

Transposable Elements in the Era of Data Science

**From Data to Discovery: Using database to TE
research**

Alexandre Paschoal, Gökhan Karakulah and Romain Guyot

EMBO Databases TEhub

<https://tehub.org/>

TE Hub

Search...

Home

Classification

Databases

Tools

Tutorials

Protocols

Conferences

Journals

Teaching Resources

Outreach

Volunteering

Working Group

About

resources / repeat_databases

Repeat databases

Description

This page compiles a list of databases for the storage of sequences and metadata associated with repetitive, mobile and selfish DNA.

It is maintained by [Tyler A. Elliott](#) . It was initiated with the help of Elizabeth Smikle and Miduna Rahulan, formerly at the [Centre for Biodiversity Genomics](#) at the [University of Guelph](#) . Suggestions, updates and error corrections are welcome. Please feel free to add missing databases into the table, that would help a lot!

For resources more tailored to [Prokaryotes](#) and [Viruses](#) , see the linked lists.

Overview of repeat databases

ResourceFind...	DOIFind...	Taxonomic GroupFind...	Repeat TypesFind...	StatusFind...
3'UTR-SIRE	https://doi.org/10.1186/1471-2105-8-274	Mammal	SINE	Absent (Apr 28 2025)
ACLAME	https://doi.org/10.1093/nar/gkp938	Archaea, Bacteria	Plasmid, Virus	Absent (Apr 28 2025)
Actinobacteriophage Database (PhagesDB)	https://doi.org/10.1093/bioinformatics/btw711	Bacteria	Prophage/Bacteriophage	Absent (Apr 28 2025)
alu_ontology	https://doi.org/10.1016/j.jbi.2016.01.010	Homo sapiens	Alu, SINE	Absent (Apr 28 2025)
APTEdb (Atlas of Plant Transposable Elements Database)	https://f1000research.com/articles/10-1194/v1	Land Plant, Algae	Eukaryotic Transposon	Present (Apr 28 2025)
ARDB (Antibiotic Resistance Genes Database)	https://doi.org/10.1093/nar/gkn656	Archaea, Bacteria	AMR/Antibiotic Resistance	Present (Apr 28 2025)
ArTEDB (Arthropod Transposable Elements Database)	https://doi.org/10.3390/genes10050338	Arthropod	Eukaryotic Transposon	
Bamboo Microsatellite Database	https://doi.org/10.1038/srep08018	Phyllostachys	Simple Sequence Repeat/Microsatellite	Absent (Apr 28 2025)
BanSatDB (Banana Satellite Database)	https://doi.org/10.1016/j.cj.2018.01.005	Musa	Simple Sequence Repeat/Microsatellite	Absent (Apr 28 2025)
BeetRepeats	https://doi.org/10.1186/s13104-024-06993-4	Beta, Patellifolia, Chenopodium, Spinacia	Eukaryotic Repeat	Absent (Apr 28 2025)
Bivalvia_TEs	https://doi.org/10.1186/s12915-023-01632-z	Bivalvia	Eukaryotic Transposon	Present (Apr 28 2025)
BMTEDb (Bombyx mori Transposable Elements Database)	https://doi.org/10.1093/database/bat055	Bombyx mori	Eukaryotic Transposon	
Boechera Microsatellite Database	https://doi.org/10.1093/database/baw169	Boechera	Simple Sequence Repeat/Microsatellite	Present (Apr 28 2025)
Brassica TE Db	https://doi.org/10.1186/1756-0500-7-379	Brassica rapa, Brassica oleracea	Eukaryotic Transposon	
BuffSatDB (Buffalo Satellite Database)	https://doi.org/10.1186/1471-2164-14-43	Bubalis bubalis	Simple Sequence Repeat/Microsatellite	Absent (Apr 28 2025)
CAGm (Comparative Analysis of Germline Microsatellites)	https://doi.org/10.1093/nar/gky969	Homo sapiens	Simple Sequence Repeat/Microsatellite	Absent (Apr 28 2025)
CARD (Comprehensive Antibiotic Resistance Database)	https://doi.org/10.1093/nar/gkz935	Archaea, Bacteria	AMR/Antibiotic Resistance	Present (Apr 28 2025)
CarnationDB	https://doi.org/10.1093/dnares/dst053	Plant, Dianthus	Eukaryotic Repeat	Absent (Apr 28 2025)
CGATE (Catalogue of Genes Affected by Transposable Elements)	https://doi.org/10.1186/1759-8753-3-9	Plant, Animal	Eukaryotic Transposon	Present (Apr 28 2025)
ChloroMitoSSRDB 2.0	https://doi.org/10.1093/database/bav084	Eukaryote	Organelle Microsatellite	Absent (Apr 28 2025)
CiArMiSatDB (Cicer arietinum Microsatellite Database)	https://doi.org/10.1186/1471-2105-15-212	Plant, Cicer	Simple Sequence Repeat/Microsatellite	Present (Apr 28 2025)

