



Search for domains

Create and manage lists

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

Amselem J, et al. RepetDB: a unified resource for transposable element references. *Mobile DNA*. 2019;10:6.

The screenshot displays the RepetDB web interface. At the top, there is a navigation bar with links for Home, Templates, Lists, QueryBuilder, Data Sources, API, MyMine, Contact Us, and Log in. A search bar on the right contains the text "e.g DHX*, GAG, Helitron" and a "GO" button. Below the navigation bar, the main content area is divided into three columns. The left column, titled "Search", contains a search form with a magnifying glass icon, instructions to search RepetDB for names, identifiers, or keywords, and a text input field containing "e.g. DHX*, GAG, Helitron". A green "SEARCH" button is at the bottom. The middle column, titled "Analyse", contains a form for entering a list of identifiers, a dropdown menu for "Consensus", and a text input field with "e.g.". A green "ANALYSE" button is at the bottom. The right column, titled "Welcome Back!", contains a welcome message and a green "TAKE A TOUR" button. Below these columns is a "Search consensus" section with various filters: "Taxon group" (All), "Wicker Classification" (All class, All order, All superfamily), "Potential chimeric" (checkboxes for Potential chimeric and Not chimeric), "Other elements" (Others...), and "Similarity feature(s)" (a text input field for protein profiles or transposon accessions). A "Search" button is at the bottom right of this section.

Request : Search for LINE R2 consensus with Reverse transcriptase + endonuclease domains

LINE	R2	RT EN	Variable	RIR	M
	RTE	APE RT	Variable	RIT	M
	Jockey	ORF1 APE RT	Variable	RIJ	M
	L1	ORF1 APE RT	Variable	RIL	P, M, F, O
	I	ORF1 APE RT RH	Variable	RII	P, M, F

1. Get the list of consensus with « reverse transcriptase » protein profile domain
2. Get the list of consensus with « endonuclease » protein profile domain
3. Intersection between list 1 & 2



Step1 : Query Builder → Select a Data Type to Begin a Query → Protein profile → Select



QueryBuilder

Advanced users can use a flexible query interface to construct their own data mining queries. The QueryBuilder lets you view the data model, apply constraints and select output. You can also export queries to share them with others.

- Browse data model
- Import query from XML
- View saved queries

Select a Data Type to Begin a Query

Click on a class name for a description or double click on a class name to create a new query starting at that class

- Ontology Term Synonym
- Organism
- Poly A
- Protein profile feature
- Protein Profile**
- Publication
- Rebase Transposon
- Ribosomal DNA
- Ribosomal DNA BlastN hit
- SSR

Select

ProteinProfile: undefined



Step2 : Protein profile → Description → Constraint : CONTAINS « reverse transcriptase »

Model browser

Browse through the classes and attributes. Click on **SUMMARY** links to add summary of fields to the results table or on **SHOW** links to add individual fields to the results. Use **CONSTRAIN** links to constrain a value in the query.

- Protein Profile **SUMMARY** **CONSTRAIN**
 - Description **SHOW** **CONSTRAIN**
 - Name **SHOW** **CONSTRAIN**
 - Accession **SHOW** **CONSTRAIN**
 - Short Name **SHOW** **CONSTRAIN**
- + Consensuses Consensus **SUMMARY** **CONSTRAIN**
- + Cross References Cross Reference **SUMMARY** **CONSTRAIN**
- + Data Sets Data Set **SUMMARY** **CONSTRAIN**
- + Located Features Location

Query Overview

no fields constrained

Constraint logic: no constraints

Constraint for Protein Profile > Description

Constrain

Choose a filter

Filter query results on this field having a specific value

Protein Profile > Description

CONTAINS reverse transcriptase Add to query

OR Contained in list: IN GAG Add to query

Filter query results on this field having any value or not

Has no value Has a value Add to query

Show empty fields

Fields selected for output



Step 3 : Consensus → Consensus identifier → show

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

Show results

Model browser

Browse through the classes and attributes. Click on **SUMMARY** links to add summary of fields to the results table or on **XSDM** links to add individual fields to the results. Use **CONSTRAIN** links to constrain a value in the query.

