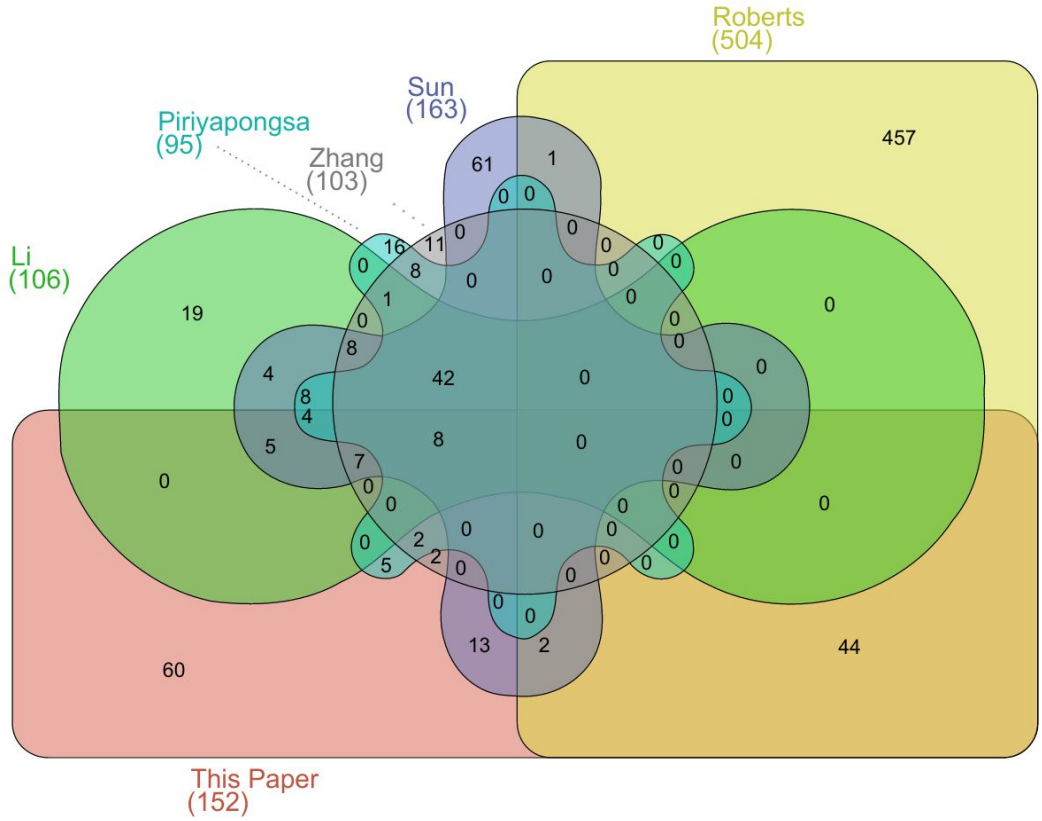
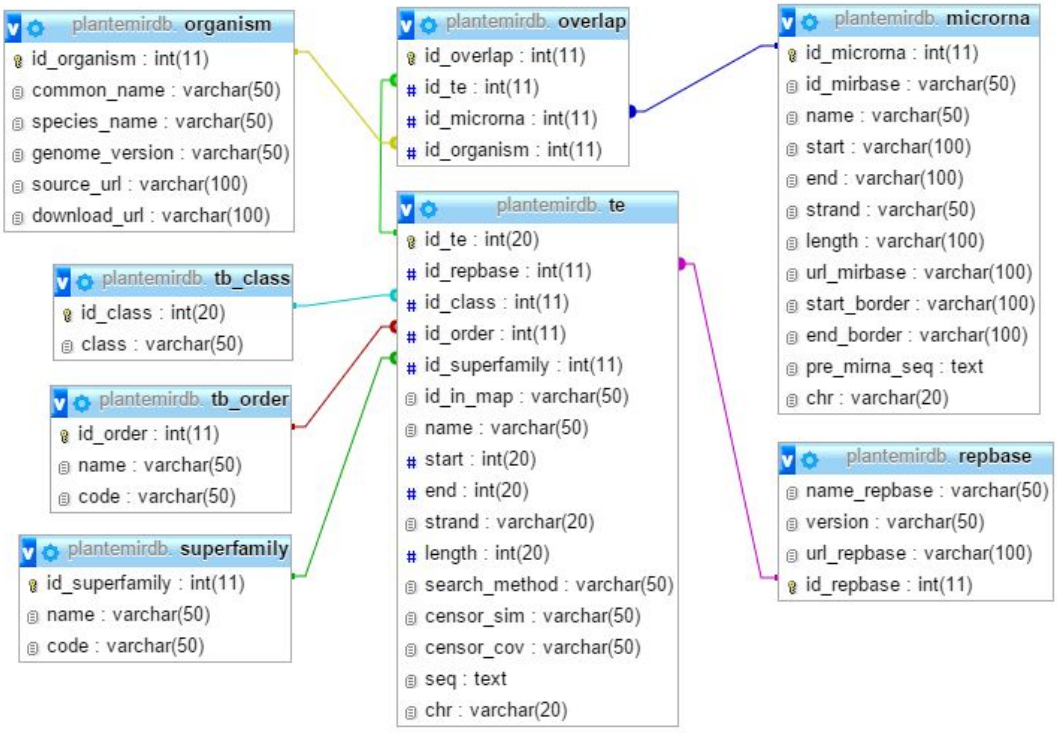
ELECTRONIC SUPPLEMENTARY INFORMATION



