

Chapter 27

PtRFdb: Plant tRNA-Derived Fragments Database



Shafaque Zahra and Shailesh Kumar

27.1 Introduction

The emergence of fast, accurate, and cost-effective high throughput Next Generation Sequencing (NGS) technology has significantly paced up the exploration of the immense repertoire of small non-coding RNA molecules (Vickers et al. 2015). The rich plethora of small RNA molecules with diverse biological activities are produced in varied organisms, in various tissues, under different conditions, and during varied stages of development. Amongst them, microRNAs (miRNAs) and small interfering RNAs (siRNAs) have been extensively studied and proved to play a significant role in different aspects of gene expression together with gene regulation. Transfer RNA-derived fragments or tRFs with length ranging from 15 to 28 nucleotides (nt), being similar to miRNAs, both structurally and functionally, has been detected in diverse species ranging from the most primitive Archaeobacteria to the most evolved human beings (Keam and Hutvagner 2015). It is reported that tRFs are present in abundance and stand next to miRNAs in the small RNA pool.

Apart from the canonical role of as adapter molecules of tRNAs during protein translation, it has been revealed that tRNAs are also shown to be involved in the regulation of cellular functionalities and metabolism (Orioli 2017). This has paved way for exploring new dimensions in tRNA biology and their implications in cellular physiology and disease. With the augmentation of the omics era, decreased cost of sequencing, and abundance of sequenced transcriptomic data, an insight into tRNA dynamics has revealed that tRNAs are also involved in generation of a novel class of small non-coding RNAs by undergoing endonucleolytic cleavage at specific positions, called tRNA-derived fragments or tRFs (Lee et al. 2009). These novel cleavage products with length ranging from 15 to 28 nucleotides have been detected in both prokaryotes and eukaryotes (Sablok et al. 2017). Some specific tRFs

S. Zahra · S. Kumar (✉)

National Institute of Plant Genome Research (NIPGR), New Delhi, India

e-mail: shailesh@nipgr.ac.in

© Springer Nature Singapore Pte Ltd. 2019

S. Kumar, C. Egbuna (eds.), *Phytochemistry: An in-silico and in-vitro Update*,

https://doi.org/10.1007/978-981-13-6920-9_27

511

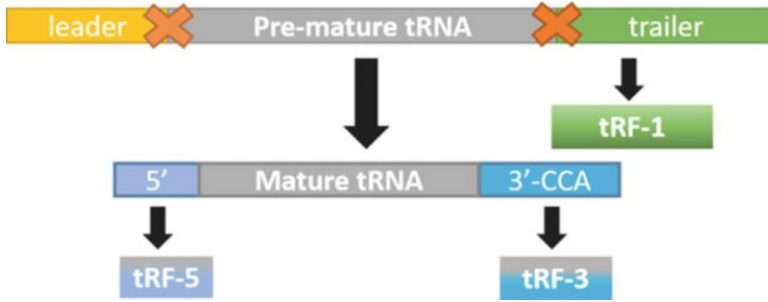


Fig. 27.1 Cleavage of pre-mature and mature tRNA for tRFs generation

overexpression in different human has unveiled their role in cancer progression and altering cellular dynamics in other organisms as well under different stress conditions (Sun et al. 2018). Although well studied in humans and other organisms, some reports in plants also confirm that tRFs are associated with Argonaute (AGO) proteins and involved in gene expression regulation under various abiotic as well as biotic stresses (Loss-Morais et al. 2013). Being functionally similar to microRNAs, recently the study of tRFs has gained momentum in plants as some tRFs have been proven to be differentially over-expressed during abiotic and biotic stresses.

In diverse organisms including plants, tRFs are synthesized by enzymatic cleavage of tRNAs in a position-specific fashion and depending on the cleavage position of tRNAs, they are classified into three types: tRF-5, tRF-3, and tRF-1 (Lee et al. 2009). The tRF-5s and tRF-3s are generated from mature tRNA from 5' and 3' ends respectively, while tRF-1s are synthesized from 3'-trailer sequences of pre-mature tRNAs (Fig. 27.1), which can be mediated by Dicer-like (DCL) proteins or by some DCL-independent process which is yet to be elucidated. The ribonucleases involved in tRF generation are not established yet but RNS1 is speculated to be responsible for tRF synthesis in the model plant, *Arabidopsis thaliana* (Alves et al. 2017). There is also evidence of the existence of organellar (mitochondrial and plastidial) tRFs from the previous literature but their potential functions have not been revealed yet (Cognat et al. 2017).