- Protein Profile **SUMMARY** **CONSTRAIN**
 - Description **SHOW** **CONSTRAIN**
 - Name **SHOW** **CONSTRAIN**
 - Accession **XSDM** **CONSTRAIN**
 - Short Name **SHOW** **CONSTRAIN**
 - Consensus Consensus **SUMMARY** **CONSTRAIN**
 - TE classification code **SHOW** **CONSTRAIN**
 - Copies Integer **SHOW** **CONSTRAIN**
 - Cumulative genome coverage Integer **SHOW** **CONSTRAIN**
 - Fragments Integer **SHOW** **CONSTRAIN**
 - Full-length copies Integer **SHOW** **CONSTRAIN**
 - Full-length fragments Integer **SHOW** **CONSTRAIN**
 - Length Integer **XSDM** **CONSTRAIN**
 - Miscellaneous classification **SHOW** **CONSTRAIN**
 - Potential Chimeric Boolean **XSDM** **CONSTRAIN**
 - Consensus identifier** **SHOW** **CONSTRAIN**
 - Classification Wicker Classif term **SUMMARY** **CONSTRAIN**
 - Copy statistics Statistic **SUMMARY** **CONSTRAIN**
 - Data Set Data Set **SUMMARY** **CONSTRAIN**
 - Data Sets Data Set **SUMMARY** **CONSTRAIN**

Show empty fields

Query Overview

Protein Profile

- Description **X**
 - CONTAINS reverse transcriptase **X** **(A)**
 - Consensus Consensus collection **X** **(A)**
 - Consensus identifier** **X**

Constraint logic: one constraint

Fields selected for output

Columns to Display

Use the **SHOW** or **SUMMARY** links to add fields to the results table. Click button below to sort results in ascending **↑** or descending **↓** order.

Protein Profile > Consensus > Consensus identifier **X**

(no description) **↑**

Show results

Step 4 : Save a List → Protein profile consensus → List name « Consensus with reverse transcriptase domain »

Trail: Query > Results

Manage Columns

Manage Filters

Manage Relationships

Save as List

Generate Python code

Export

Showing 1 to 25 of 5,274 rows

1

Protein Profile > Consensus (5,274 Consensus)

Protein Profile (8 Protein Profiles)

Pick items from the table

Create List

Add to List

Consensus

Consensus identifier

chim_VvTEdenovoV2-B-R15883-Map5

chim_VvTEdenovoV2-B-R15898-Map5

DHX-comp-chim_MCL17_Tpar_TEdenovo-B-R1678-Map5_reversed

DHX-incomp-chim_MCL178_Alyr_TEdenovo-B-R643-Map9_reversed

DHX-incomp-chim_MCL278_Alyr_TEdenovo-B-R441-Map17

DHX-incomp-chim_MCL52_Alyr_TEdenovo-B-G3929-Map18

DHX-incomp-chim_VvTEdenovoV2-B-R17442-Map10

DHX-incomp-chim_Zmv3_300M-B-R1209-Map5_reversed

DHX-incomp_MCL278_Alyr_TEdenovo-B-P692.237-Map3

DHX-incomp_MCL52_Alyr_TEdenovo-B-G4563-Map3

DTX-hAT_denovoMDO_kr-B-R4007-Map8_reversed

DTX-incomp_3h_itr1_300M.R-G24172-Map4_reversed

Create a new List of 5,274 Consensus

List Name

Consensus with reverse transcriptase domain

Optional attributes

List Description

Enter a description

NO TAGS

Add a new tag

add

Close

3
Create List



Relaunch Steps 1 to 4 to search for consensuses with endonuclease domain

Step 1 : Query Builder → Select a Data Type to Begin a Query → Protein profile → Select

Step 2 : Protein profile → Description → Constraint : CONTAINS « endonuclease »

Step 3 : Consensus → Consensus identifier → show

Step 4 : Save a List → Protein profile consensuses → List name « Consensus with endonuclease domain »



Step 5 : Lists

- Select “Consensus with reverse transcriptase domain”
- Select “Consensus with endonuclease domain”
- intersect
- Enter a new List name : “RT + EN” → Save



Lists

View your own and public lists, search by keyword and compare or combine the contents of lists. Click on a list to view graphs and summaries in an analysis page, select lists using checkboxes to perform set operations. Click 'Upload' above to import a new list.

