PLMview

Preferentially **L**ocated **M**otif View for Plants

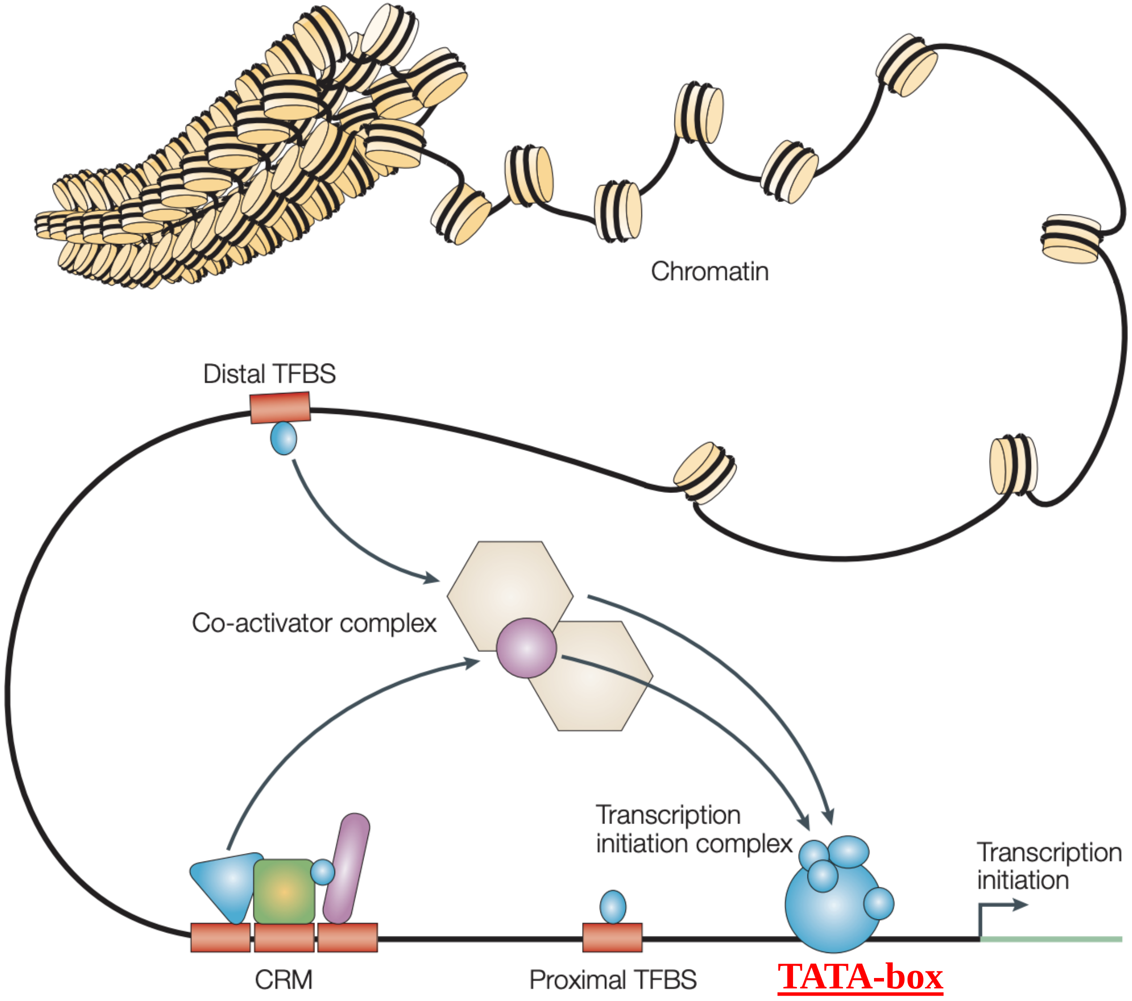
PLMview a database and web interface for PLMdetect, an *in silico* tool to predict cis-regulatory motifs

Véronique Brunaud IPS2- GNet

Journées du PEPI-IBIS, 15-16 oct 2025

IPS2
Institute of Plant Sciences
Paris - Saclay

Transcriptional regulation and cis-regulatory motifs

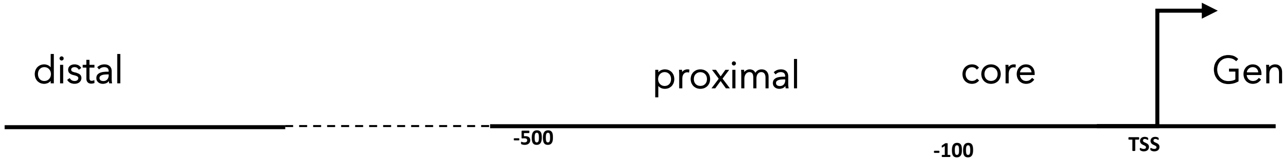


from Wasserman and Sandelin, 2004

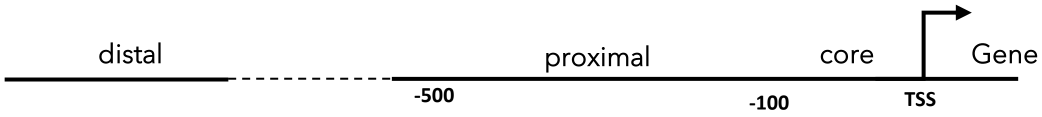
Cis-regulatory motifs are 4 to 15 bases sequences bound by Transcription Factors (TF)