EMBO Databases TEhub

<https://tehub.org/>

TE Hub				
Search...				
resources / repeat_databases				
Repeat databases				
Description				
This page compiles a list of databases for the storage of sequences and metadata associated with repetitive, mobile and selfish DNA.				
It is maintained by Tyler A. Elliott . It was initiated with the help of Elizabeth Smikle and Miduna Rahulan, formerly at the Centre for Biodiversity Genomics at the University of Guelph . Suggestions, updates and error corrections are welcome. Please feel free to add missing databases into the table, that would help a lot!				
For resources more tailored to Prokaryotes and Viruses , see the linked lists.				
Overview of repeat databases				
Resource: Find...	DOI: Find...	Taxonomic Group: Find...	Repeat Types: Find...	Status: Find...
3'UTR-SIRE	https://doi.org/10.1186/1471-2105-8-274	Mammal	SINE	Absent (Apr 28 2025)
ACLAME	https://doi.org/10.1093/nar/gkp938	Archaea, Bacteria	Plasmid, Virus	Absent (Apr 28 2025)
Actinobacteriophage Database (PhagesDB)	https://doi.org/10.1093/bioinformatics/btw711	Bacteria	Prophage/Bacteriophage	Absent (Apr 28 2025)
alu_ontology	https://doi.org/10.1016/j.jbi.2016.01.010	Homo sapiens	Alu, SINE	Absent (Apr 28 2025)
APTEdb (Atlas of Plant Transposable Elements Database)	https://f1000research.com/articles/10-1194/v1	Land Plant, Algae	Eukaryotic Transposon	Present (Apr 28 2025)
ARD (Archaeal Repeat Database)	https://doi.org/10.1093/nar/gkz935	Archaea, Bacteria	Archaeal Repeat, Plasmid, Virus	Present (Apr 28 2025)
ARTEdb (Arthropod Transposable Elements Database)	https://doi.org/10.3390/genes10050338	Arthropod	Eukaryotic Transposon	Absent (Apr 28 2025)
Bamboo Microsatellite Database	https://doi.org/10.1038/srep08018	Phyllostachys	Simple Sequence Repeat/Microsatellite	Absent (Apr 28 2025)
BanSatDB (Banana Satellite Database)	https://doi.org/10.1016/j.cj.2018.01.005	Musa	Simple Sequence Repeat/Microsatellite	Absent (Apr 28 2025)
BeetRepeats	https://doi.org/10.1186/s13104-024-06993-4	Beta, Patellifolia, Chenopodium, Spinacia	Eukaryotic Repeat	Absent (Apr 28 2025)
Bivalvia_TEs	https://doi.org/10.1186/s12915-023-01632-z	Bivalvia	Eukaryotic Transposon	Present (Apr 28 2025)
BMTEdb (Bombyx mori Transposable Elements Database)	https://doi.org/10.1093/database/bat055	Bombyx mori	Eukaryotic Transposon	
Boechera Microsatellite Database	https://doi.org/10.1093/database/baw169	Boechera	Simple Sequence Repeat/Microsatellite	Present (Apr 28 2025)
Brassica TE Db	https://doi.org/10.1186/1756-0500-7-379	Brassica rapa, Brassica oleracea	Eukaryotic Transposon	
BuffSatDB (Buffalo Satellite Database)	https://doi.org/10.1186/1471-2164-14-43	Bubalis bubalis	Simple Sequence Repeat/Microsatellite	Absent (Apr 28 2025)
CAGm (Comparative Analysis of Germline Microsatellites)	https://doi.org/10.1093/nar/gky969	Homo sapiens	Simple Sequence Repeat/Microsatellite	Absent (Apr 28 2025)
CARD (Comprehensive Antibiotic Resistance Database)	https://doi.org/10.1093/nar/gkz935	Archaea, Bacteria	AMR/Antibiotic Resistance	Present (Apr 28 2025)
CarnationDB	https://doi.org/10.1093/dnares/dst053	Plant, Dianthus	Eukaryotic Repeat	Absent (Apr 28 2025)
CGATE (Catalogue of Genes Affected by Transposable Elements)	https://doi.org/10.1186/1759-8753-3-9	Plant, Animal	Eukaryotic Transposon	Present (Apr 28 2025)
ChloroMitoSSRDB 2.0	https://doi.org/10.1093/database/bav084	Eukaryote	Organelle Microsatellite	Absent (Apr 28 2025)
CiArMiSatDB (Cicer arietinum Microsatellite Database)	https://doi.org/10.1186/1471-2105-15-212	Plant, Cicer	Simple Sequence Repeat/Microsatellite	Present (Apr 28 2025)

CONSENSUS vs individual sequences (redundant or not)

Databases — consensus —

Rebase a collection of consensus sequences

- The largest collection of eukaryotic transposons and repetitive sequences
- Includes over 44,000 sequences (mostly family consensus)
- Covers over a hundred model organisms and species of interest including animals, plants and fungi
- Subject to extensive manual curation and ongoing updates with the aim of recovering all repeat families in each species
- Sequences are systematically classified according to the nature of the repeats
- Essential for well-known tools such as CENSOR, RepeatMasker, REPET
- Recognized and used as a standard around the world

But Subscription required !