Filter: Filter: ☆ NY -- filter by a tag -- Reset

Actions: Union | Intersect | Subtract | Asymmetric Difference | Options: Show descriptions Show Tags

- Consensus with endonuclease **Intersect** 8 Consensuss
- Consensus with reverse transcriptase domain ☆ 5274 Consensuss

Intersect x

Enter a new List name:

RT+EN Save



List Analysis for RT+EN (369 Consensuss)

Manage Columns Manage Filters

Manage Relationships

Save as List

Generate Python code

Export

External Links
No external links.

Showing rows 1 to 25 of 369

Rows per page: 25

Consensus identifier	Consensus Length	Consensus TE classification code	Consensus Miscellaneous classification	Consensus Potential Chimeric	Consensus Cumulative genome coverage	Consensus Fragments	Consensus Full-length fragments	Consensus Copies	Consensus Full-length copies
DTX-incomp-chim_3b_itr1_300M-L-B1034-Map1_reversed	11461	Class II : TIR : ?		true	524702	923	1	852	1
DTX-incomp-chim_VvTedenovoV2-B-R9898-Map16	10331	Class II : TIR : ?		true	81120	134	2	107	2
DTX-incomp-chim_Zmv3_300M-B-R1440-Map10_reversed	15409	Class II : TIR : ?		true	720328	2782	1	2554	1
RIX_Bgh_Satine2	6014	Class I : LINE : ?		false	427933	463	4	288	14
RIX_BgraDH14-B-G4001-Map5_reversed	1315	Class I : LINE : ?		false	34122	64	3	58	3
RIX_BgraDH14-B-G4016-Map4	1224	Class I : LINE : ?		false	32081	77	2	74	2
RIX_BgraDH14-B-G4066-Map3_reversed	1462	Class I : LINE : ?		false	45075	89	2	76	3
RIX_BgraDH14-B-G4211-Map3	1305	Class I : LINE : ?		false	60022	125	3	115	4
RIX_BgraDH14-B-G4797-Map3	1784	Class I : LINE : ?		false	59420	107	4	93	4
RIX_BgraDH14-B-G4981-Map3_reversed	2451	Class I : LINE : ?		false	37866	85	1	66	2
RIX_BgraDH14-B-G5093-Map5_reversed	2434	Class I : LINE : ?		false	24574	57	1	52	1

Lists
- Select and display "RT + EN"

Result : A LINE consensus



Consensus : RIX_BgraDH14-B-G4211-Map3

Length	1305 pb FASTA	TE classification code	Class I : LINE : ?
Potential Chimeric	false	Classification	Class I : LINE
Cumulative genome coverage	60022 pb	Fragments	125
Full-length fragments	3	Copies	115
Full-length copies	4		

