

Compilation of tRNA sequences and sequences of tRNA genes

Mathias Sprinzl* and Konstantin S. Vassilenko¹

Laboratorium für Biochemie, Universität Bayreuth, 95440 Bayreuth, Germany and ¹Institute of Protein Research, Russian Academy of Sciences, 142290 Puschchino, Moscow Region, Russia

Received September 15, 2004; Accepted September 15, 2004

ABSTRACT

Maintained at the Universität Bayreuth, Bayreuth, Germany, the Compilation of tRNA Sequences and Sequences of tRNA Genes is accessible at the URL <http://www.tRNA.uni-bayreuth.de> with mirror site located at the Institute of Protein Research, Pushchino, Russia (<http://alpha.protres.ru/trnadbse>). The compilation is a searchable, periodically updated database of currently available tRNA sequences. The present version of the database contains a new Genomic tRNA Compilation including the sequences of tRNA genes from genomic sequences published up to July 2003. It consists of about 5800 tRNA gene sequences from 111 organisms covering archaea, bacteria, higher and lower eukarya. The former Compilation of tRNA Genes (up to the end of 1998) and the updated Compilation tRNA Sequences (561 entries) are also supported by the new software. The database can be explored by using multiple search criteria and sequence templates. The database provides a service that allows to obtain statistical information on the occurrences of certain bases at given positions of the tRNA sequences. This allows phylogenetic studies and search for identity elements in respect to interactions of tRNAs with various enzymes.

INTRODUCTION

The remarkable progress in large-scale automated DNA sequencing and the development of computer algorithms for fast automatic identification of the location of tRNA genes in sequenced genomes lead to the exponential increase of the available tRNA sequence information. To date, more than a hundred genomes from Eukarya, Bacteria and Archaea have been completely sequenced and the sequencing of many genomes is in progress. The comprehensive information on the

sequences, structure and occurrences of tRNA genes in different genomes allows reliable systematic analysis of phylogenetic dependencies, anticodon usage, characteristic structural features of tRNA, etc. In the light of this, the tRNA compilation was completely renewed. New version of the database now includes the separate collection of tRNA genes from published complete genomes. Thus the current compilation keeping all the advantages of the previous versions was extended to cover the needs of modern genomics.

DATABASE CONTENT AND ORGANIZATION

The new Compilation of tRNA Sequences and Sequences of tRNA genes contains in addition to 3279 sequences of the last edition from 1998 (1) the completely new Genomic tRNA Compilation including the sequences of tRNA genes from complete genomes published up to July 2003. The current database consists of three parts:

- (i) Genomic tRNA Compilation is a new addition to the database. This is the most complete compilation of the sequences of cytoplasmic tRNA genes derived from complete genome sequences included into DNA databases. Since sequences of tRNA genes originating from cellular organelles frequently cannot be processed to the general cloverleaf scheme, they were not included in the Genomic tRNA Compilation. There are specialized databases dealing with these sequences (2,3). Current Genomic tRNA Compilation consists of about 5800 tRNA gene sequences from 111 organisms covering archaea, bacteria, higher and lower eukarya. The database includes the tRNA gene sequences collected in GtRDB (4) as well as those from the additional complete genomes found in DNA databases. If the genomes of the different strains of the same organism were sequenced, the corresponding tRNA genes were added to the database independently.
- (ii) Compilation of tRNA Sequences, is a summary of tRNA sequences, including modified bases and references of the publications. The references are restricted to the first publication of the complete sequence unless additional

*To whom correspondence should be addressed. Tel: +49 921 552420; Fax: +49 921 552432; Email: Mathias.Sprinzl@uni-bayreuth.de

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use permissions, please contact journals.permissions@oupjournals.org.

information (e.g. base modification, corrections, etc.) was later obtained. In such cases, additional references were added. This compilation is updated up to December 2002. The table contains the known tRNA sequences of all organisms including organelles. This is the continuation of the original tRNA compilation first published in 1978.

- (iii) **Compilation of tRNA Genes**, is a summary of the sequences of tRNA genes published in the literature and databases up to the end of 1998. It contains tRNA genes of all organisms and organelles, but is not updated since January 1999. This table contains about 350 sequences of cytoplasmic tRNA genes that are not included in the Genomic tRNA Database. Most of the tRNA gene entries in this table have references of the publications in which the sequence was communicated.

