

Analysis of Repeats on RepeatExplorer Galaxy Server

RepeatExplorer

Discover repeats in your next generation sequencing data



<https://repeatexplorer-elixir.cerit-sc.cz>

Laboratory of Molecular Cytogenetics

Jiri Macas

Pavel Neumann

Nina Hostakova

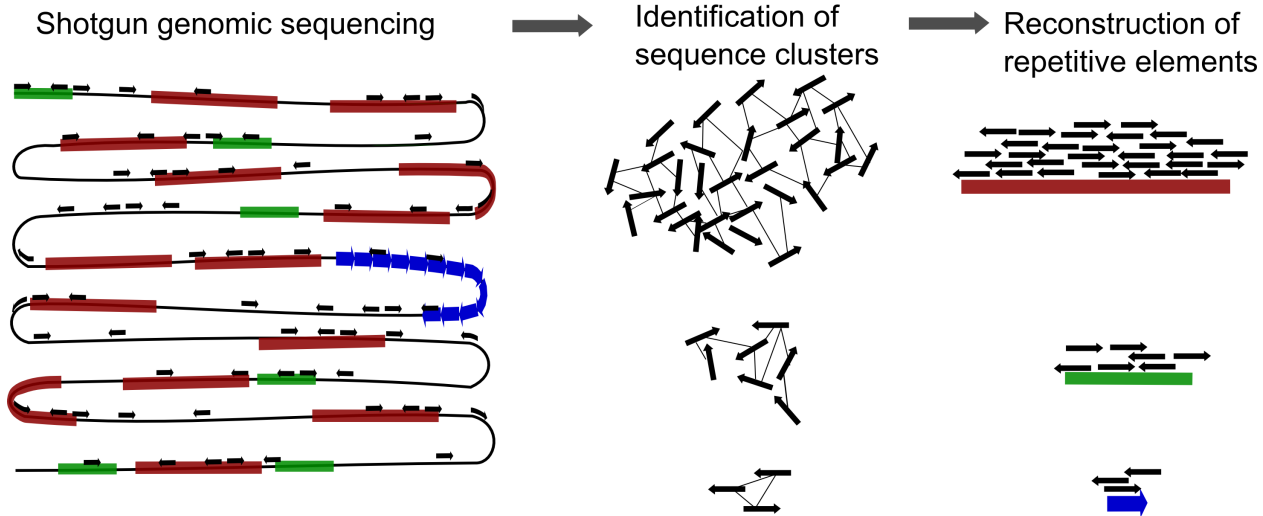
Petr Novak



Biology Centre
Czech Academy of
Sciences



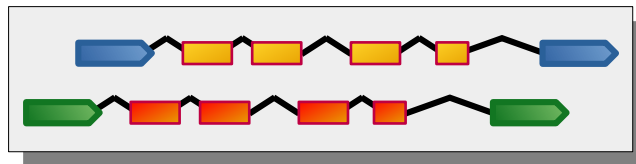
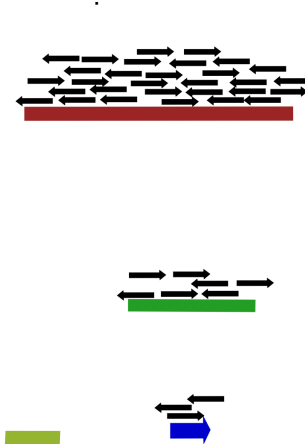
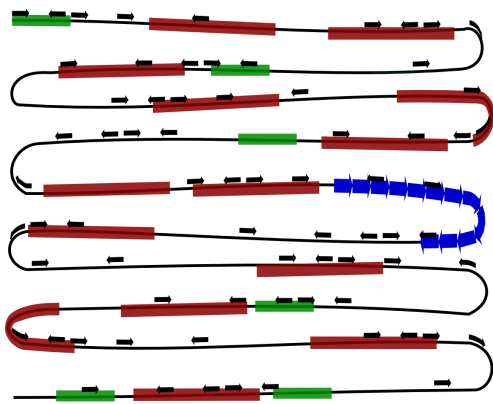
RepeatExplorer tools for genome annotation