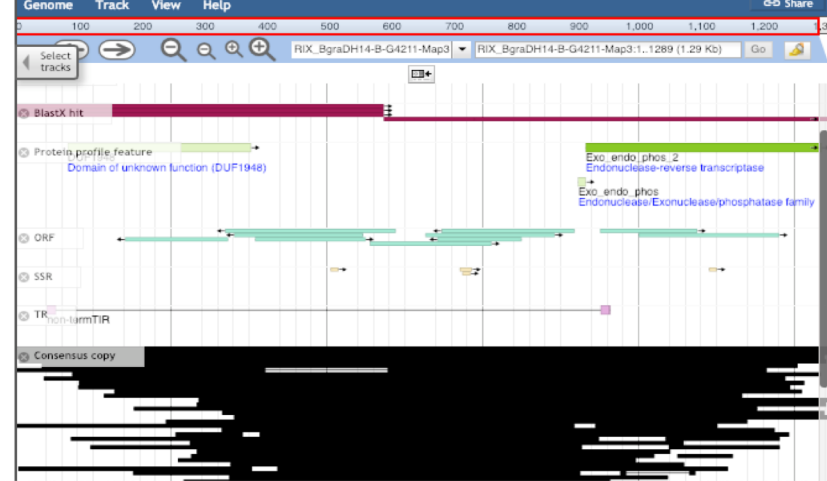
Material and Method

Organism	Blumeria graminis f. sp. hordei DH14	Genome assembly	Name: v3-contig sequences, http://www.blugen.org/index.php?page=data
Software used	TEdenovo v1.3 TEannot v2.1 PastedClassifier v1.0		
Comments	<p>The TEdenovo was launched on Blumeria graminis hordei genome contigs (10726 sequences > 1250 bp, Assembly Version 3.0) provided a TE consensus library of 2251 sequences.</p> <p>After a first TEannot using this TE consensus library, we filtered out consensus sequences without at least one full length copy in the genome. The 1465 remaining TE consensus were used to launched a second TEannot pipeline (v2.1) to annotate the genome. Using the result of this second annotation, these TE consensus were manually filtered from chimeras, duplicates to obtain a library of 733 consensus sequences used to launch a third TEannot.</p> <p>We provide here a new refined classification (v2017) with PastedClassifier (v1.0) to obtain the TE consensus library of B. graminis hordei inserted in RepetDB.</p> <p>Genome annotation files and TE consensus ID mapping between previous and new classification are available for download.</p>		
Contact	Joelle Anselem (mail to urgi-contact)		
2 Publications			

Consensus copy statistics

	Mean	Standard deviation	Minimum	Q25	Median	Q75	Maximum
Copies coverage over consensus	40.00 %	26.52 %	1.53 %	18.24 %	36.02 %	56.25 %	100.38 %
Copies length	521.97 pb	346.05 pb	20 pb	238.00 pb	470.00 pb	734.00 pb	1310 pb
Identity	81.96 %	9.26 %	65.32 %	74.91 %	78.53 %	92.12 %	100.00 %

Features browser



Lists

- This Consensus is in 4 lists:
- RT+EN (369)
 - Consensus List (List of consensus containing int: (11083))
 - Consensus with endonuclease domain (828)
 - Consensus with reverse transcriptase domain (5274)
- Add this Consensus to one of your lists:
- DIRS*
-

External Links

No external links.

Result

- A chimeric consensus : LINE + LTR Gypsy

Consensus : RIX-chim_TmeI28-B-G14973-Map7_reversed

Length	16195 pb FASTA...	TE classification code	Class I : LINE : ?
Potential Chimeric	true	Classification	Class I : LINE
Cumulative genome coverage	1318137 pb	Fragsments	499
Full-length fragments	1	Copies	293
Full-length copies	14		

Material and Method

Organism	Tuber melanosporum	Genome assembly	Name: ASM15164v1 Accession: GCA_000151645.1
Software used	TEdenovo v1.3 TEannot v2.1 PastedClassifier v1.0		
Comments	<p>The TEdenovo (v1.3) pipeline launched on Tuber melanosporum (398 scaffolds) provided a TE consensus library of 2597 sequences. After a first TEannot using this TE consensus library, we filtered out consensus sequences without at least one full length fragment in the genome.</p> <p>The 905 remaining TE consensus were used to launched a second TEannot pipeline (v2.1) to annotate the genome.</p> <p>We provide here a new refined classification (v2017) with PastedClassifier (v1.0) to obtain the TE consensus library of T. melanosporum inserted in RepetDB. This TE library has not been manually curated. It is provided "as is", i.e. a direct output of the REPET pipeline.</p> <p>Genome annotation files and TE consensus ID mapping between previous and new classification are available for download.</p>		

Contact [Joelle Anselem \(mail to urgi-contact\)](#)

[2 Publications](#)

Consensus copy statistics

	Mean	Standard deviation	Minimum	Q25	Median	Q75	Maximum
Identity	86.00 %	7.20 %	65.16 %	79.12 %	86.68 %	90.51 %	97.83 %
Copies length	4902.13 pb	4939.44 pb	21 pb	430.00 pb	2022.00 pb	8197.00 pb	16224 pb
Copies coverage over consensus	27.80 %	30.50 %	0.13 %	2.66 %	12.49 %	50.61 %	100.18 %

Lists

This Consensus is in 7 lists:

- RT+EN (369)
- Consensus List (List of consensus containing int (11083)
- List of consensus containing GAG (9500)
- Consensus List of consensus containing rnaesh (7203)
- RT (7529)
- Consensus with endonuclease domain (828)
- Consensus with reverse transcriptase domain (5274)

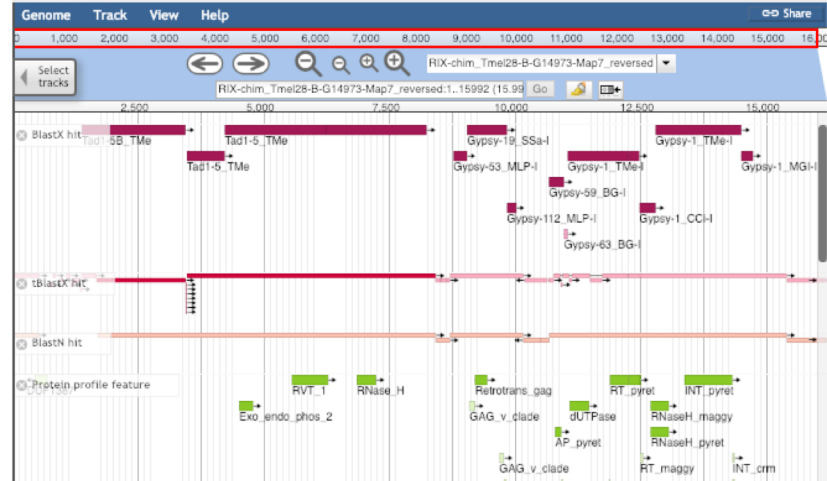
Add this Consensus to one of your lists:

DIRS*

External Links

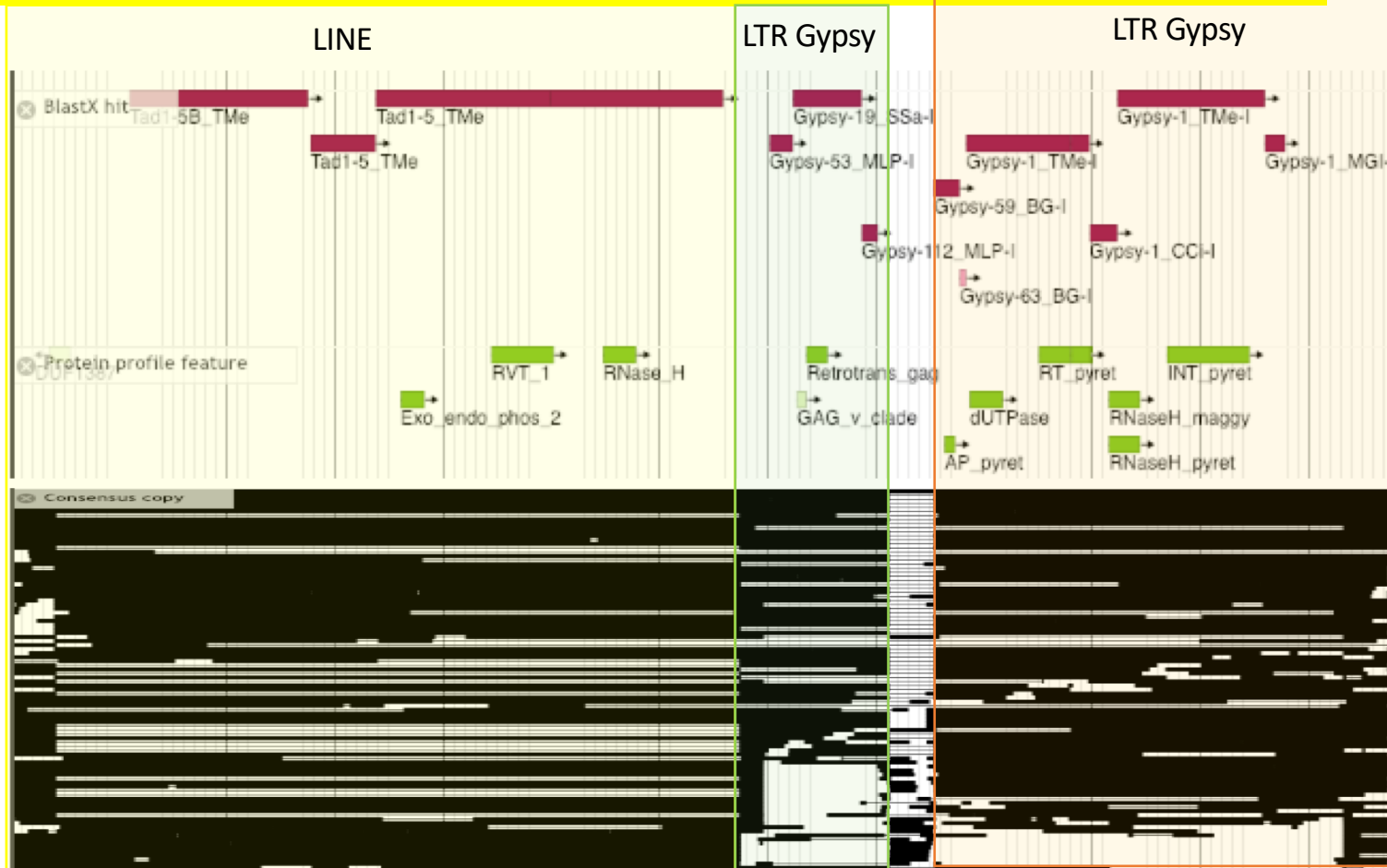
No external links.