PLMdetect – find enrichment of cis-regulatory element in proximal regions of genes

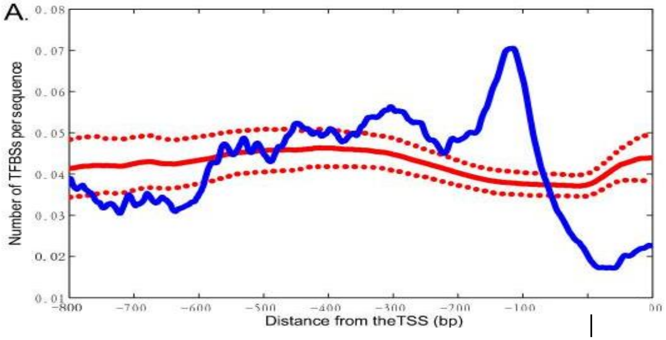
Hypothesis: gene-proximal regions enriched with cis-regulatory motifs, evolutionary conservation



PLMdetect – find enrichment of cis-regulatory element in proximal regions of genes

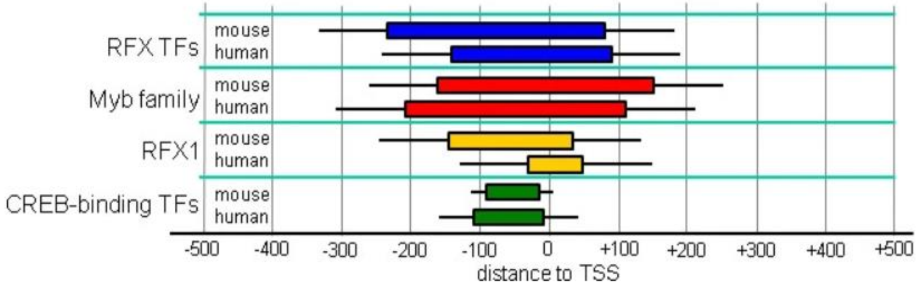


Why focus on this proximal region?



TSS

Chez la levure dès 2004 (Harbison *et al.*)
et en 2010 Lin *et al.* (extraction figure)

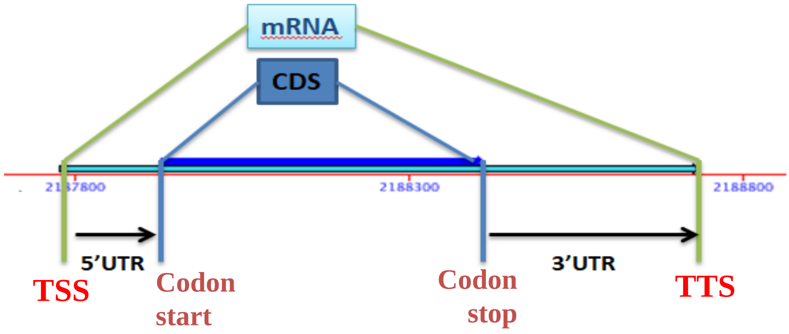


Article de 2013 Vandenbon *et al.* chez la souris et l'humain

In Arabidopsis, Yu et al. 2016 confirms this enrichment: 86% of TFBSs in the [-1000,+200] region and 63% in the [-400, +200] region.

PLMdetect method

1- Extract gene-proximal sequences



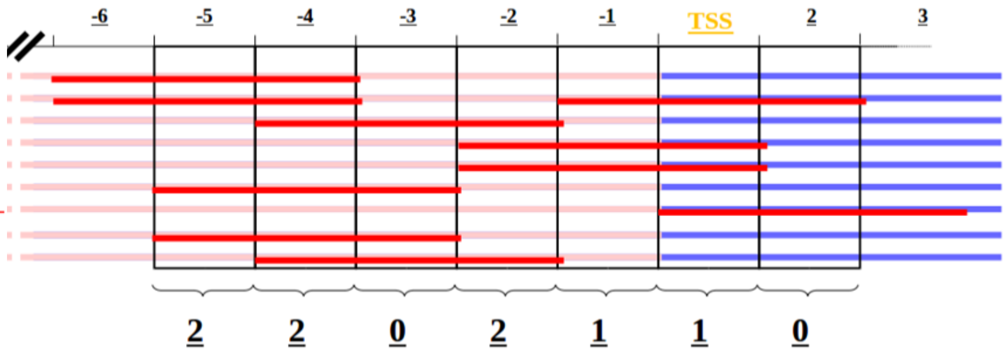
- ✓ UTR required
- ✓ [-1000,+500] TSS
- ✓ [-500,+1000] TTS