Characterization of repeat from low-pass shogun sequencing

RepeatExplorer tools for genome annotation

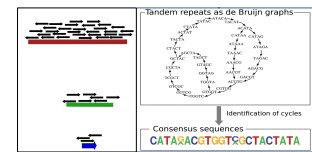
Genome Assembly



DANTE LTR



DANTE



Library based annotation

Complete assembly annotation workflow

- 1) Run RepeatExplorer on un-assembled sequence reads
- 2) Create library of repeats from RepeatExplorer output
- 3) Run RepeatMasker using RE library to annotate genome assembly
- 4) Annotate TE protein domains in assembly using DANTE
- 5) Run DANTE_LTR to identify full length LTR retrotransposons
- 6) Use intact LTR retrotransposons as second custom library for RepeatMasker

RepeatExplorer on un-assembled sequence reads

Galaxy Analyze Data Workflow Visualize Shared Data Admin Help User

Tools ☆ ↑

search tools ×

Get Data

RepeatExplorer2

TAREAN

ChIP-Seq Mapper

RepeatExplorer Utilities

DANTE

PROFREP

Text Manipulation

NGS: QC and manipulation

JBrowse

FASTA manipulation

Experimental Tools

Published Histories

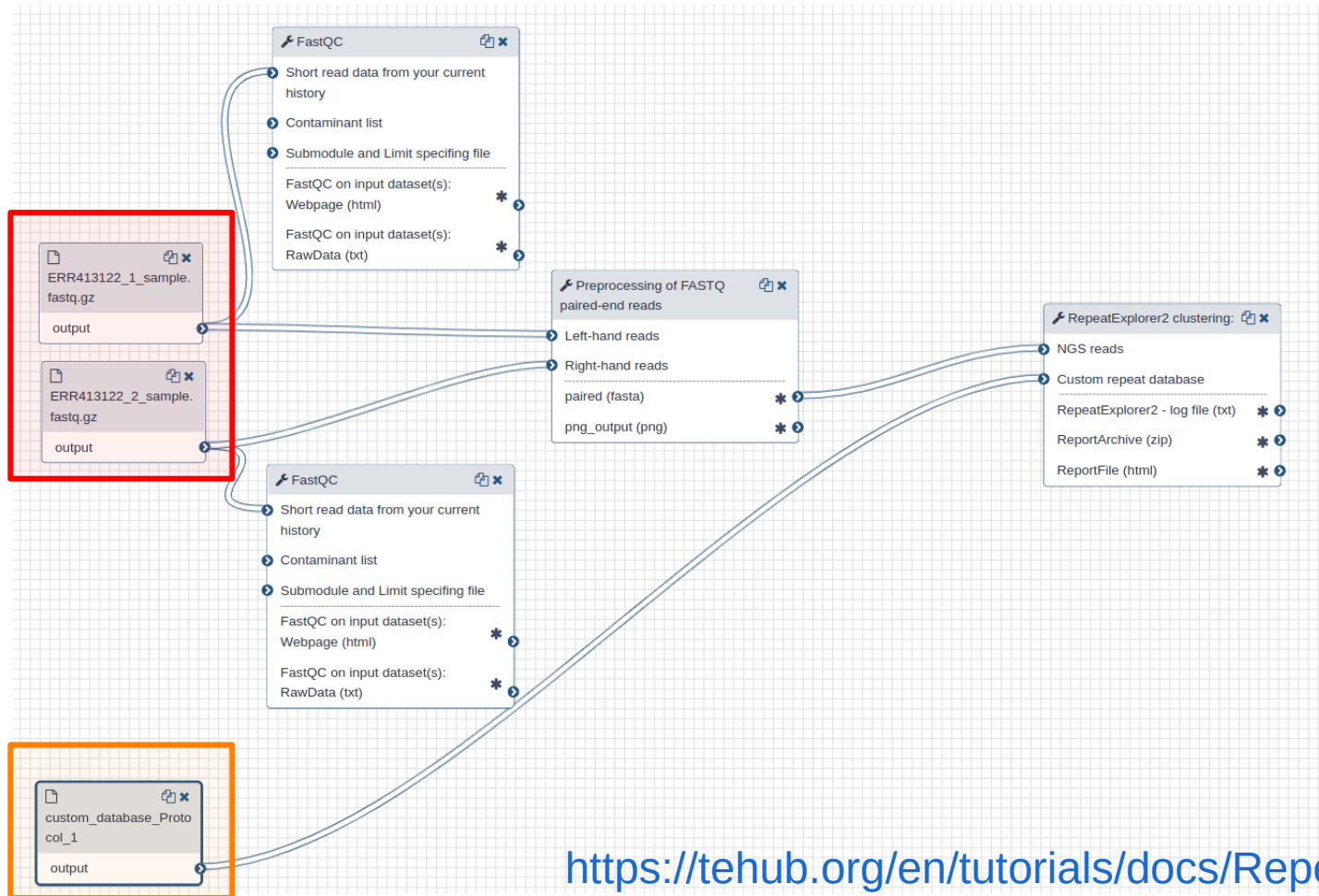
workshop search name, annotation, owner, and t Q