Features browser





Zoom on Similarity evidence + consensus genome copies



Request : Search for classI LTR complete consensus

1. Get the list of consensus with « GAG » protein profile domain
2. Get the list of consensus with « Asp, AP ... » protein profile domain
3. Get the list of consensus with « Integrase, INT...» protein profile domain
4. Get the list of consensus with « RT, reverse transcriptase ... » protein profile domain
5. Get the list of consensus with « RNaseH » protein profile domain
6. Get the list of consensus with LTR structure
7. Intersection between list 1,2,3,4,5

Step1 : Query Builder → Select a Data Type to Begin a Query → Protein profile → Select

Step2 : Protein profile → Description → Constraints

Step 3 : Consensus → Consensus identifier → show

Once you get the table, you can browse and check the different Profile description

Protein Profile

Description

CONTAINS GAG (A)

!= GAGA binding protein-like family (B)

!= GAGA factor (C)

!= GAGE protein (D)

Consensuses Consensus collection

Consensus identifier

Constraint logic: A and B and C and D

A and B and C and D


Trail: Query > Results

Manage Columns Manage Filters Manage Relationships

Showing 1 to 25 of 6,002 rows

Protein F	Consensuses
Description	Consensus identifier
gag gene	RIX-incomp_MCL449_Tp
gag gene	RLC_BgraDH14-B-R178
gag gene	RLG_BgraDH14-B-G559
gag gene	RLX-incomp_3b_itr1_30k
gag-polyp	chim-PotentialHostGene
gag-polyp	chim_VvTEdenovoV2-B-I
gag-polyp	chim_VvTEdenovoV2-B-I
gag-polyp	chim_VvTEdenovoV2-B-I
gag-polyp	DXH-incomp-chim_deno
gag-polyp	DXH-incomp-chim_deno
gag-polyp	DTX-incomp-chim_3b_It
gag-polyp	DTX-incomp-chim_3b_It
gag-polyp	DTX-incomp-chim_3b_It
gag-polyp	DTX-incomp-chim_3b_It
gag-polyp	DTX-incomp-chim_3b_It
gag-polyp	DTX-incomp-chim_3b_It
gag-polyp	DTX-incomp-chim_3b_It

11 Protein Profile Descriptions



Protein Profile Description	Count
Retrotransposon gag protein	2,333
GAG-pre-integrase domain	1,608
gag-polypeptide of LTR copia-type	1,582
gag-polyprotein putative aspartyl protease	491
GAG-polyprotein viral zinc-finger	27
Spumavirus gag protein	9
Gag polyprotein, inner coat protein p12	5
gag gene protein p17 (matrix protein)	2
gag gene protein p24 (core nucleocapsid protein)	2
Retroviral GAG p10 protein	2

Step 4 : Save a List → Protein profile consensus → List name « Consensus with GAG domain »

