



HAL
open science

RepetDB: A TE database

Nicolas Francillonne, Mariène Wan, Nathalie Choisne, Françoise Alfama,
Raphaël Flores, Joelle Amselem, Johann Confais, Hadi Quesneville

► **To cite this version:**

Nicolas Francillonne, Mariène Wan, Nathalie Choisne, Françoise Alfama, Raphaël Flores, et al..
RepetDB: A TE database. International congress on transposable elements ICTE, Apr 2024, Saint
Malo, France. hal-04570060

HAL Id: hal-04570060

<https://hal.science/hal-04570060v1>

Submitted on 6 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Nicolas FRANCILLONNE^{1,2}, Mariène WAN^{1,2}, Nathalie CHOISNE^{1,2}, Françoise ALFAMA^{1,2}, Raphaël FLORES^{1,2}, Joelle AMSELEM^{1,2}, Johann CONFAIS^{1,2} and Hadi QUESNEVILLE^{1,2}

¹ Université Paris-Saclay, INRAE, URGI, 78026, Versailles, France

² Université Paris-Saclay, INRAE, BioinfOmics, Plant bioinformatics facility, 78026, Versailles, France

Transposable elements (TEs) are major players of structure and evolution of eukaryote genomes. Thanks to their ability to move around and to replicate within genomes, they are probably the most important contributors to genome plasticity. **Their detection and annotation are considered essential and must be undertaken in any genome sequencing project.**

To annotate transposable elements in genome, consensus are needed and RepetDB aims to make available these sequence to make easier the annotation of closely related genomes. You can select consensus of interest in the database and download fasta sequence.

- I wish to make an annotation of a Brassicaceae species close from *Arabidopsis thaliana*. - I go to repetdb website and check the related species of interest <https://urgi.versailles.inrae.fr/repetdb>

- I have an interesting sequence that I suspect is a transposable element and I wish to check if a match exist in the database. I go to PlantbioinfoPF blast services <https://urgi.versailles.inrae.fr/blast/> and select in group « Repeats » a close related species

- I can also select multiple species, from example all Brassicaceae or just one close related species like *Arabidopsis thaliana*

- If a match is found I get a direct link to the consensus card

- Selection of specifics class, order or superfamily is also available to refine a search

By selecting one consensus I can access a consensus card describing its classification

Consensus search result

Back to home screen

Search criteria: Taxon group: Arabidopsis thaliana

Showing 1 to 25 of 406 rows

| Consensus Identifier | Consensus Copies | Consensus Full-length copies | Consensus Length | Consensus TE classification code | Consensus Manual validation | Consensus Confused | Consensus Comments |
|---|------------------|------------------------------|------------------|----------------------------------|-----------------------------|--------------------|--------------------|
| MCL102_AthaCol0_TEdenovo-B-G1208-Map3 | 20 | 0 | 471 | Class I : LINE : ? | classification | no | NO VALUE |
| MCL102_AthaCol0_TEdenovo-B-R178-Map3 | 122 | 8 | 2017 | Class I : LINE : ? | classification | no | NO VALUE |
| MCL103_AthaCol0_TEdenovo-B-G1214-Map20 | 23 | 11 | 449 | Class II : Helitron : ? | classification | no | NO VALUE |
| MCL103_AthaCol0_TEdenovo-B-P161.100-Map16 | 21 | 11 | 452 | Class II : Helitron : ? | classification | no | NO VALUE |

- I can obtain statistics and filter some values out or download my result and consensus fasta sequences