Advanced Search

| Name | Annotation | Pages | Community Rating | Community Tags | Last Updated |
|--|------------|-----------|------------------|--|----------------|
| Assembly annotation example | | kavonrtep | ☆☆☆☆☆ | uppsala workshop | 15 seconds ago |
| Workshop_2021_Protocol_1 [single species] | | jimacas | ☆☆☆☆☆ | workshop | May 09, 2022 |
| Elixir-cz workshop - complete results | | kavonrtep | ☆☆☆☆☆ | workshop elixir | Nov 04, 2021 |
| Workshop_2021_Protocol_2 [comparative analysis] | | jimacas | ☆☆☆☆☆ | workshop | May 10, 2021 |
| Workshop_2021_Protocol_4-ChIP-Seq | | kavonrtep | ☆☆☆☆☆ | workshop | May 06, 2021 |
| Workshop_2021_Protocol_3-TAREAN | | kavonrtep | ☆☆☆☆☆ | workshop | May 04, 2021 |
| workshop_TEdomains | | pavel | ☆☆☆☆☆ | | Feb 13, 2013 |

RepeatExplorer on un-assembled sequence reads

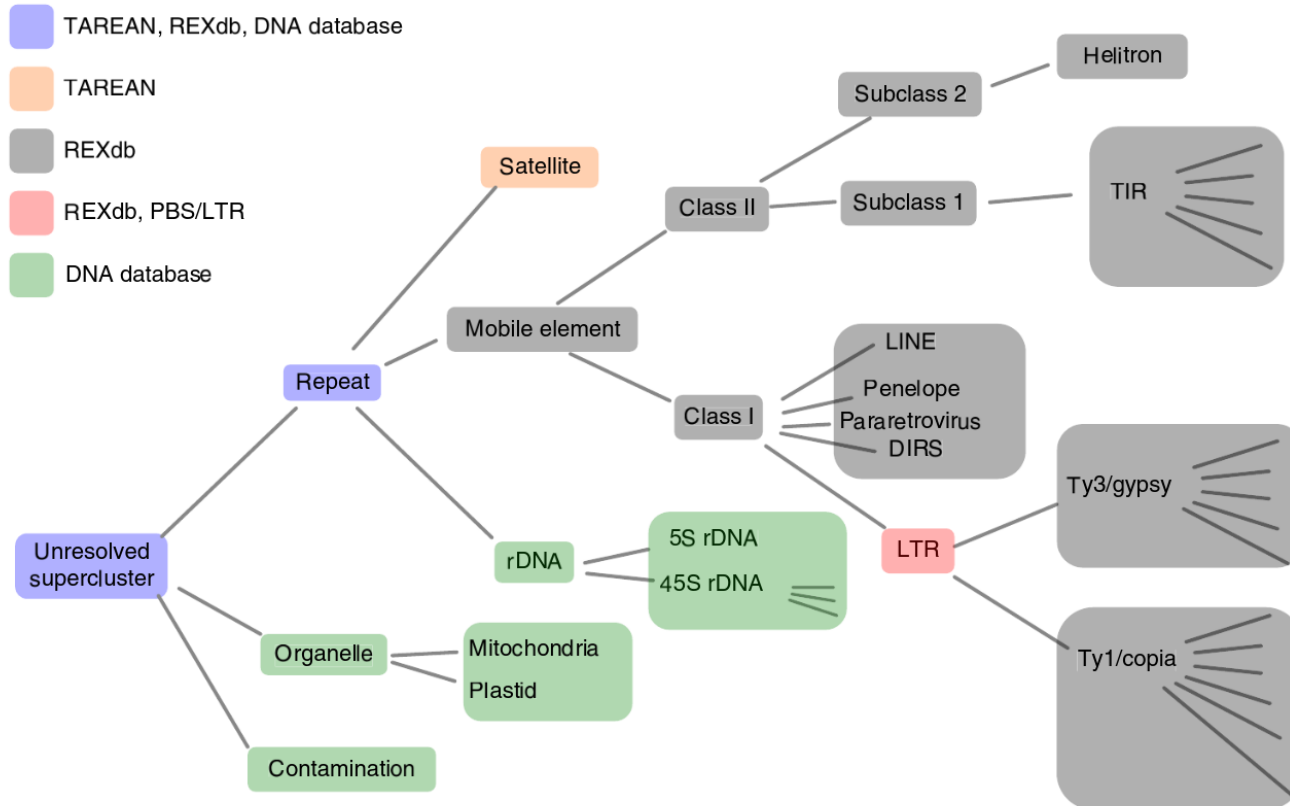
Paired-end reads



Custom database of repeats (optional)

RepeatExplorer output, annotation curation

<https://repeatexplorer-elixir.cerit-sc.cz/galaxy/u/jimacas/h/workshop2021protocol1-single-species-1>



Decision tree for automatic annotation

Annotation table

| Cluster | Supercluster | Size | Size_adjusted | Automatic_annotation | TAREAN_annotation | Final_annotation |
|---------|--------------|-------|---------------|--|--------------------------------------|--|
| 1 | 1 | 29329 | 29329 | All/repeat | Putative satellites (low confidence) | All/repeat/satellite |
| 2 | 1 | 8234 | 8234 | All/repeat | Other | All/repeat/satellite |