It has now been established that tRFs have a differential accumulation in different plants under various abiotic and abiotic stresses. They are believed to be acting as stress-related gene expression regulators. Specific tRFs viz. Ala^{AGC}, Arg^{CCT}, Arg^{TCG}, and Gly^{TCC} were reported to be overexpressed in drought and salt stress, Val^{CAC}, Tyr^{UGU}, Thr^{GUA}, and Ser^{UGA} during heat stress, Asp^{GTC} and Gly^{TCC} during phosphate deficiency, Arg^{CCT} in cold stress, Arg^{TCG} and Tyr^{GTA} during oxidative stress, Ile^{AAT}, Arg^{ACG}, and Ala^{CGC} during pathogen infection (Hsieh et al. 2009; Loss-Morais et al. 2013; Wang et al. 2016a, b; Alves et al. 2017).

tRFs are also been associated with some AGO proteins like AGO1, AGO2, AGO4, and AGO7 during stress conditions (Sablok et al. 2017). It has been discovered that tRFs also interfere with ribosomal proteins and affect translational activity. Apart from translational repression, the tRFs affect genome stability by

governing the post-transcriptional activity of retrotransposons (Martinez et al. 2017). Some recent research suggests that tRFs facilitate in the root nodule formation as well as aid arbuscular mycorrhiza growth in leguminous plants (Jin et al. 2018).

For further exploration of the domain of tRFs, some currently available web-based portals are MINTbase (Pliatsika et al. 2016), tRFdb (Kumar et al. 2015), tRex (Thompson et al. 2018) and tRF2Cancer (Zheng et al. 2016). However, very little information is available for plant tRFs with the exception of tRex which is web-portal holding information about tRFs detected in *Arabidopsis* only. This chapter discusses about our recently developed database, 'PtRFdb' (www.nipgr.res.in/PtRFdb) containing complete information of tRFs detected in 10 evolutionarily dissimilar and diversified plant species (Gupta et al. 2018). This database is believed to be highly resourceful for gaining numerous useful information about different tRF types in diverse plants species. PtRFdb will be useful to elucidate new pathways of gene expression regulation in plant genomics and better comprehensive understanding of the cross-talks between other small non-coding RNAs and their downstream target molecules.

27.2 Materials

The tRNA genes of the ten plant species viz. *Physcomitrella patens* (Version 1.1), *Brachypodium distachyon* (JGI v1.0 8X), *Populus trichocarpa* (January 2010 Version 2.0), *O. sativa* (v7.0), *Sorghum bicolor* (Version 1.0), *M. truncatula* (March 2009 Version 3.0), *A. thaliana* (TAIR10 February 2011), *Glycine max* (Wm82.a2), *Vitis vinifera* (Grapevine 12X) and *Zea mays* (Version 5b.60) were downloaded from GtRNAdb (Chan and Lowe 2009). Additionally, the reference genomes of the plants were also fetched from their respective genomic portals. Thus, in the FASTA format, the sequences of tRNA genes for each of the considered plant species were extracted as per the strand information. For generating pre-tRNAs, we extracted the sequences 40 nt upstream and downstream at the terminal ends of the mature tRNA genes. As mature RNA contains 'CCA' at the 3'-end, CCA was added to tRNA sequences obtained from tRNAscan (Lowe and Eddy 1997). By combining pre-mature and mature tRNA sequences, a reference database was created for each plant species using the option 'makeblastdb' script of BLAST (Basic Local Alignment Search Tool) (Altschul et al. 1990). This reference database was utilized for the prediction of three different tRF types (tRF-5s, tRF-3s, and tRF-1s).

