Tree Project (LTP) team announces the first LTP release in 2011. The LTP database compiles the sequences of all validly named species with a high quality entry in the public repositories. The current release LTPs104 is the fifth update of the database that contains all new classified taxa with names validly published up to the December 2010 issue (vol. 60) of the International Journal of Systematic and Evolutionary Microbiology (IJSEM). The previous update of the LTP (LTPs102) contained type strain sequences with a valid name publication effective up to February 2010 (IJSEM 60(2)). During the intervening 10-month period, a total of 530 new species had been classified and their names validly published. Among them, 32 were new combinations of already classified taxa, and two lacked good quality 16S rRNA entries. Additionally, nine SSU entries were submitted after EMBL release 104 and will be included in the next LTP update. No further updating has been carried out on the 23S rRNA-based All-Species Living Tree due to the low number of newly sequenced type-strain LSU genes since LTP release 102.

The current release accounts for 8545 sequences of high quality. The sequences of the 545 newly classified species had been added to the LTP, and their alignments were improved for their genealogy reconstructions. All of them had been manually checked to emend possible errors or outdated information in their respective entries. Following previous releases [1,2] manual cross-check was undertaken for the sequence-associated information provided by SILVA and the up-to-date List of Prokaryotic Names with Standing in Nomenclature (LPSN, http://www.bacterio.cict.fr). It is remarkable that about 74% of the entries carried outdated information in important fields for taxonomy, such as the species name, and/or mistakes in their strain numbers (Table S1). Consequently, corrected and/or updated species names with valid publications have been written to the field "fullname_ltp" of the LTPs104 ARB database.

LTP provides a start-up tree for taxonomic calculations. The published tree provided in LTPs102 reflected the accepted global topology, since it was validated by means of extensive studies using type and non-type sequences and a large collection of supporting individual phylogenetic reconstructions [2]. For the current release, we preferred to maintain the tree topology from LTPs102, and all new sequences were added to it using the parsimony algorithm implemented in the ARB software package [3]. This tool allows local insertion of new sequences in a tree without global modification of the pre-existing topology. However, it is important to remark that the recognition of new unique branches that may indicate putative new taxa should not only be the result of a parsimony insertion, but should also be due to an accurate treeing approach.

Changes in the SSU dataset

IJSEM issues 60(3)–60(12) were tracked in order to build a list of new validly published species with their corresponding type strain ribosomal SSU sequence. From the 530 new species with validly published names that appeared from March to December 2010, 32 of them were new combinations of species already present in the tree. In such cases, only updates of their names were needed. Eleven species published during the update period had to be omitted due to various reasons: (I) nine species were submitted after SILVA launched its last release (SILVA 104), so their sequences are not available in the desired format: GU584097, *Meiothermus granaticius* (Albuquerque et al. 2010, sp. nov.); EU672803, *Flavobacterium glycines* (Madhaiyan et al. 2010, sp. nov.); FN386750, *Marmoricola scoriae* (Lee and Lee 2010, sp. nov.); EU672804, *Mucilaginibacter gossypii* (Madhaiyan et al. 2010, sp. nov.); EU672805, *Mucilaginibacter gossypiicola* (Madhaiyan et al. 2010, sp. nov.); EU672801, *Enterobacter arachidis* (Madhaiyan et al. 2010, sp. nov.); EU912487, *Microbacterium azadirachtae* (Madhaiyan et al. 2010, sp. nov.); GU065210, *Photobacterium jeanii* (Chimetto et al. 2010, sp. nov.); FM882231, *Dactylosporangium darangshiense* (Seo and Lee 2010, sp. nov.); (II) one species showed SSU sequence entries below the LTP quality criteria: FJ168539, with 1239 nucleotides, for *Pseudomonas chlororaphis* subsp. *piscium* (Burr et al. 2010, subsp. nov.); and (III) one species was considered 'orphan' (lack of SSU sequences in the

* Corresponding author.
  E-mail address: pablo.yarza@uib.es (P. Yarza).

[1] These authors have contributed equally to this work. Between 2007 and 2010, P. Yarza was in charge of the Living Tree Project as Editorial Assistant of the journal Systematic and Applied Microbiology (ELSEVIER GmbH). From January 2011, both scientific and editorial responsibilities have been transferred to R. Munoz.

public repositories): *Xanthomonas dyei* (Young et al. 2010, sp. nov.). Nevertheless, apart from these exceptions, a total of 58 species, omitted from the last LTP release, have now been included in the database, since they were either recently (re)sequenced and now provided high quality SSU entries, or their SSU sequences were submitted within the EMBL release 103 (hence they were missing during the creation of SILVA 102, but present in SILVA 104). Supplementary Table S2 summarizes the changes performed on the current release by means of additions, deletions or modifications in the dataset.