| 3 | 2 | 7977 | 7977 | All/repeat/rDNA/45S_rDNA | Other | All/repeat/rDNA/45S_rDNA |
| 4 | 3 | 6391 | 6391 | All/repeat/mobile_element/Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila | Other | All/repeat/mobile_element/Class_I/LTR/Ty3_gypsy/non-chromovirus/OTA/Athila |
| 5 | 2 | 6368 | 6368 | All/repeat/rDNA/45S_rDNA | Other | All/repeat/rDNA/45S_rDNA |
| 6 | 2 | 5853 | 5853 | All/repeat/rDNA/45S_rDNA | Other | All/repeat/rDNA/45S_rDNA |
| 7 | 4 | 5577 | 5577 | All/repeat/rDNA/5S_rDNA | rDNA | All/repeat/rDNA/5S_rDNA |
| 8 | 5 | 5434 | 5434 | All/organelle/plastid | Other | All/organelle/plastid |
| 9 | 6 | 5363 | 5363 | All/organelle/plastid | Other | All/organelle/plastid |
| 10 | 2 | 5254 | 5254 | All/repeat/rDNA/45S_rDNA | Other | All/repeat/rDNA/45S_rDNA |
| 11 | 2 | 4380 | 4380 | All/repeat/rDNA/45S_rDNA | Other | All/repeat/rDNA/45S_rDNA |
| 12 | 7 | 2786 | 2786 | All/repeat/mobile_element/Class_I/LTR/Ty3_gypsy/chromovirus/Tekay | Other | All/repeat/mobile_element/Class_I/LTR/Ty3_gypsy/chromovirus/Tekay |
| 13 | 8 | 2472 | 2472 | All/repeat/satellite | Putative satellites (low confidence) | All/repeat/satellite |
| 14 | 9 | 2347 | 2347 | All/organelle/plastid | Other | All/organelle/plastid |
| 15 | 10 | 2208 | 2208 | All/repeat/mobile_element/Class_II/Subclass_1/TIR/EnSpm_CACTA | Other | All/repeat/mobile_element/Class_II/Subclass_1/TIR/EnSpm_CACTA |
| 16 | 1 | 2201 | 2201 | All/repeat | Other | All/repeat/satellite |