27.3 Methods

27.3.1 Data Procurement

Datasets constituting small RNA sequencing reads and small RNA/microRNA data comprising of unique sequences with clonal frequency were downloaded from

NCBI-SRA (<https://www.ncbi.nlm.nih.gov/sra>) and NCBI-GEO (<http://www.ncbi.nlm.nih.gov/geo/>) respectively. This data was further processed for the identification of tRFs.

27.3.2 tRFs Identification

The fragments with the length ranging from 15 to 28 nt, with a clonal frequency greater than 9 (>9) were selected and BLASTN was performed against reference database aforementioned. Only those reads were considered that mapped along 100% length to the database for further study. Similarly, raw reads were processed by using stringent filters as described in our published research cited earlier. Reads were also filtered using tDRmapper software (Selitsky and Sethupathy 2015) and only reads with a quality score of >28 were accepted. Further, for eliminating false positive reads, reads with identity equal to 100% and without a gap (0%) were selected. The reads length ranging from 15 to 28 nt were selected for incorporation in PtRFdb.

27.3.3 The Web Interface of PtRFdb

After the collection and compilation of all the information, our PtRFdb database was developed on an Apache Hypertext Transfer Protocol (HTTP) Server together with MySQL at the backend, providing commands for data storage and retrieval into the database. Hypertext Markup Language or HTML, JavaScript, in addition to Hypertext Pre-processor (PHP) for front-end of web interface designing. PHP and PERL languages were used for writing in-built scripts. The home page of PtRFdb is represented in Fig. 27.2.

27.3.4 PtRFdb Features and Tools

PtRFdb holds detailed information related to tRFs identified in different plants. It holds data at two levels- primary and secondary level. At the primary level, basic information pertaining to each tRF entry like tRF type, tRNA name, gene coordinates, plant, tissue, PubMed ID, anticodon of corresponding tRNA and GSM number are provided. At the secondary level, sequence length, mapping position, relevant frequency, publication and sequencing study was incorporated. For ease in the retrieval of information, we have provided user-friendly search modules: 'Basic search', 'Advanced search' and 'Browse' (Figs. 27.3, 27.4, and 27.5). For each query, up to a maximum of ten different fields can be displayed. The columns for 'GSM number', 'Sequence', and 'PMID' are again linked with their parallel information related to experimental details, research publications, and sequence details as highlighted in the search result of Fig. 27.4.

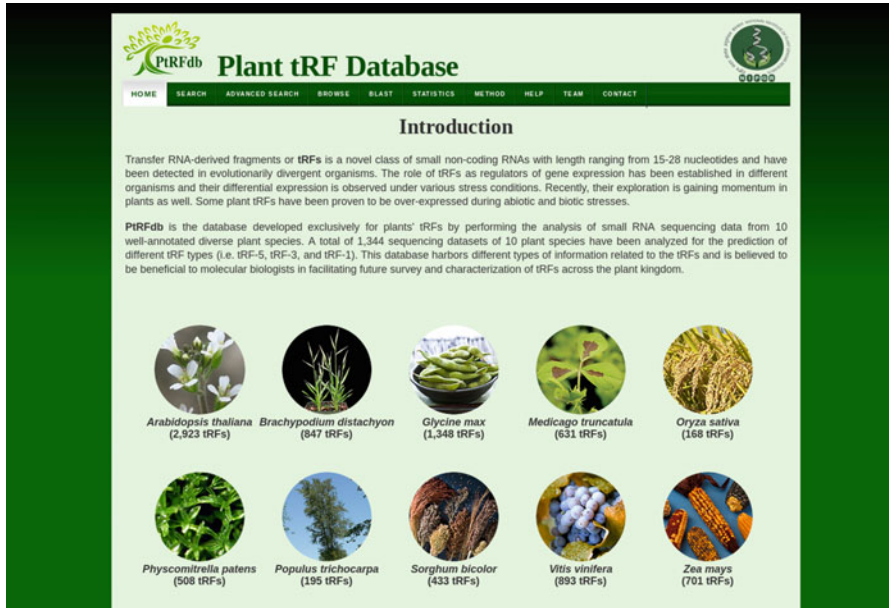


Fig. 27.2 Homepage of PtRFdb (<http://www.nipgr.res.in/PtRFdb/>)