The database is organized as MS Excel[®] workbooks. All the information collected is split into different indexed tables according to the type of data (specificity, sequence, organism, etc.) and the descriptions of certain genes are summarized in the main worksheet that includes the relations between the data tables. The information can be obtained by filling the query form that allows to enter the simple search criteria and to select the type of data to be displayed. The result of search is presented as a table containing the description of the sequences found. This includes amino acid specificity, anticodon sequence, organism name, strain, literature reference, PubMed ID, sequence, base-pairing and additional comments. Description of tRNA genes in Genomic tRNA Compilation also includes full organism taxonomy, position of the gene in the genome and index of general database record used as a source of the data.

In order to facilitate a computer analysis, an alignment of sequences is used, which is most compatible with the tRNA phylogeny and known three-dimensional structures of tRNA (5,6). The corresponding numbering system is described in (1). Positions in particular sequence which are not filled (gaps in the generalized structure) are indicated by a dash. The compilations use a one-letter code for all nucleotides including modified ones. To designate modified nucleotides, the other ASCII signs other than A, C, G, T and U are employed. Terminology and structure of the modified nucleosides occurring in tRNAs were used according to (7) and (8). All nucleotide insertions are commented and denoted by underlining at the place of insertion.

In addition to the plain text table one can explore the result of search by presenting the sequences in a cloverleaf form. It is possible to scroll the found sequences one by one or to select directly the sequence of interest from the result table. The presentation supports colour code for different structural features in the canonical cloverleaf model.

Simple statistical information on the occurrences of certain bases at given positions and the preferences in base-pairing also can be obtained on a special data sheet.

Compilation of tRNA Sequences and Compilation of tRNA Genes also exist in the classical tabular form described in (1). For user convenience, tables were converted to MS Excel[®] format.

ACCESS

The data are freely accessible for research purposes at <http://www.tRNA.uni-bayreuth.de> and <http://alpha.protres.ru/trnad-base>. This article should be cited in research projects assisted by the use of the compilation. Comments, corrections and new entries are welcome.

ACKNOWLEDGEMENTS

This project was supported by Fonds der Chemischen Industrie and Universität Bayreuth. We are grateful for advise, cooperation and help with data collection to Todd Michael Johnson Lowe, Genetics, Stanford University, California, Dr Carlos Hoyo-Vadillo, Departamento de Farmacología y Toxicología, Cinvestav, Mexico City, Mexico, Yvonne Baberowski, Mark Dürr and Bernhard Thielen, Universität Bayreuth.

REFERENCES

1. Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A. and Steinberg, S. (1998) Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.*, **26**, 148–153.
2. Rainaldi, G., Volpicella, M., Licciulli, F., Liuni, S., Gallerani, R. and Ceci, L.R. (2003) PLMitRNA, a database on the heterogeneous genetic origin of mitochondrial tRNA genes and tRNAs in photosynthetic eukaryotes. *Nucleic Acids Res.*, **31**, 436–438.
3. Helm, M., Brulé, H., Friede, D., Giegé, R., Putz, D. and Florentz, C. (2000) Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA*, **6**, 1356–1379.
4. Lowe, T.M. and Eddy, S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.
5. Schimmel, P.R., Söll, D. and Abelson, J.N. (eds) (1979) *Proposed Numbering System in tRNAs Based on Yeast tRNA^{Phe} in Transfer-RNA: Structure, Properties and Recognition*. Cold Spring Harbor Laboratory, pp. 518–519.
6. Steinberg, S.V. and Kisselev, L.L. (1992) Comparison of dissimilarity patterns of *E.coli*, yeast and mammalian tRNAs. *Biochimie*, **74**, 337–351.
7. Limbach, P.A., Crain, P.F. and McCloskey, J.A. (1994) Summary: the modified nucleosides of RNA. *Nucleic Acids Res.*, **22**, 2183–2196.
8. Crain, P.F. and McCloskey, J.A. (1997) The RNA modification database. *Nucleic Acids Res.*, **25**, 126–127.