Databases — consensus —

Rebase a collection of consensus sequences

39 divisions (ref files)

For plants: 6 divisions: athrep, dcotrep, grasrep, mcotrep, oryrep, plnrep

athrep.ref 1136
dcotrep.ref 840
grasrep.ref 6422
mcotrep.ref 371
oryrep.ref 3456
plnrep.ref 3840

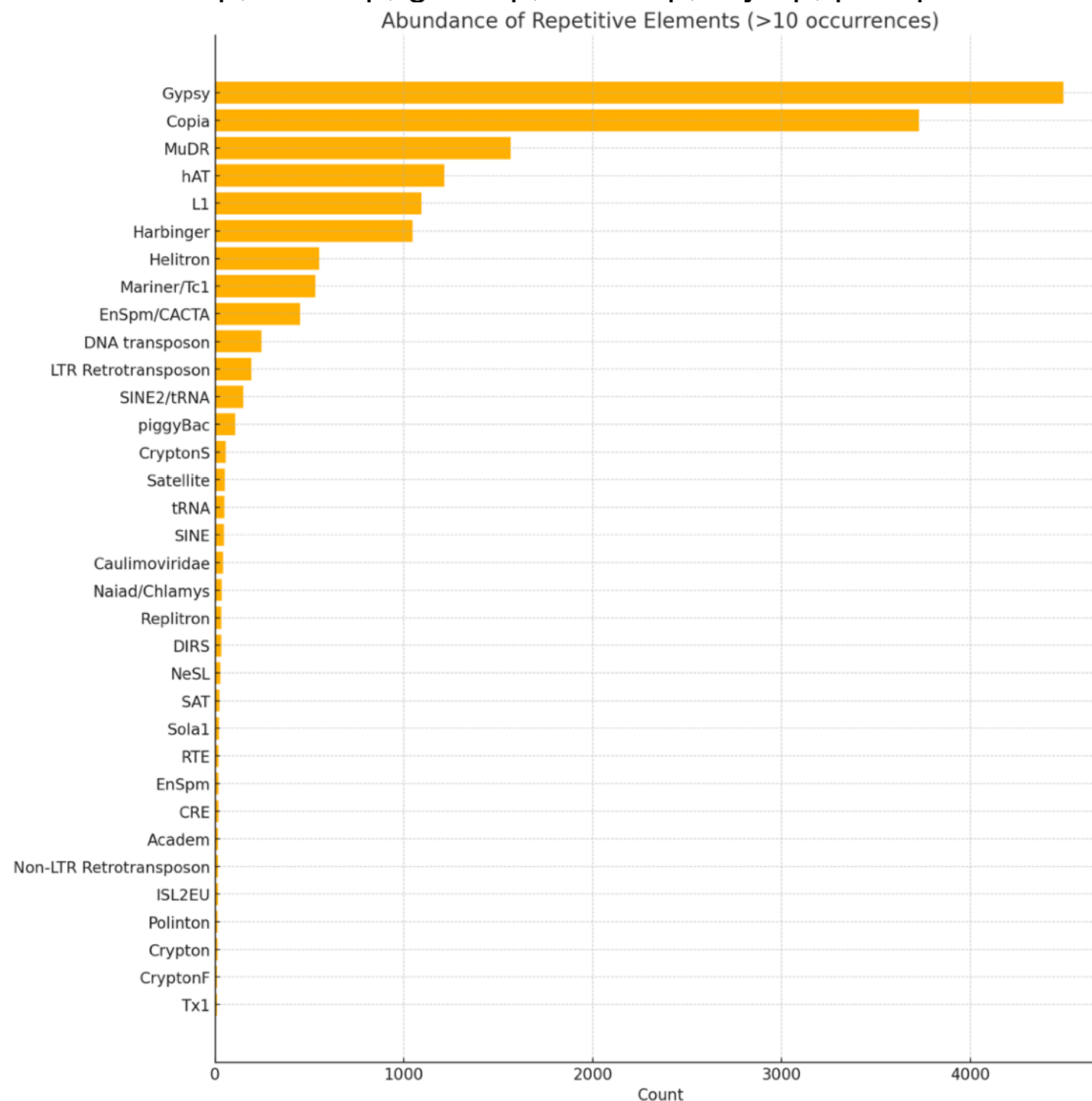
Databases — consensus —

Rebase a collection of consensus sequences

39 divisions (ref files)