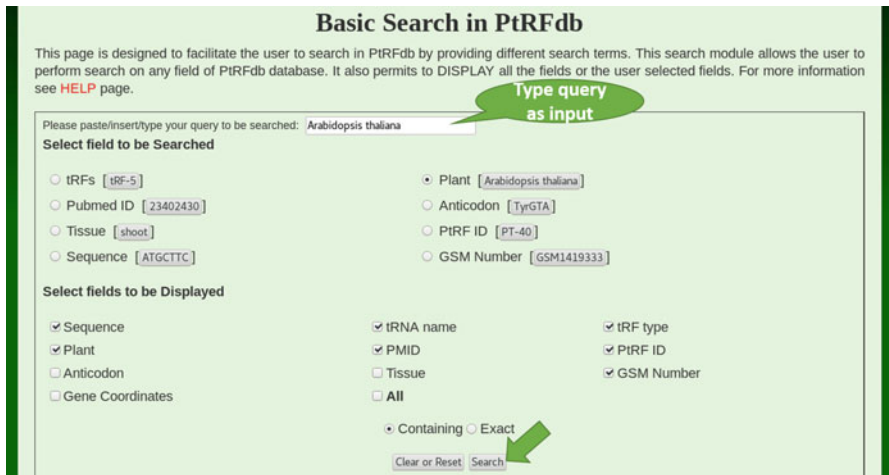


Fig. 27.3 Basic search page of PtRFdb (<http://www.nipgr.res.in/PtRFdb/search.php>)

Advanced search is supported with conditional and Boolean operators for user-built customized search. The ‘Browse’ section of PtRFdb facilitates the user to browse in three different manners: with respect to individual plants, by tRF types (i.e. tRF-5, 3, and 1) and also by anticodon type.

Advanced Search in PtRFdb

This page is designed to facilitate the user to search PtRFdb using multiple queries. You can use AND/OR. For more information see [HELP](#) page.

Explanation for Advanced Search

S. No.	Field to be Searched	Condition	Fill/Select Query	AND/OR	Add	Remove
1	tRF Type	=	tRF-1	AND	+	-
2	Tissue	LIKE	shoot	AND	+	-
3	Organism	LIKE	Arabidopsis	AND	+	-
4	PMID	=	19854858	AND	+	-

Clear Submit

Advanced Search Results

User can Click on icon to view the results in either ascending or descending order in the desired column. Click on the download icon on the right side to download the results displayed.

Show 10 entries

PtRF Id	Plant Name	tRF Type	Tissue	Gene Coordinates	tRNA Name	Anticodon	GSM Number	PMID	Sequence
PT-2285	Arabidopsis thaliana	tRF-1	shoot	chrCp:52016-52168	chrCp.tma7-MetCAT	MetCAT	GSM442935	19854858	TAACCTATTAGACACCATGA
PT-3127	Arabidopsis thaliana	tRF-1	shoot	chr1.2107358-2107508	chr1.tma237-GlyGCC	GlyGCC	GSM442934	19854858	TGAGCTGTGAAGAAATTGGCT
PT-3203	Arabidopsis thaliana	tRF-1	shoot	chr1.2107358-2107508	chr1.tma237-GlyGCC	GlyGCC	GSM442935	19854858	TGAGCTGTGAAGAAATTGGCTT
PT-3338	Arabidopsis thaliana	tRF-1	shoot	chr1.2111108-2111258	chr1.tma236-GlyGCC	GlyGCC	GSM442935	19854858	TGAGCTGTGATGAAATTGGCTT
PT-3339	Arabidopsis thaliana	tRF-1	shoot	chr1.2111108-2111258	chr1.tma236-GlyGCC	GlyGCC	GSM442934	19854858	TGAGCTGTGATGAAATTGGCTT

Fig. 27.4 Advanced search and its corresponding result page in PtRFdb (<http://www.nipgr.res.in/PtRFdb/cond.php>)

Plant tRF Database

HOME SEARCH ADVANCED SEARCH **BROWSE** BLAST STATISTICS METHOD HELP TEAM CONTACT

User can browse for tRF sequences by either of the options provided that can be easily accessed by clicking on and .