Assembly annotation

The screenshot shows the Galaxy web interface. The top navigation bar includes 'Galaxy' and menu items like 'Analyze Data', 'Workflow', 'Visualize', 'Shared Data', 'Admin', 'Help', and 'User'. On the left, there is a 'Tools' sidebar with a search box and a list of tool categories: 'Get Data', 'RepeatExplorer2', 'TAREAN', 'CHIP-Seq Mapper', 'RepeatExplorer Utilities', 'DANTE', 'PROFREP', 'Text Manipulation', and 'NGS: QC and manipulation'. The main content area is titled 'Published Histories' and features a search bar with the text 'workshop' and a search icon. Below the search bar is an 'Advanced Search' section. A table lists several published histories with columns for Name, Annotation, Owner, Community Rating, Community Tags, and Last Updated. The first row, 'Assembly annotation example', is highlighted with a red box. The table data is as follows:

| Name | Annotation | Owner | Community Rating | Community Tags | Last Updated |
|---|------------|-----------|------------------|------------------|--------------|
| Assembly annotation example | | kavonrtep | ★★★★★ | Uppsala workshop | 6 hours ago |
| Workshop_2021_Protocol_1 [single species] | | jimacas | ★★★★★ | workshop | May 09, 2022 |
| Elixir-cz workshop - complete results | | kavonrtep | ★★★★★ | workshop elixir | Nov 04, 2021 |
| Workshop_2021_Protocol_2 [comparative analysis] | | jimacas | ★★★★★ | workshop | May 10, 2021 |
| Workshop_2021_Protocol_4-CHIP-Seq | | kavonrtep | ★★★★★ | workshop | May 06, 2021 |
| Workshop_2021_Protocol_3-TAREAN | | kavonrtep | ★★★★★ | workshop | May 04, 2021 |

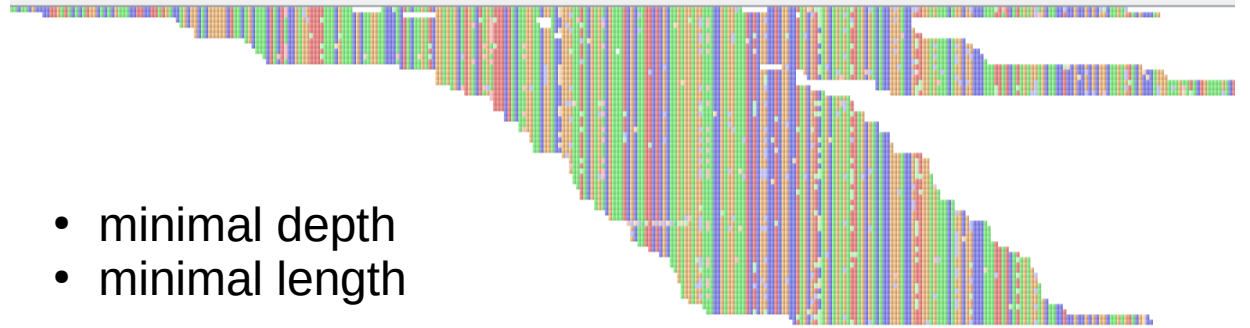
https://repeatexplorer-elixir.cerit-sc.cz/galaxy/histories/list_published

