154–155 *Nucleic Acids Research, 1998, Vol. 26, No. 1*

A FastA based compilation of higher plant mitochondrial tRNA genes

A. Sagliano, M. Volpicella, R. Gallerani and L. R. Ceci[1,*]

Dipartimento di Biochimica e Biologia Molecolare, Universita' di Bari, Via Amendola 165/A, 70126 Bari, Italy and [1]Centro di Studio sui Mitocondri e Metabolismo Energetico, CNR, Sezione di Trani, Via Corato 17, 70059 Trani, Italy

Received August 28, 1997; Revised and Accepted September 30, 1997

ABSTRACT

A new version of the compilation of higher plant mitochondrial tRNA genes (http://www.ebi.ac.uk/service) has been obtained by means of the FastA program for similarity searching in nucleotide sequence Databases. This approach improves the previous collection, which was based on literature data analysis. The current compilation contains 158 sequences with an increase of 43 units. In this paper, some interesting features of the new entries are briefly presented.

INTRODUCTION

tRNA molecules active in higher plant mitochondria are encoded by: (i) 'genuine' or 'native' mitochondrial genes derived from the original endosymbiont genome; (ii) 'chloroplast-like' genes present on chloroplast DNA regions which are thought to have been inserted in the mitochondrial genomes during their evolution; (iii) nuclear genes, in the case of imported tRNAs. In no case has redundancy of functional tRNAs of different genetic origin been observed (1).

In a previous compilation of higher plant mitochondrial tRNA genes (2) we reported 115 sequences in two different lists. The first contained complete gene sequences, the latter contained truncated or pseudo genes. In each list, the genes were further classified in the two classes 'genuine' and 'chloroplast-like'.

This compilation contains 158 sequences. The increase is due both to the availability of the complete *Arabidopsis thaliana* mitochondrial genome sequence (3) and also to the different approach used for sequence retrieval. While the first compilation was essentially based on analysis of the literature data, this compilation has been obtained by analysing nucleotide sequence databases with the FastA program (4). The program allows a reliable identification for any given sequence of the most similar sequences in the databases. tRNA genes reported in Tables 1 and 2 of the first compilation were used as query sequences. The analysis was performed directly at the European Bioinformatics Institute (EBI) World Wide Web server by means of the 'Similarity Searches' service (http://www.ebi.ac.uk/services/services.html). It was carried out against the whole EMBL Database in July 1997.

In this way several improvements have been obtained with respect to the previous compilation: (i) an increase in the number of sequences (eight more plants appear in the compilation); (ii) the cross-checking of the alignments among the genes; (iii) the identification of sequences containing tRNA genes not previously identified or searched for.

The sequence alignments for either the complete and truncated genes have been deposited as two separated lists (Lists 1 and 2) in the database 'PLMItRNA' at the European Bioinformatics Institute (EBI) Data Library. Information on how to obtain the compilation can be requested by electronic mail at the netserv@ebi.ac.uk address, by sending the command 'help PLMItRNA'. The compilation is also available via anonymous FTP from ftp.ebi.ac.uk in the directory pub/databases/PLMItRNA or from the EBI WWW server via the 'EBI Databases' service (http://www.ebi.ac.uk/service). A diskette version is also available from the authors.

DESCRIPTION OF THE COMPILATION AND DISCUSSION

Sequences of tRNA genes have been reported in the same way used in the previous collection (2): List 1 contains sequences of complete genes, List 2 contains sequences of truncated or pseudo genes. In each list the genes have been further classified as 'genuine' or 'chloroplast-like' and their sequences subdivided into the classical tRNA domains. Nucleotide positions have been numbered from 0 to 76. Positions of nucleotides which are not always present (0, 17, 17a, 20a, 20b, the nucleotides of the extra-loop, 47 and 74–76) are indicated by dashes. The genes are indicated by capital letters corresponding to the one letter code of the charged amino acid. Anticodons have been reported in small letters. Initials of the Latin name indicate the plant species: A.t., *Arabidopsis thaliana* (thale cress); B.v., *Beta vulgaris* (sugar beet); B.n., *Brassica napus* (rapeseed); B.o., *Brassica oleracea* (cauliflower); G.m., *Glycine max* (soybean); H.a., *Helianthus annuus* (sunflower); L.l., *Lupinus luteus* (lupine); L.e., *Lycopersicon esculentum* (tomato); O.b., *Oenothera berteriana* (primrose); P.h., *Petunia hybrida* (petunia); P.v., *Phaseolus vulgaris* (bean); P.s., *Pisum sativum* (pea); R.s., *Raphanus sativus* (horseradish); S.t., *Solanum tuberosum* (potato) for the dicotyledons, and L.m.. *Lolium multiflorum* (Italian ryegrass); Lo.e., *Lophopyrum elongatum* (tall wheatgrass); O.s., *Oryza sativa* (rice); S.b., *Sorghum bicolor* (sorghum); S.c., *Secale cereale* (rye); S.v., *Sorghum vulgare* (sorghum); T.a., *Triticum aestivum* (wheat); Z.m., *Zea mays* (maize) for the monocotyledons. La.l., *Larix leptoeuropaea* (larch) for the sole gymnosperm. The reference for each sequence is reported, when possible, as its DDBJ/EMBL/GenBank Database accession number.

Since in angiosperms a different distribution of genuine and chloroplast-like genes is observed between monocotyledonous

*To whom correspondence should be addressed. Tel: +39 80 544 3311; Fax: +39 80 544 3317; Email: luigi.ceci@area.ba.cnr.it

and dicotyledonous plants, sequences belonging to dicotyledons are reported separately from those of monocotyledons. For each gene a 'd' or 'm' symbol indicates the first of the dicots or monocots respectively. A 'g' indicates the sole gymnosperm in the case of the genuine *trn*H gene.