- Browse by Plant Name
- Browse by tRF type
- Browse by Anticodon

AlaAGC

- Arabidopsis thaliana
- Brachypodium distachyon
- Glycine max
- Medicago truncatula
- Oryza sativa
- Physcomitrella patens
- Populus trichocarpa
- Sorghum bicolor
- Vitis vinifera
- Zea mays

Fig. 27.5 Browse section of PtRFdb (<http://www.nipgr.res.in/PtRFdb/browse.php>)

In the ‘BLAST’ page of the PtRFdb (Fig. 27.6), BLASTN of any query nucleotide sequence can be performed against a particular plant species or over entire available datasets which can be selected as per the user’s requirement.

Fig. 27.6 Blast page of PtRFdb (<http://www.nipgr.res.in/PtRFdb/blast.php>)

To know the significance of the BLAST match, different ‘Expect value or E value’ ranging from 0.001 to 100 can be selected. The ‘Method’ section highlights sequencing steps for identification of tRFs in our study right from downloading raw datasets till tRFs prediction. In the ‘Statistics’ page, graphical and tabular representation of the overall distribution of different tRF types in individual plant species is provided. Lastly, the ‘Help’ section guides the users for simply understanding and navigating different modules of PtRFdb.

27.4 Notes

1. The total of 1344 sequencing datasets of ten plant species were used for the identification of tRFs.
2. Information associated with all analyzed GEO samples in our study was fetched by using the ‘SRAdb’ as well as ‘GEOmetadb’ libraries of the Bioconductor software package (<http://www.bioconductor.org>) and united for each entry of our database, PtRFdb.
3. For extraction of the mapping coordinates, in house PERL scripts were used.
4. This database holds information about 487,765 entries of tRFs (258,439 tRF-5s, 225,380 tRF-3s and 3946 tRF-1s).
5. The total number of 5607 unique tRFs sequences is incorporated in PtRFdb (2580 tRF-5s, 2269 tRF-3s, 758 tRF-1s).
6. The majority of the tRFs had a length in the range of 18–24 nt and tRF-5 were most abundant of all tRFs.

7. In the advanced search option, the conditional operators ‘=’ and ‘Like’ coupled with the two logical operators ‘AND’ and ‘OR’ were incorporated for user-built customized search.
8. For providing flexibility for search options, the ‘containing’ and ‘exact’ options have been provided.
9. BLAST-version 2.6.0 was utilized in the PtRFdb.
10. As Apache, PHP, and MySQL are free, open-source software, and are platform-independent, so they were preferably utilized for our database development.
11. In the future, attempts will be made for updating our database by further addition of more data.
12. For more details, related to PtRFdb, refer to our published paper (Gupta, N., Singh, A., Zahra, S., and Kumar S. PtRFdb: a database for plant transfer RNA-derived fragments. Database (Oxford). 2018 Jan 1; 2018. doi: 10.1093/database/bay063. PMID: 29939244).