Supplementary Fig. 1. Venn diagram showing the quantity of TE-related pre-miRNAs found by other authors compared to our study (Piriyapongsa; Jordan, 2008; Li et al., 2011; Zhang; Jiang; Gao, 2011; Sun et al., 2012; Roberts et al., 2013). Only miRBase reference data from Roberts et al. (2013) were used to compare results. Venn diagram was plotted using InteractiVenn (Heberle et al., 2015).



Supplementary Fig. 2. Entity-Relationship Diagram (ERD) of PlanTE-MIR DB.

Species Name	Common Name	Version	URL
Arabidopsis thaliana	thale cress	TAIR v10.0	ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Athaliana/assembly/Athaliana_167.fa.gz
Arabidopsis lyrata	lyre-leaved rock cress	JGI v1.0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000004255.1_v.1.0/GCA_000004255.1_v.1.0_genomic.fna.gz
Brachypodium distachyon	purple false brome	JGI v1.0	ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Bdistachyon/assembly/Bdistachyon_192.fa.gz
Glycine max	soybean	Glyma v1.0	ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Gmax/assembly/Gmax_189.fa.gz
Lotus japonicus	miyakogusa	Lj v2.5	ftp://ftp.kazusa.or.jp/pub/lotus/lotus_r2.5/pseudomolecule/Lj2.5_pseudomol.fna.gz
Malus domestica	apple	maldom pseudo v1.0	http://www.rosaceae.org/system/files/apple_genome/Malus_x_domestica.v1.0-primary.pseudo.fa.gz
Medicago truncatula	barrel medic	MedtrA17 v3.5	ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Mtruncatula/assembly/Mtruncatula_198.fa.gz
Oryza sativa	rice	MSU v7.0	ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Osativa/assembly/Osativa_204.fa.gz
Physcomitrella patens	moss	JGI v1.1	ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Ppatens/assembly/Ppatens_152.fa.gz
Populus trichocarpa	poplar	JGI Poptr2.0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000002775.2_Poptr2_0/GCA_000002775.2_Poptr2_0_genomic.fna.gz
Prunus persica	peach	JGI v1.0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000346465.1_Prupe1_0/GCA_000346465.1_Prupe1_0_genomic.fna.gz
Solanum lycopersicum	tomato	SL v2.40	ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Slycopersicum/assembly/Slycopersicum_225.fa.gz
Solanum tuberosum	irish potato	SolTub v3.0	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000226075.1_SolTub_3.0/GCA_000226075.1_SolTub_3.0_genomic.fna.gz
Sorghum bicolor	sorghum	JGI Sb v1.0	ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Sbicolor/assembly/Sbicolor_79.fa.gz
Vitis vinifera	grape	genoscope march 2010	ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/Vvinifera/assembly/Vvinifera_145.fa.gz

Supplementary Table 1. Fifteen species assessed by our analysis. Versions were retrieved from genome assembly source repositories.

Table format description

Alongside with detailed information showed on the website, data can be downloaded as a tab-separated table format. It was made this way for easy parsing of entries. Each one of twenty seven columns refers to following attributes:

1. Species Name: Species binomial nomenclature.
2. Common Name: Organism's popular name.
3. Assembly Version: Genome assembly version (see list in Supplementary Table 1).
4. Assembly URL: Download genome assemblies available at source website.
5. TE Name: Names were given to each annotated TE which had intersection with any pre-miRNA. Nomenclature rules and classification system were used according to Wicker and co-workers (2007).
6. TE Class: Higher level of most recent classification system for TEs. Based on Repbase.
7. TE Order: Another classification level (may be absent). Based on Repbase.
8. TE Superfamily: Lower level of classification system. Based on Repbase.
9. Repbase Name: Repbase's TE consensus name. Used to identify TE consensus in Repbase.
10. Repbase Consensus URL: Direct link to access TE consensus notes in Repbase.
11. Repbase Version: Version of Repbase's library used for CENSOR's annotation.
12. Search Method: Search method employed by CENSOR to find repetitions in genome assemblies.
13. Coverage (CENSOR): Fraction of TE consensus aligned to genome assemblies (values between 80% and 100%).
14. Similarity (CENSOR): Similarity between TE consensus and genome assemblies (values between 0.8 and 1.0). More information is available at CENSOR's help webpage: http://www.girinst.org/censor/help.html.
15. TE Chromosome or scaffold: Accession in genome assembly FASTA files (e.g. chr1).
16. TE Start Position: One-based start coordinate of annotated TE.
17. TE End Position: One-based end coordinate of annotated TE.
18. TE Strand: "+" (plus or sense) and "-" (minus or antisense).
19. Overlapping pre-miRNA: pre-miRNA name available at miRBase.
20. pre-miRNA ID: pre-miRNA accession available at miRBase.
21. pre-miRNA URL: pre-miRNA information URL at miRBase.
22. pre-miRNA Chromosome or scaffold: Accession in genome assembly FASTA files (e.g. chr1).
23. pre-miRNA Start position: One-based start coordinate of annotated TE.
24. pre-miRNA End position: One-based end coordinate of annotated TE.
25. pre-miRNA Strand: "+" (plus or sense) and "-" (minus or antisense).
26. pre-miRNA Sequence: pre-miRNA complete sequence (RNA sequence).
27. TE Sequence: annotated TE complete sequence (DNA sequence).