SSU-tree reconstruction

The LTPs102 dataset was used as the reference for the automatic alignment of the 545 new sequences using the SINA aligner (http://www.arb-silva.de [4]) followed by a manual curation on misplaced bases. The sequences were inserted in the LTPs102 tree using the ARB-parsimony tool implemented in the ARB software package (http://www.arb-home.de/ [3]). The whole SSU-based All-Species Living Tree of the current LTPs104 is represented in an 83 page pdf document (Fig. S1).

Features of the ARB database

The following seven fields have been implemented into the LTP database: (*fullname_ltp*), corrected species names according to LPSN (http://www.bacterio.cict.fr/) (*rel_ltp*), name of the LTP release where an entry appeared for the first time (*hi_tax_ltp*), name of the family of a taxon, except for unclassified genera which are explicitly labelled as 'Unclassified <taxon>' (*type_ltp*), type species receive the label 'type sp.' in this field (*riskgroup_ltp*), European Risk Group classification of micro organisms inherited from DSMZ (http://www.dsmz.de), according to the German BG Chemie List (TRBA 466 (list A), 2002); (*tax_ltp*), the last update of the accepted taxonomy according to LPSN (http://www.bacterio.cict.fr/classifphyla.html); (*url_lpsn_ltp*), contains the variable part of the URLs of LPSN's species files at http://www.bacterio.cict.fr.

Five filters (SAIs) of conserved positions have been created with the complete dataset of type strain SSU sequences in the LTPs104 release. During this process, alignment positions below 10%, 20%, 30%, 40% and 50% positional conservatory were removed from the alignment. The use of filters is a common practice to decrease the harmful effect of sequencing errors and areas where positional orthology cannot be unambiguously recognized. The number of allowed positions by each filter is, 1435, 1435, 1434, 1392 and 1290, respectively.

To obtain additional information for some SILVA and LTP fields, ARB users must perform some property adjustments in the *ARB_NT → properties → search world wide web (WWW)* window. In the first instance, a user may need to set up his/her web browser, which will be automatically launched from ARB. For Firefox users, they should fill in the bottom slot with the following command: *(firefox -remote 'openURL($(URL))' || firefox '$(URL)') &*. Some examples of ARB's web functionality are cited below:

I. to open LPSN files, type the following command into an empty slot: "*http://www.bacterio.cict.fr*";*readdb (url_lpsn_ltp)*
II. to open STRAININFO files type: "*http://www.straininfo.net/taxonGet.jsp?taxon=*"; *readdb(fullname_ltp);*"*&restrict=1*"
III. to open NCBI-taxonomy records type: "*http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=*"; *readdb(tax_xref_embl)*

"S.O.S." initiative, Sequencing the Orphan Species

Eleven international culture collections (DSMZ, CCUG, NBRC, CIP, LMG, CECT, ATCC, JCM, NCCB, ICMP and BZF) have launched a joint effort, coordinated by the LTP-team, to sequence the 16S rRNA gene of all orphan type strains (i.e. nearly 7% of all classified species with validly published names [2]). It is expected that we will soon complete the remaining 30% (i.e. 117 of 390 total reachable type strains in culture collections) of our task list in order to achieve the complete set of SSU sequences for all classified species of *Bacteria* and *Archaea* with validly published names.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.syapm.2011.03.001.

References

[1] Yarza, P., Richter, M., Peplies, J., Euzéby, J., Amann, R., Schleifer, K.-H., Ludwig, W., Glöckner, F.O., Rosselló-Móra, R. (2008) The All-Species Living Tree Project: a 16S rRNA-based phylogenetic tree of all sequenced type strains. Syst. Appl. Microbiol. 31, 241–250.
[2] Yarza, P., Ludwig, W., Euzéby, J., Amann, R., Schleifer, K.-H., Glöckner, F.O., Rosselló-Móra, R. (2010) Update of the all-species living tree project based on 16S and 23S rRNA sequence analyses. Syst. Appl. Microbiol. 33, 291–299.
[3] Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., Konig, A., Liss, T., Lussmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N., Vilbig, A., Lenke, M., Ludwig, T., Bode, A., Schleifer, K.-H. (2004) ARB: a software environment for sequence data. Nucleic Acids Res. 32, 1363–1371.
[4] Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glockner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35, 7188–7196.