References

- Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol.* 1990;215:403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Alves CS, Vicentini R, Duarte GT, et al. Genome-wide identification and characterization of tRNA-derived RNA fragments in land plants. *Plant Mol Biol.* 2017;93:35–48. <https://doi.org/10.1007/s11103-016-0545-9>.
- Chan PP, Lowe TM. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* 2009;37:D93–7. <https://doi.org/10.1093/nar/gkn787>.
- Cognat V, Morelle G, Megel C, et al. The nuclear and organellar tRNA-derived RNA fragment population in *Arabidopsis thaliana* is highly dynamic. *Nucleic Acids Res.* 2017;45:3460–72. <https://doi.org/10.1093/nar/gkw1122>.
- Gupta N, Singh A, Zahra S, Kumar S. PtRFdb: a database for plant transfer RNA-derived fragments. *Database.* 2018;2018:63.
- Hsieh L-C, Lin S-I, Shih AC-C, et al. Uncovering small RNA-mediated responses to phosphate deficiency in *Arabidopsis* by deep sequencing. *Plant Physiol.* 2009;151:2120–32. <https://doi.org/10.1104/pp.109.147280>.
- Jin D, Meng X, Wang Y, et al. Computational investigation of small RNAs in the establishment of root nodules and arbuscular mycorrhiza in leguminous plants. *Sci China Life Sci.* 2018;61(6):706–17. <https://doi.org/10.1007/s11427-017-9203-7>.
- Keam SP, Hutvagner G. tRNA-derived fragments (tRFs): emerging new roles for an ancient RNA in the regulation of gene expression. *Life (Basel, Switzerland).* 2015;5:1638–51. <https://doi.org/10.3390/life5041638>.
- Kumar P, Mudunuri SB, Anaya J, Dutta A. tRFdb: a database for transfer RNA fragments. *Nucleic Acids Res.* 2015;43:D141–5. <https://doi.org/10.1093/nar/gku1138>.
- Lee YS, Shibata Y, Malhotra A, Dutta A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* 2009;23:2639–49. <https://doi.org/10.1101/gad.1837609>.
- Loss-Morais G, Waterhouse PM, Margis R. Description of plant tRNA-derived RNA fragments (tRFs) associated with argonaute and identification of their putative targets. *Biol Direct.* 2013;8:6. <https://doi.org/10.1186/1745-6150-8-6>.
- Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 1997;25:955–64.

- Martinez G, Choudury SG, Slotkin RK. tRNA-derived small RNAs target transposable element transcripts. *Nucleic Acids Res.* 2017;45:5142–52. <https://doi.org/10.1093/nar/gkx103>.
- Orioli A. tRNA biology in the omics era: stress signalling dynamics and cancer progression. *BioEssays.* 2017;39:1600158. <https://doi.org/10.1002/bies.201600158>.
- Pliatsika V, Loher P, Telonis AG, Rigoutsos I. MINTbase: a framework for the interactive exploration of mitochondrial and nuclear tRNA fragments. *Bioinformatics.* 2016;32:2481–9. <https://doi.org/10.1093/bioinformatics/btw194>.
- Sablok G, Yang K, Chen R, Wen X. tRNA derived smallRNAs: smallRNAs repertoire has yet to be decoded in plants. *Front Plant Sci.* 2017;8:1–4. <https://doi.org/10.3389/fpls.2017.01167>.
- Selitsky SR, Sethupathy P. tDRmapper: challenges and solutions to mapping, naming, and quantifying tRNA-derived RNAs from human small RNA-sequencing data. *BMC Bioinforma.* 2015;16:354. <https://doi.org/10.1186/s12859-015-0800-0>.
- Sun C, Fu Z, Wang S, et al. Roles of tRNA-derived fragments in human cancers. *Cancer Lett.* 2018;414:16–25. <https://doi.org/10.1016/J.CANLET.2017.10.031>.
- Thompson A, Zielezinski A, Plewka P, et al. tRex: a web portal for exploration of tRNA-derived fragments in *Arabidopsis thaliana*. *Plant Cell Physiol.* 2018;59:e1. <https://doi.org/10.1093/pcp/pcx173>.
- Vickers KC, Roteta LA, Hucheson-Dilks H, et al. Mining diverse small RNA species in the deep transcriptome. *Trends Biochem Sci.* 2015;40:4–7. <https://doi.org/10.1016/J.TIBS.2014.10.009>.
- Wang Q, Li T, Xu K, et al. The tRNA-derived small RNAs regulate gene expression through triggering sequence-specific degradation of target transcripts in the oomycete pathogen *Phytophthora sojae*. *Front Plant Sci.* 2016a;07:1938. <https://doi.org/10.3389/fpls.2016.01938>.
- Wang Y, Li H, Sun Q, Yao Y. Characterization of small RNAs derived from tRNAs, rRNAs and snoRNAs and their response to heat stress in wheat seedlings. *PLoS One.* 2016b;11:e0150933. <https://doi.org/10.1371/journal.pone.0150933>.
- Zheng L-L, Xu W-L, Liu S, et al. tRF2Cancer: a web server to detect tRNA-derived small RNA fragments (tRFs) and their expression in multiple cancers. *Nucleic Acids Res.* 2016;44:W185–93. <https://doi.org/10.1093/nar/gkw414>.