For plants: 6 divisions: athrep, dcotrep, grasrep, mcotrep, oryrep, plnrep

athrep.ref 1136
dcotrep.ref 840
grasrep.ref 6422
mcotrep.ref 371
oryrep.ref 3456
plnrep.ref 3840



Databases — consensus —

Rebase a collection of consensus sequences

39 divisions (ref files)

For plants: 6 divisions: athrep, dcotrep, grasrep, mcotrep,

oryrep, plnrep

athrep.ref 1136

dcotrep.ref 840

grasrep.ref 6422

mcotrep.ref 371

oryrep.ref 3456

plnrep.ref 3840





EMBO

Databases — consensus —

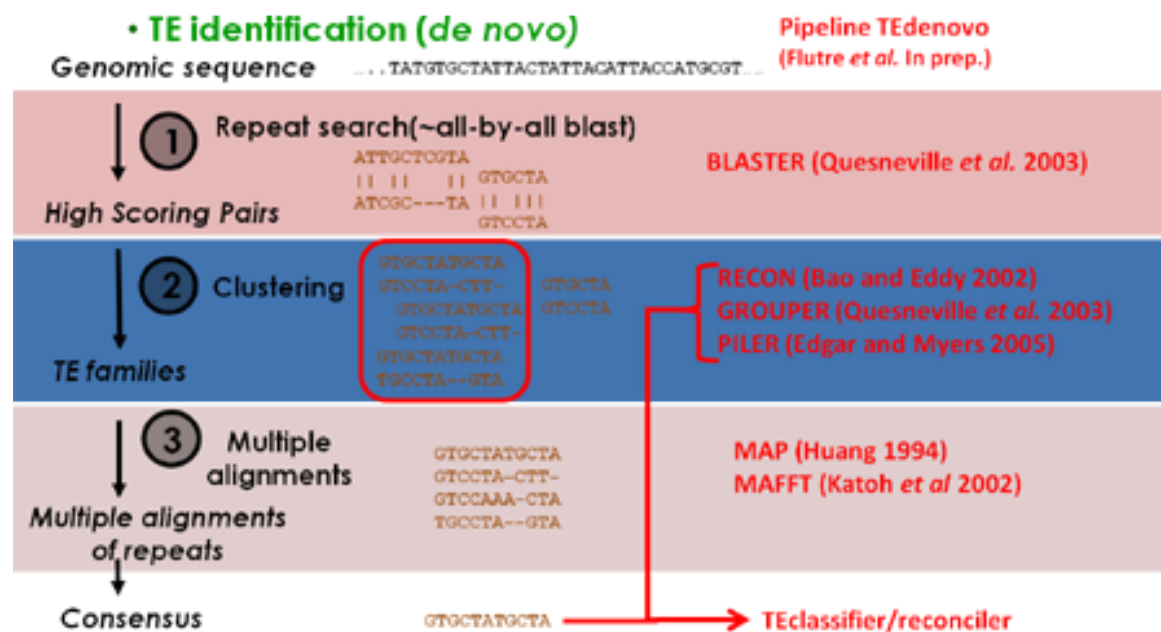
REPETdb a collection of consensus sequences

<https://urgi.versailles.inra.fr/repetdb/begin.do>

Produce with REPET TEdenovo.

Tutorial of REPET: <https://forgemia.inra.fr/urgi-anagen/wiki-repet/-/wikis/REPET-V3.0-tutorial>

Classification following the Wicker's acronyms



RepetDB v2 Transposable element Consensus database

Home Templates Lists QueryBuilder Data Sources API MyMine

Contact Us | Log In

Search: e.g. DHX*, GAG, Helitron GO

Search

Search RepetDB. Enter names, identifiers or keywords for consensus, protein profiles, Wicker classification term, etc.

e.g. DHX*, GAG, Helitron

SEARCH

Analyse

Enter a list of identifiers.

Consensus

e.g.

advanced

ANALYSE

Welcome Back!

RepetDB provides repeat consensus detected and classified by TEdenovo and used by TEannot to annotate copies in genomes.