Library Based Repeat Annotation



- **Custom library**

Contigs from clustering results



TAREAN consensus

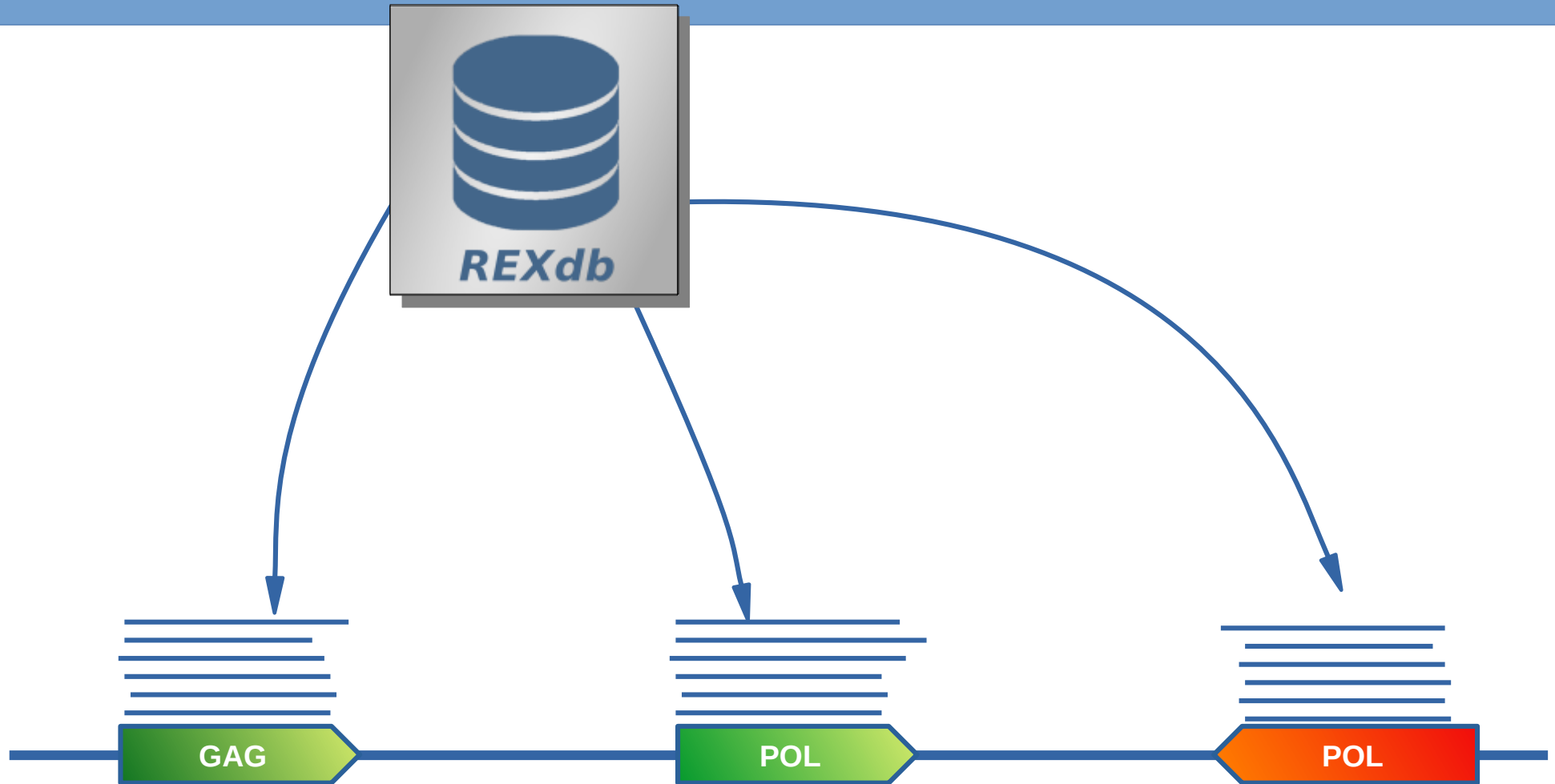
cAGTAGAATGATTTTCTATTCTTATGACTCTGGGAAAAATGGAACTGATTTTTGGAATATTTAGAGTCTAAAAAATTACATTTGAGAAATCTCAGA