Generic Feature Format Version 3 (GFF3) description

GFF3 files are well known by scientific community and are described at http://www.sequenceontology.org/gff3.shtml. For better understanding, tab-separated column content are described below (bold text and quotes emphasize the attributes):

TE GFF3:

Column #1: TE Chromosome or scaffold
Column #2: Search Method
Column #3: Key (transposable_element in this case)
Column #4: TE Start Position
Column #5: TE End Position
Column #6: Score (in this case "." means not available)
Column #7: TE Strand
Column #8: Phase (in this case "." means not available)
Column #9: Features (comprises following additional information)
ID="**Repbase Name**";Name="**TE Name**";Alias="**Class**":"**Order**":"**Superfamily**";Note="**Similarity (CENSOR)**":"**Coverage (CENSOR)**":"**Repbase Version**"

E.g.:
ID=**ATMU3N1**;Name=**DTx_ATMU3N1_Chr1-1**;Alias=**Class II (DNA transposons) - Subclass I**:**TIR**:**Undefined**;Note=**0.9529**:**100.00**:**Repbase19.04**

pre-miRNA GFF3:

Column #1: pre-miRNA Chromosome or scaffold
Column #2: Source (miRBasev21 in this case)
Column #3: Key (ncRNA in this case)
Column #4: pre-miRNA Start Position
Column #5: pre-miRNA End Position
Column #6: Score (in this case "." means not available)
Column #7: pre-miRNA Strand
Column #8: Phase (in this case "." means not available)
Column #9: Features (comprises following additional information)
ID="**pre-miRNA ID**";Name="**pre-miRNA Name**"

E.g.:
ID=**MI0019229**;Name=**ath-MIR5635b**

FASTA header description

Respecting the presented titles, FASTA files for TEs contain the following header organization:
>"**TE Name**" "**Repbase Name**" intersects "**pre-miRNA Name**" ["**Species Name**"]

E.g.:
>**RLG_ATHILA4C_LTR_Chr2-1 ATHILA4C_LTR** intersects **ath-MIR8175** [**Arabidopsis thaliana**]

Using the same rules, the headers for pre-miRNAs follow this sample:
>"**pre-miRNA Name**" "**pre-miRNA ID**" intersects "**TE Name**" ["**Species Name**"]

E.g.:
>**ath-MIR8175 MI0026805** intersects **RLG_ATHILA4C_LTR_Chr2-1** [**Arabidopsis thaliana**]

Data visualization

We suggest using Artemis software for data visualization. Entire datasets and assemblies can be downloaded at Downloads section and loaded in Artemis. All files are ready to use, except for *Populus trichocarpa* assembly, which may need corrections in the FASTA file headers in order to match GFF3 accessions. The scaffold names must be "scaffold_" followed by their respective number (e. g. scaffold_1).

References

HEBERLE, H. et al. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. **BMC bioinformatics**, v. 16, n. 1, p. 169, 2015.

LI, Y. et al. Domestication of transposable elements into microRNA genes in plants. **Plos one**, v. 6, n. 5, p. e19212, 2011.

PIRIYAPONGSA, J.; JORDAN, I. K. Dual coding of siRNAs and miRNAs by plant transposable elements. **Rna**, v. 14, n. 5, p. 814-821, 2008.

ROBERTS, Justin T. et al. Continuing analysis of microRNA origins: Formation from transposable element insertions and noncoding RNA mutations. **Mobile genetic elements**, v. 3, n. 6, p. e27755, 2013.

SUN, J. et al. Characterization and evolution of microRNA genes derived from repetitive elements and duplication events in plants. **PloS one**, v. 7, n. 4, p. e34092, 2012.

WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nature Reviews Genetics**, v. 8, n. 12, p. 973-982, 2007.

ZHANG, Y.; JIANG, W.; GAO, L. Evolution of microRNA genes in Oryza sativa and Arabidopsis thaliana: an update of the inverted duplication model. **PloS one**, v. 6, n. 12, p. e28073, 2011.