PLMItRNA, a database for higher plant mitochondrial tRNAs and tRNA genes

Luigi R. Ceci*, Mariateresa Volpicella[1], Sabino Liuni, Vito Volpetti[1], Flavio Licciulli[2] and Raffaele Gallerani[1]

Centro di Studio sui Mitocondri e Metabolismo Energetico, CNR, Via Amendola 165/A, 70126 Bari, Italy, [1]Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, 70126 Bari, Italy and [2]Area di Ricerca, CNR, 70126 Bari, Italy

ABSTRACT

The PLMItRNA database contains information and multialignments of tRNA genes and molecules detected in higher plant mitochondria. It has been developed from a previous compilation of higher plant mitochondrial tRNA genes [Sagliano,A., Volpicella,M., Gallerani,R. and Ceci,L.R. (1998) *Nucleic Acids Res.,* **26, 154–155] and implemented with data and sequences of tRNA molecules retrieved from the literature. The current version of the database reports information on 171 genes and 16 tRNA molecules from 24 plants. PLMItRNA is accessible via WWW at http://bio-www.ba.cnr.it:8000/srs/**

INTRODUCTION

PLMItRNA is a database developed to facilitate retrieval of information on the organization of tRNA molecules and genes in higher plant mitochondria.

More than 15 years of studies have revealed a complex and still unexplainable distribution of tRNA genes and molecules with different genetic origin in higher plant mitochondria. tRNAs active in these organelles are the result of a selective recruitment of genes and molecules occurred in the precursor of the present-day mitochondrion which also involved the chloroplast and the cytosol of the host cell. Three classes of tRNAs with different genetic origin can be distinguished in higher plant mitochondria: (i) 'genuine' or 'native' mitochondrial tRNAs, encoded by genes derived from the original endosymbiont genome; (ii) 'chloroplast-like' tRNAs, encoded by genes present in chloroplast DNA insertions in the mitochondrial genomes; (iii) cytosolic tRNAs, encoded by nuclear genes and imported into the mitochondrion. The overall scenario is further complicated by the different distribution among plants of tRNAs belonging to the different classes.

The PLMItRNA database has been developed from a previous compilation of higher plant mitochondrial tRNA genes (1). It was updated in August 1998 by analysing the whole EMBL database (release 55) with the FastA program (2). The analysis was performed directly at the European Bioinformatics Institute (EBI) World Wide Web server by means of the 'Similarity Searches' service at the address http://www2.ebi.ac.uk/fasta3/

In addition to the gene sequences present in the previous compilation, the current version of the PLMItRNA database also contains sequences of higher plant mitochondrial tRNAs. Several papers have been published so far concerning the sequences of tRNA molecules in higher plant mitochondria. Besides the first papers which date from the early '80s, the interest in such studies has been raised in the last decade since the discovery of the import of tRNAs in higher plant mitochondria (3) and of RNA editing for the tRNA molecules (4). Updating of PLMItRNA for the tRNA sequences has been based on the analysis of the literature since many of their sequences have not been deposited in the nucleic acids databases.

DESCRIPTION OF THE DATABASE

Each entry in the database contains an informative and a multialignment section (Fig. 1). In the first part different characteristics of the tRNA molecule or gene have been reported. In the multialignment section all the homologous sequences detected in higher plant mitochondria have been aligned.

The informative section slightly differs in the cases of tRNA genes or molecules. In the case of genes the information on 'Gene structure' became necessary to distinguish between the complete genes and the several pseudo or truncated genes which have been detected. These sequences appear to be particularly diffuse in higher plant mitochondrial genomes as a probable consequence of the numerous rearrangements which occurred during the evolution of these complex genomes.

In the case of tRNA sequences we report as specific entries only sequences obtained by direct RNA sequencing or cDNA sequences for which the corresponding gene was not sequenced, while cDNA sequences of genes already sequenced are mentioned in the 'Note' section of the gene entry. When possible tRNA sequences have been aligned under their gene sequence. Symbols of modified nucleotides in the tRNA sequences are according to ref. 5 with the exception of the Å character that represents either i^6A or $ms^2 i^6A$ modified nucleotides, and for the £ character that represents a lysidine-like hypermodified nucleotide.