Step1 : Query Builder → Select a Data Type to Begin a Query → Protein profile → Select

Step2 : Protein profile → Description → Constraints

Step 3 : Consensus → Consensus identifier → show

Once you get the table, you can browse and filter the domains you don't want

Query Overview

Protein Profile

- Description
 - CONTAINS Aspartyl (A)
- Name
 - CONTAINS AP_ (B)
- Consensuses Consensus collection
 - Consensus identifier

Constraint logic: A or B

A or B

Protein Profile Description	Consensuses Consensus identifier
Adenylate cyclase associated (CAP) C terminal	RLX-comp_denovoPyrus-B-R2838-Map3
Adenylate cyclase associated (CAP) N terminal	DTX-incomp-chim_3b_itr1_300M-B-R3247-Map16_reversed
Adenylate cyclase associated (CAP) N terminal	noCat_denovoFragaria-B-R2123-Map19
Adenylate cyclase associated (CAP) N terminal	RIX-incomp_MCL1599_Brap_TEdenovo-B-R1848-Map3
Adenylate cyclase associated (CAP) N terminal	RLX-comp_3b_itr1_300M-L-B147-Map1
Adenylate cyclase associated (CAP) N terminal	RLX-comp-chim_3b_itr1_300M-B-R449-Map20_reversed
Adenylate cyclase associated (CAP) N terminal	RLX-comp-chim_Zmv3_300M-L-B3969-Map1_reversed
Adenylate cyclase associated (CAP) N terminal	RLX-comp_Zmv3_300M-L-B10104-Map1_reversed
Adenylate cyclase associated (CAP) N terminal	RLX-comp_Zmv3_300M-L-B3895-Map1
Adenylate cyclase associated (CAP) N terminal	RLX-comp_Zmv3_300M-L-B9119-Map1
Adenylate cyclase associated (CAP) N terminal	RLX-incomp_3b_itr1_300M-B-G23218-Map3
Adenylate cyclase associated (CAP) N terminal	RLX-incomp_3b_itr1_300M-B-G23487-Map20
Adenylate cyclase associated (CAP) N terminal	RLX-incomp_3b_itr1_300M-L-B1360-Map1
Adenylate cyclase associated (CAP) N terminal	RLX-incomp_3b_itr1_300M-L-B1655-Map1_reversed
Adenylate cyclase associated (CAP) N terminal	RLX-incomp-chim_Zmv3_300M-L-B2241-Map1
A-kinase anchor protein 110 kDa (AKAP 110)	RLX-incomp-chim_Zmv3_300M-L-B6793-Map1

108 Protein Profile Names

133 Items Selected

Filter values

- AP_endonuc_2 7
- AP_galea 7
- AP_pepsins_A1a 7
- AP_bel 6
- AP_microopia_mdg3 6
- AP_sinbad 6
- Asp 6
- FSAP_sig_propep 6
- Herpes_DNAp_acc 6
- Jnk-SapK_ap_N 6
- Alpha-2-MRAP_C 5
- AP_cog5550 5

Filter

Step 4 : Save a List → Protein profile consensuses → List name « Consensus with AP domain »



3. Get the list of consensus with « Integrase, INT...» protein profile domain
4. Get the list of consensus with « RT, reverse transcriptase ... » protein profile domain
5. Get the list of consensus with « RNaseH » protein profile domain

For each search :

Browse and check description and/or name to filter those not specific of Class I LTR elements



Get the list of consensus with LTR structure

Step1 : Query Builder → Select a Data Type to Begin a Query → Consensus structural feature → Select

Step2 : Type → Constraints

Step 3 : Consensus → Consensus identifier → show You can filter on termLTR or not

Step 4 : Save a List → Consensus structural features → Consensus → List name « Consensus with termLTR domain »

Consensus Structural Feature Type	Count
<input type="checkbox"/> LTR	28,350
<input checked="" type="checkbox"/> termLTR	25,347



7. Intersection between list 1,2,3,4,5,6

Note that the results could be under representative of the Complete LTR because one of the domain could not have similarity with PFAM or Gypsy DB known domains

If you select only termLTR, you also could miss LTR found at at more than 10 base from the 5' or 3' end