HOW TO USE REPETDB

Search consensus

Taxon group All

Wicker Classification All classes All orders All superfamilies

Confused classification ☐ Only confused ☐ Only not confused

☐ Unclassified or not TE ☐ Only TE

Manual Validation All

Similarity feature(s)

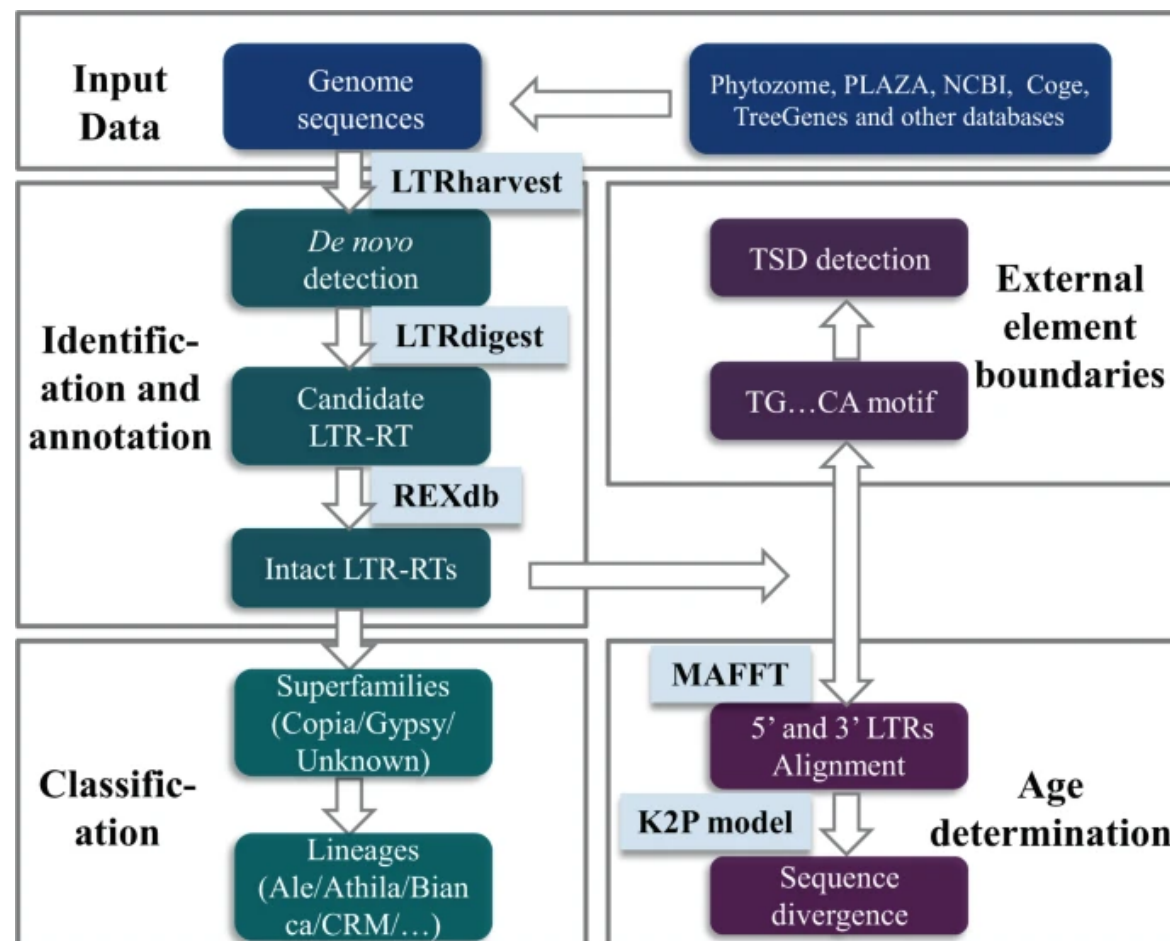
comma-separated list of protein profiles (accessions from GyDB or PFAM) or transposon (accessions from Repbase). Only exact identifier can be queried (example: PF13650.5, INT_csm1,PF13976.5). Version identifier is mandatory

Search

Databases. —individual sequences—

A comprehensive annotation dataset of intact LTR retrotransposons of 300 plant genomes