When redundancy of sequences is found, only the accession number of the longest sequence is reported, while shorter sequences are mentioned in brackets in the footnotes.

A + or – sign in List 1 indicates the result of transcription analysis when available.

Discrepancies between database sequences and literature data reported in the previous compilation are no longer reported.

In the course of this survey several points of interest for further studies and discussion on the organization of tRNA genes in higher plant mitochondria emerged.

The FastA analysis allowed the identification of complete or truncated tRNA genes which were not identified in the original sequences. These genes have been reported in Table 1 together with the accession numbers of the original sequences and the coordinates of their position. These results increase even more the number of the truncated genes, which appear to be particularly diffuse in higher plant mitochondrial genomes. They might be a consequence of the numerous rearrangements which occurred during the evolution of these complex genomes.

Portions of an original genuine *trn*P-(cgg) gene have been revealed in the dicotyledonous *A.thaliana*, *B.napus* and *O.berteriana* (Table 1). This finding seems particularly interesting since tRNA-Pro(cgg) molecules or *trn*P-(cgg) genes have never been described in mitochondria of any eukaryotes. It is not possible from the data currently available to ascertain whether the presumable *trn*P-(cgg) genes might have coded for functional tRNA-Pro(cgg) molecules or for post-transcriptional edited tRNA-Pro(ugg) molecules. Editing in the anticodon sequence has never been described in plant mitochondria, but has been reported for marsupial mitochondria (5).

Several modifications in the anticodon sequence of the genuine pseudo *trn*F genes reported in List 2 have been described that would change the tRNA specificity. Anyway these sequences have not been reported as different genes owing to the few differences that they show with respect to the original *trn*F gene. On the contrary four *trn*Y sequences can be detected in the *A.thaliana* mitochondial genome (cfr. the *A.thaliana trn*Y genes in List 2 and related notes) which, even if they are thought to derive from an original *trn*F gene, show relevant modifications in the acceptor stem sequences.

The FastA analysis was carried out against the whole EMBL Database Library. It is worthwhile to note that not all the sequences reported in the compilation can be found in the EMBL 'Organelle Library'. Some sequences obtained in the course of the development of sequencing projects for the *A.thaliana* genome correspond in fact to mitochondrial sequences. The *A.thaliana* genomic sequence B08367 is almost identical to a region of the *A.thaliana* mitochondrial sequence Y08502 (27 392–28 142). It contains two tRNA genes: the genuine *trn*E gene (coordinates 306–377) and a pseudo *trn*Y gene (coordinates 193–267). One *A.thaliana* EST (Expressed Sequence Tag) sequence, Z17763, the function of which was not previously described, contains part of the genuine mitochondrial *trn*S(gct) gene and 270 nt of its 5′ flanking region. This gene is present in double copy in the *A.thaliana* mitochondrial genome (3) with the two copies of the gene sharing a 5′ flanking region of 22 nt. These data indicate that at least one copy of the *trn*S gene is transcribed.

Table 1. tRNA genes not identified in the original sequences.

Complete genes		
Genuine		
*trn*F(gaa)	*P.sativum*	X14409: 124–196
Chloroplast-like		
*trn*C(gca)	*Z. mays*	M74160: 2030–2100
trnH(gtg)	*L.multiflorum*	D28336: 421–495
*trn*V(gac)	*H.annuus*	X82386: 3226–3297[a]
Truncated genes		
Genuine		
*trn*I(cat)	*A.thaliana*	Y08502: 144 612–144 639
*trn*fM(cat)	*A.thaliana*	Y08501: 93 352–93 422[a]
*trn*fM(cat)	*B.napus*	S47089: 133–155
*trn*P(cgg)	*A.thaliana*	Y08502: 33 112–33 218
*trn*P(cgg)	*B.napus*	D13699: 3819–3924
*trn*P(cgg)	*O.berteriana*	X69555: 2576–2688
*trn*P(tgg)	*P.sativum*	X14409: 473–538
*trn*Q(ttg)	*A.thaliana*	Y08501: 184 527–184 575[a]
*trn*Y(gta)	*A.thaliana*	Y08502: 62 130–62 210[a]
*trn*Y(gta)	*A.thaliana*	Y08501: 191 953–192 025
*trn*Y(gta)	*A.thaliana*	Y08501: 12 147–12 204
*trn*Y(gta)	*A.thaliana*	Y08502: 133 650–133 731[a]
Chloroplast-like		
*trn*M(cat)	*B.napus*	U10428: 1–42
*trn*R(acg)	*Z.mays*	M36716: 1773–1839
*trn*R(acg)	*Z.mays*	M12582: 2296–2364
*trn*R(acg)	*Z.mays*	X06667: 742–804

[a]Coordinates allow the identification of the gene on the complementary strand.

REFERENCES

1 Maréchal-Drouard,L., Weil,J.-H. and Dietrich,A. (1993) *Annu. Rev. Plant Physiol. Plant Mol. Biol.*, **44**, 13–32.
2 Veronico,P., Gallerani,R. and Ceci,L.R. (1996) *Nucleic Acids Res.*, **24**, 2199–2203.
3 Unseld,M., Marienfeld,J.R., Brandt,P. and Brennicke,A. (1997) *Nature Genet.*, **15**, 57–61.
4 Pearson,W.R. and Lipman,D.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
5 Morl,M., Dorner,M. and Paabo,S. (1995) *Nucleic Acids Res.*, **23**, 3380–3384.