In the multialignments, sequences have been subdivided into the classical tRNA domains. Nucleotide positions have been numbered from 0 to 76. Positions of nucleotides which are not

---

*To whom correspondence should be addressed. Tel: +39 080 544 3311; Fax: +39 080 544 3317; Email: luigi.ceci@area.ba.cnr.it

```
ID:              16
Molecule:        DNA
Gene structure:  Complete gene
Genetic origin:  Mitochondrial
Plant group:     Viridiplantae; Charophyta/Embryophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Magnoliopsida.
Plant:           Bean (Phaseolus vulgaris)
Amino acid:      Phenylalanine
Anticodon:       GAA
Transcription:   +
Link:            EMBL;X73285
Note:            The gene has also been sequenced as cDNA. C to U editing for nucleotide C4 has been observed. Marechal-Drouard L.,
                 Ramamonjisoa D., Cosset A., Weil J.H. and Dietrich A. (1993) Nucleic Acids Res. 21: 4909-4914.
Multialignment:
            Acceptor |  |            D-domain          |    | Anticodon domain       |  |   Variable region      |    |    T-domain    |Acceptor|
            stem     |  |stem|  loop           |stem| stem | loop  | stem  |  |  extra arm            |    |stem| loop  | stem| stem   |
            .        |  |.   |  --    .--     |    |    .  |       |  .    |  |------------------|  - |.   |    .|     |   .    |---
            0           10         20            30         40              |                      |        50         60         70    76

   13 H.a.   D -GTTCAGG|TA|GCTC|AGCT-GGTT-A|GAGC|AAAGGA|CTGAAAA|TCCTTG|T|------------------|GTC|AGTGG|TTCGAAT|CCACT|TCTAAGCG|---
   14 O.b.   D -.......|..|....|....-....-.|....|......|.......|......|.|------------------|...|.....|.......|.....|........|---
   15 P.s.   D -.......|..|....|....-....-.|....|......|.......|......|.|------------------|...|.....|.......|.....|........|---
   16 P.v.   D -.......|..|....|....-....-.|....|......|.......|......|.|------------------|...|.....|.......|.....|........|---
  160 P.v.   R -...U...|..|....|...D-#.XD-.|....|......|......-A|......|.|------------------|7X.|.....|.P.....|.....|........|CCA
   17 S.t.   D -.......|..|....|...D-..DX-.|....|......|......-A|......|.|------------------|...|.....|.......|.....|......A|---
  161 S.t.   R -...U...|..|....|...D-..DX-.|....|......|......-A|......|.|------------------|7X.|.....|.P.....|.....|......A|CCA
//
```

Figure 1. Example of a complete entry for a tRNA gene.

Table 1. Plant taxonomic groups present in the PLMItRNA database

1.	Viridiplantae
1.1.	Charophyta/Embryophyta
1.1.2.	Embryophyta (Higher plants)
1.1.2.3.	Trachephyta (Vascular plants)
1.1.2.3.1.	Spermatophyta (Seed plants)
1.1.2.3.1.1.	Conipherophyta
1.1.2.3.1.1.1.	Coniferopsida
1.1.2.3.1.2.	Magnoliophyta=Angiospermae (Angiosperms, Flowering plants)
1.1.2.3.1.2.1.	Liliopsida=Monocotyledonae (Monocots, Monocotyledons)
1.1.2.3.1.2.2.	Magnoliopsida=Dicotyledonae (Dicots, Dicotyledons)