<https://github.com/sszhou9/intact-LTR-RTs>

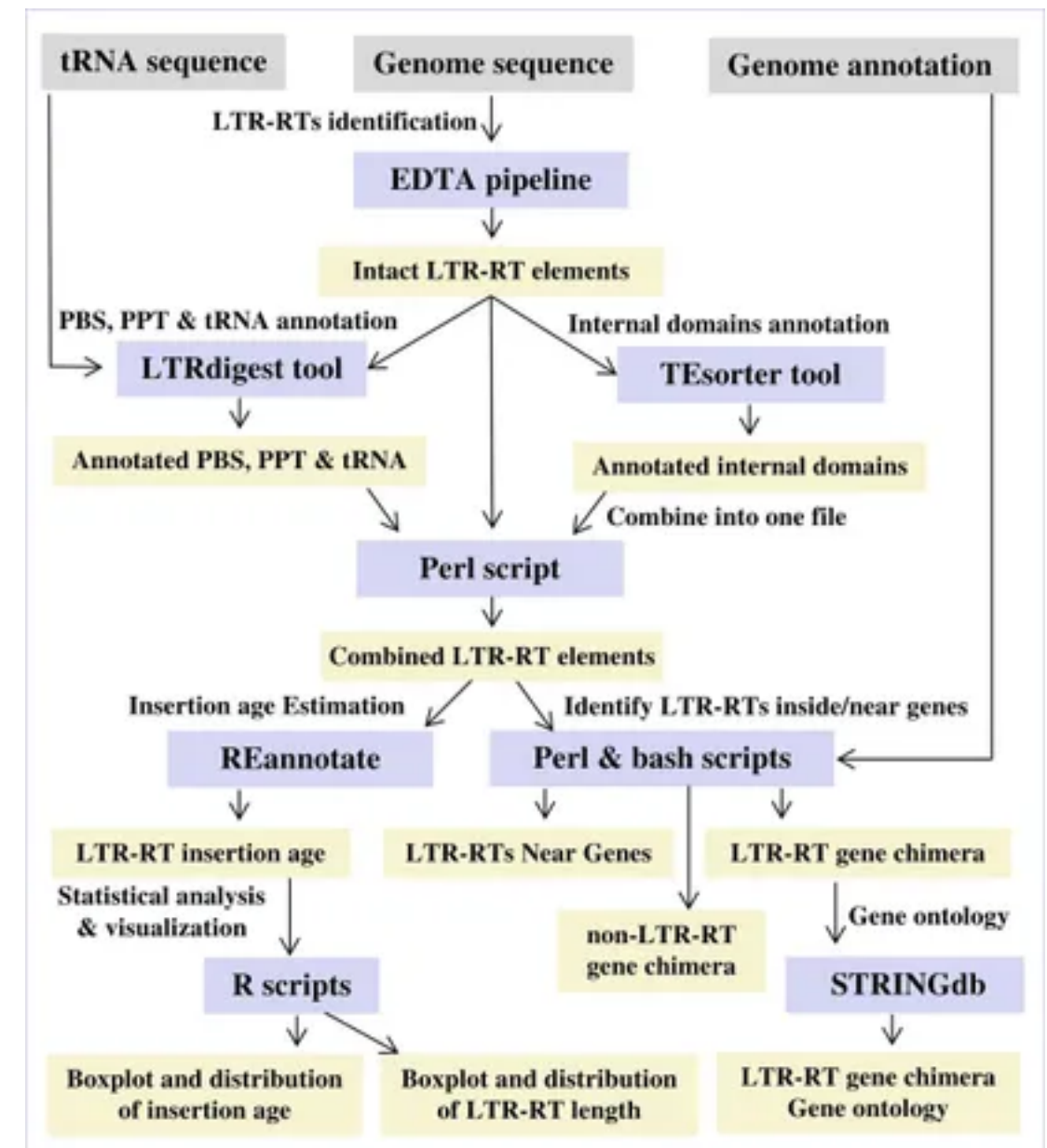
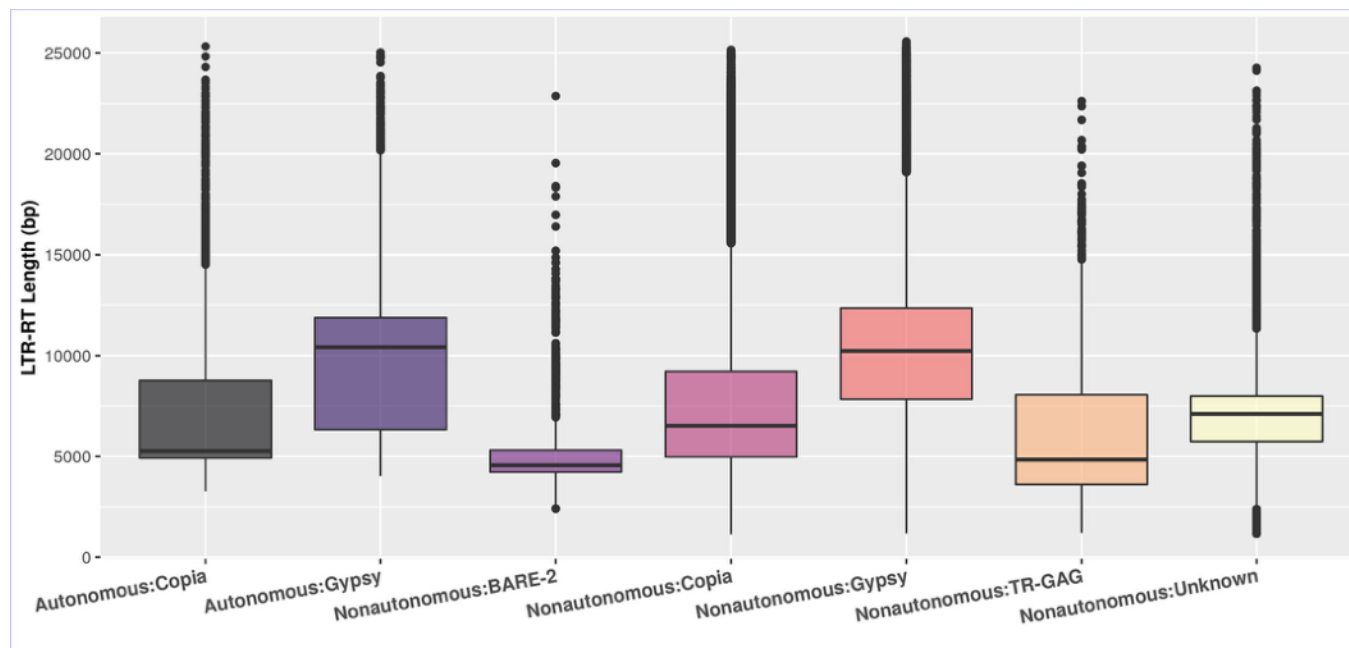