- as dimers

Complete assembly annotation workflow

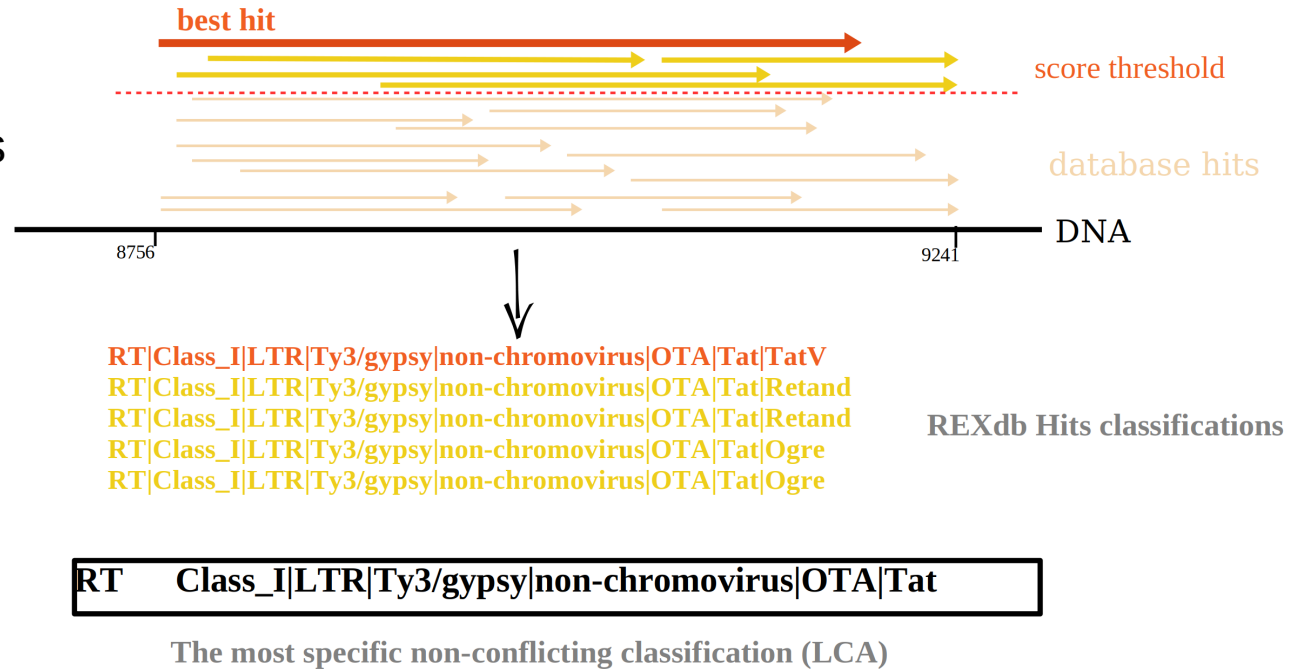
- 1) Run RepeatExplorer on un-assembled sequence reads
- 2) **Create library of repeats from RepeatExplorer output**
- 3) **Run RepeatMasker using RE library to annotate genome assembly**
- 4) **Annotate TE protein domains in assembly using DANTE**
- 5) **Run DANTE_LTR to identify full length LTR retrotransposons**
- 6) **Use intact LTR retrotransposons as second custom library for RepeatMasker**