Table 2. Plants present in the PLMItRNA database

	Latin name	English name	PLMItRNA code
Angiosperms			
dicotyledons	*Arabidopsis thaliana*	thale cress	A.t.
	Beta vulgaris	sugar beet	B.v.
	Brassica napus	rapeseed	B.n.
	Brassica oleracea	cauliflower	B.o.
	Glycine max	soybean	G.m.
	Helianthus annuus	sunflower	H.a.
	Lupinus luteus	lupine	L.l.
	Lycopersicon esculentum	tomato	L.e.
	Oenothera berteriana	primrose	O.b.
	Petunia hybrida	petunia	P.h.
	Phaseolus vulgaris	bean	P.v.
	Pisum sativum	pea	P.s.
	Raphanus sativus	horseradish	R.s.
	Solanum tuberosum	potato	S.t.
	Vigna radiata	mung bean	V.r.
monocotyledons	*Lolium multiflorum*	Italian ryegrass	L.m.
	Lophopyrum elongatum	tall wheatgrass	Lo.e.
	Oryza sativa	rice	O.s.
	Secale cereale	rye	S.c.
	Sorghum bicolor	sorghum	S.b.
	Sorghum vulgare	sorghum	S.v.
	Triticum aestivum	wheat	T.a.
	Zea mays	maize	Z.m.
Gymnosperms			
	Larix leptoeuropaea	larch	La.l.

always present (0, 17, 17a, 20a, 20b, nucleotides in the extra-arm, 47, and 74–76) are indicated by dashes. Each sequence is also preceded by some informative characters: the ID number, the initials of the Latin name of the plant species, a D or R letter for DNA or RNA sequences, respectively. In each multialignment, dicotyledonous plants have been aligned before monocotyledons. This allows easy comparison of possible nucleotide differences between the two classes.

The database allows the selection of a list of entries which satisfy the characteristics specified in the query form. A selection by single plant name (English or Latin name) or plant taxonomic groups is also possible. Groups are according to the ACNUC retrieval program (6). For the sequences currently available each of the taxonomic groups reported in Table 1 can be selected. In Table 2 Latin and English names of plants present in the database are listed.

It is also possible to select entries by indicating specific keywords in the 'All text' field of the query form. A keyword is any word present in the 'Note' line of the entries. We have introduced standard sentences to allow the retrieval of entries with common features. For example, the keyword 'not described' allows the retrieval of genes that have not been described by authors in the primary database accession numbers, and that we have detected by means of the FastA analysis (28 sequences). Sequences that are not present in nucleic acids libraries (five genes and 12 tRNAs) can be selected by the keyword 'not available'. Other useful keywords can be 'cDNA' and 'RNA editing'.

When possible, entries contain links to the EMBL/GenBank libraries accession numbers.

DATABASE AVAILABILITY AND CITATION

Access to the PLMItRNA database is possible by World Wide Web at the address http://bio-www.ba.cnr.it:8000/srs/

The collection of sequences present in PLMItRNA is also available for database searching at the address http://bio-www.ba.cnr.it:8000/BioWWW/fasta.htm . The FastA output file is designed in order to allow links with the PLMItRNA entries.

Users of the database should cite the present publication as reference. Comments, corrections and new entries are welcome.

REFERENCES

1 Sagliano,A., Volpicella,M., Gallerani,R. and Ceci,L.R. (1998) *Nucleic Acids Res.*, **26**, 154–155.
2 Pearson,W.R. and Lipman,D.J. (1988) *Proc. Natl Acad. Sci. USA*, **85**, 2444–2448.
3 Dietrich,A., Weil,J.H. and Maréchal-Drouard,L. (1992) *Annu. Rev. Cell. Biol.*, **8**, 115–131.
4 Dietrich,A., Small,I., Cosset,A., Weil,J.H. and Maréchal-Drouard,L. (1996) *Biochimie*, **78**, 518–529.
5 Sprinzl,M., Horn,C., Brown,M., Ioudovitch,A. and Steinberg,S. (1998) *Nucleic Acids Res.*, **26**, 148–153.
6 Gouy,M., Gautier,C., Attimonelli,M., Lanave,C. and Di Paola,G. (1985) *Comput. Applic. Biosci.*, **1**, 167–172.