2,593,685 intact LTR-RTs from genomes of 300 plant species representing 93 families of 46 orders

Databases – individual sequences –

PlantLTRdb

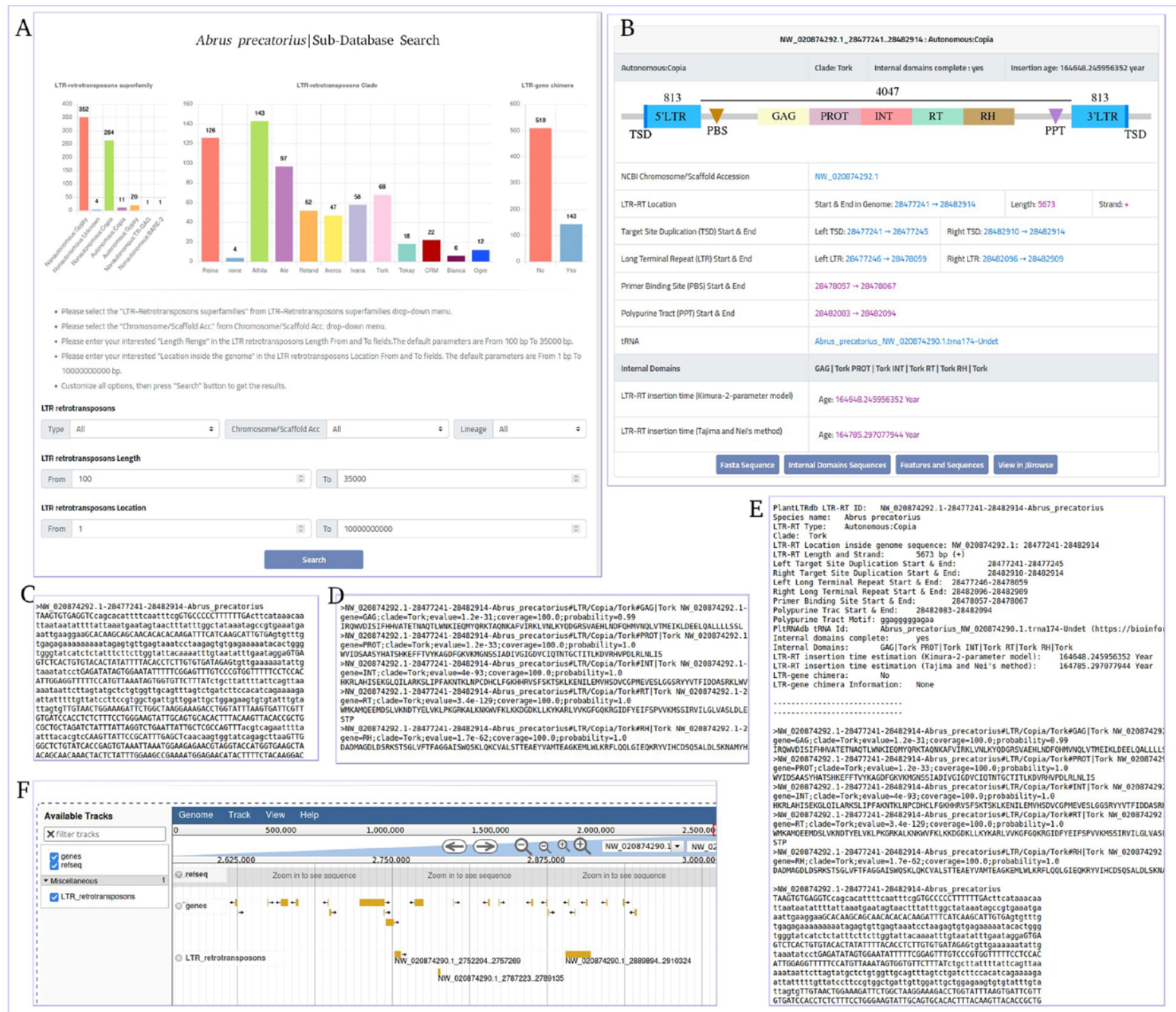
<https://bioinformatics.um6p.ma/PlantLTRdb/>