DANTE – domain based annotation of transposable elements

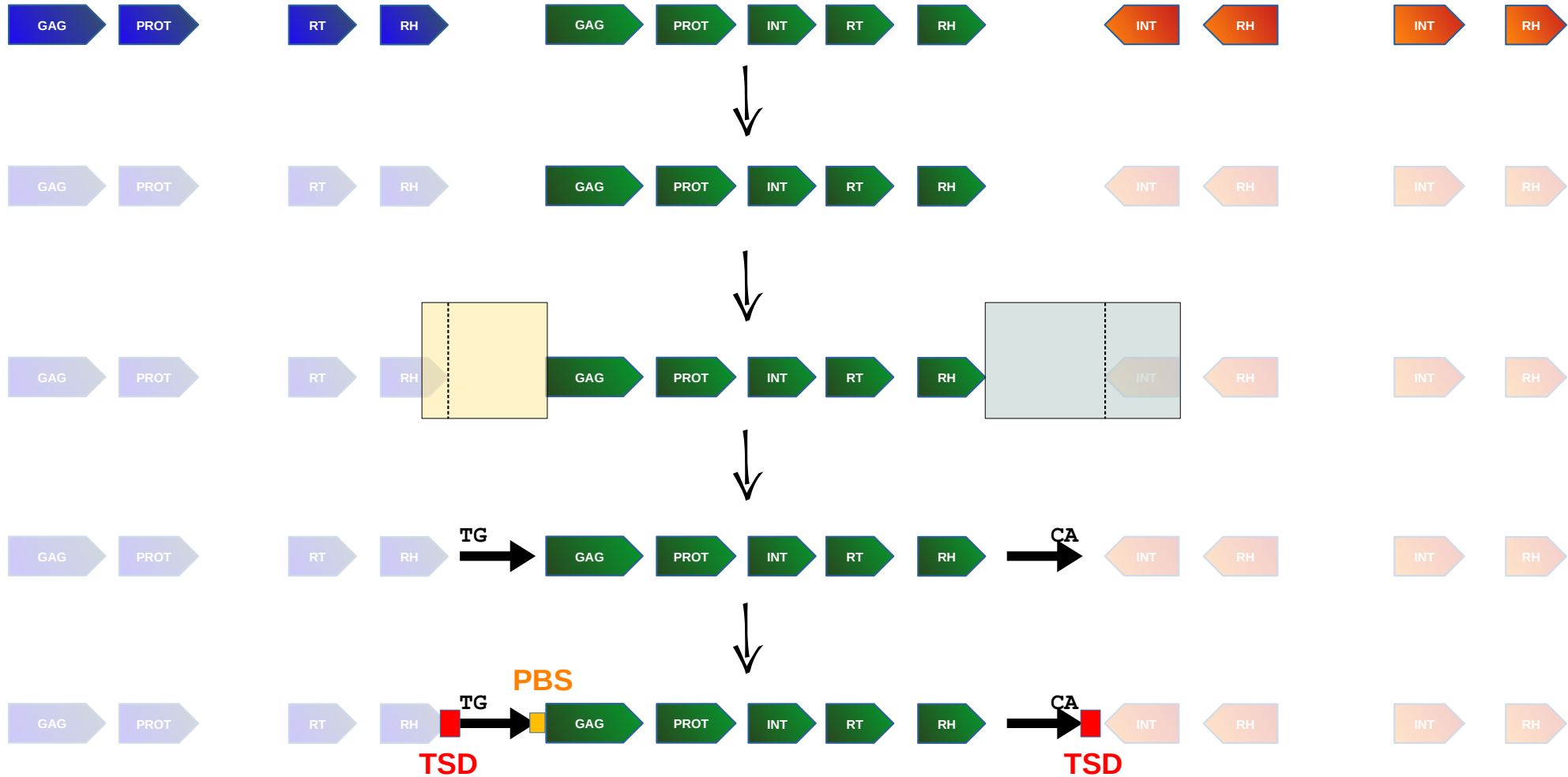


DANTE – domain based annotation of transposable elements

Resulting transposable elements classification is based on the multiple similarity hits against REXdb database



DANTE LTR - WORKFLOW



Final Assembly annotation



Custom
library
Annotation
(RE + RM)



DANTE
annotation



DANTE_LTR
annotation