520,194 intact elements from 195 plant genomes

EMBO Databases – individual sequences – PlantLTRdb

<https://bioinformatics.um6p.ma/PlantLTRdb/>

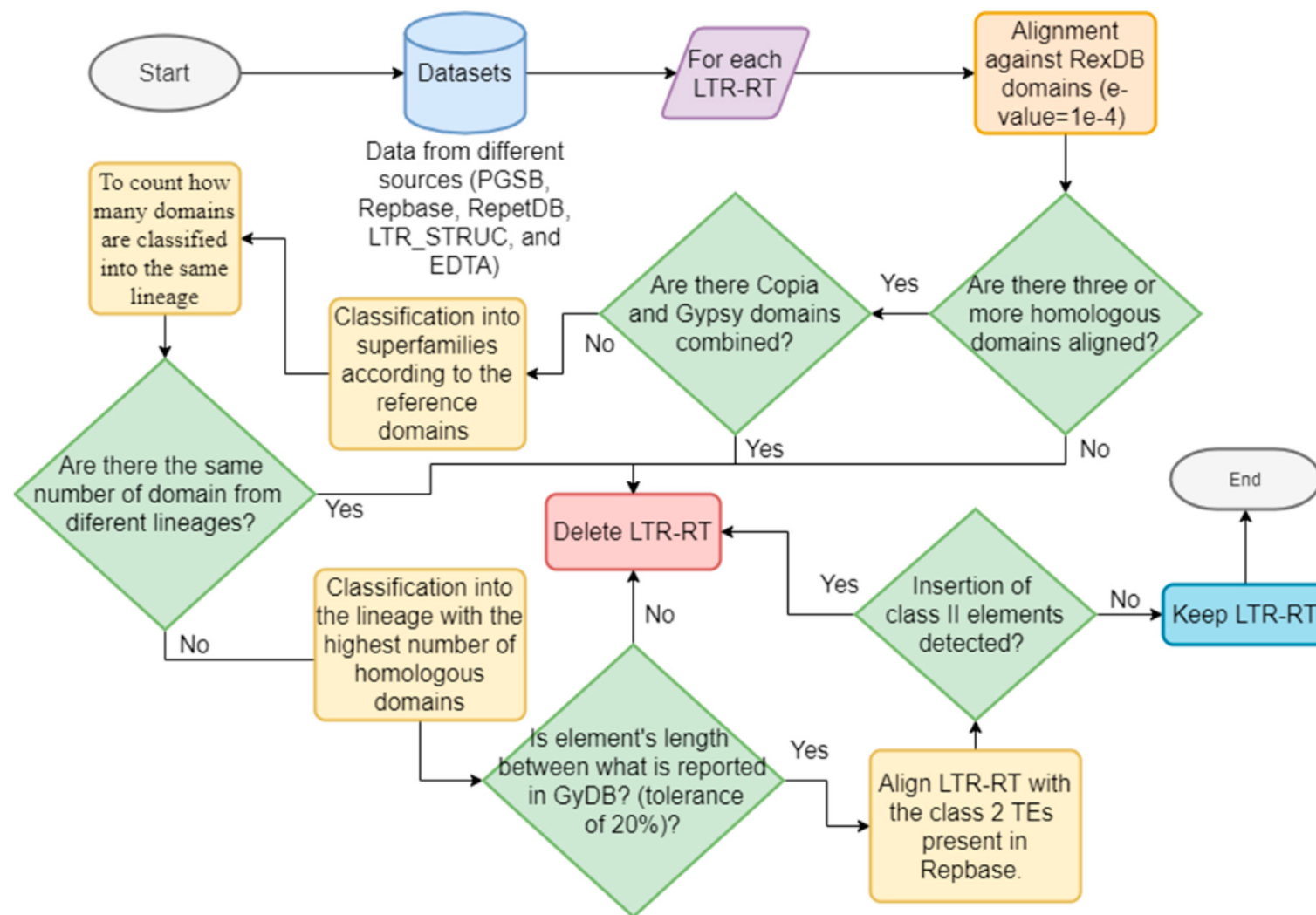


520,194 intact elements from 195 plant genomes

Databases –individual sequences–

An expert-reviewed and comprehensive database of elements : InpactorDB

<https://inpactordb.github.io/>



130,439 elements from 195 plant genomes from NCBI classified into Superfamily and family