- Applied to coding genes
- Need UTR annotations
- Not all regions processed, only utr defined

2- Align sequences on TSS/TTS

Motif size = 3 bases

Sliding window size = 1 base

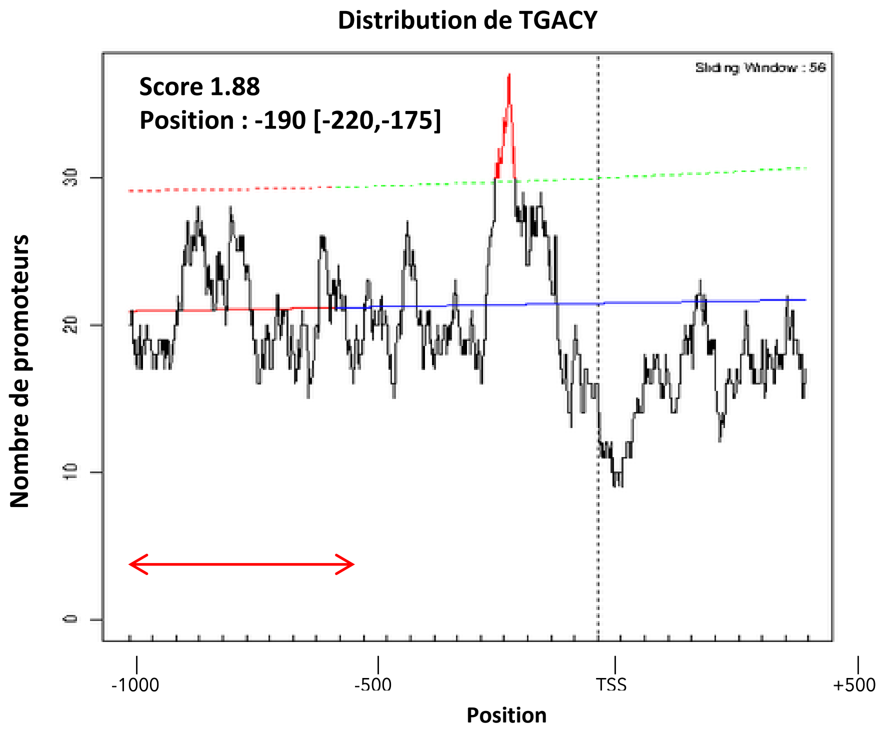


3- Calculate motif frequency

PLM Preferentially Located Motif

4- Detect a motif-enriched region ?

- Use the region [-1000, -500] of sequences to define the linear average frequency of each motif
- Calculate a confidence interval (0.99)
- A peak is defined if > confidence interval



Distribution de TGACY

PLM TGACY (W-box) defined by

- ✓ Preferential Position -190
- ✓ Enriched region [-220,-175]
- ✓ Gene list with PLM in the enriched region
- ✓ Score 1.88

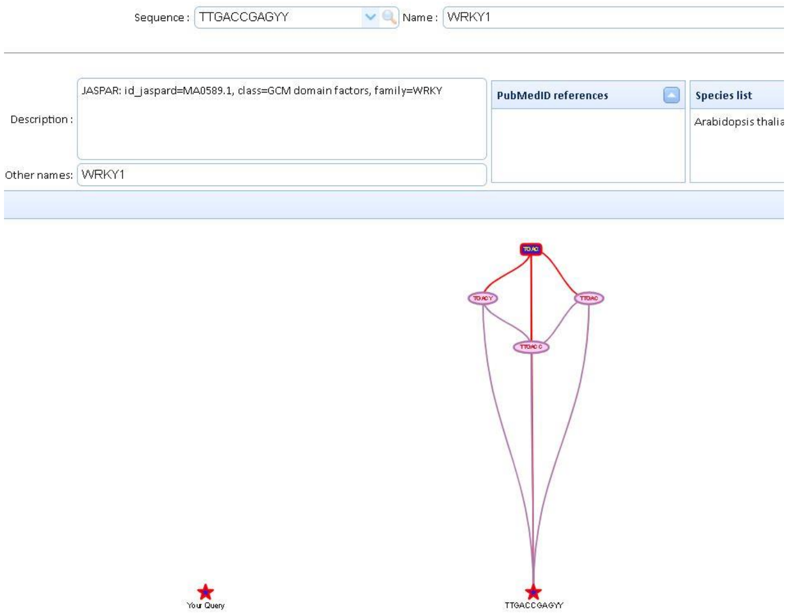
PLMdetect - INPUT

Gene-proximal regions (5' or 3')

de novo DNA motifs

OR

Known TFBS



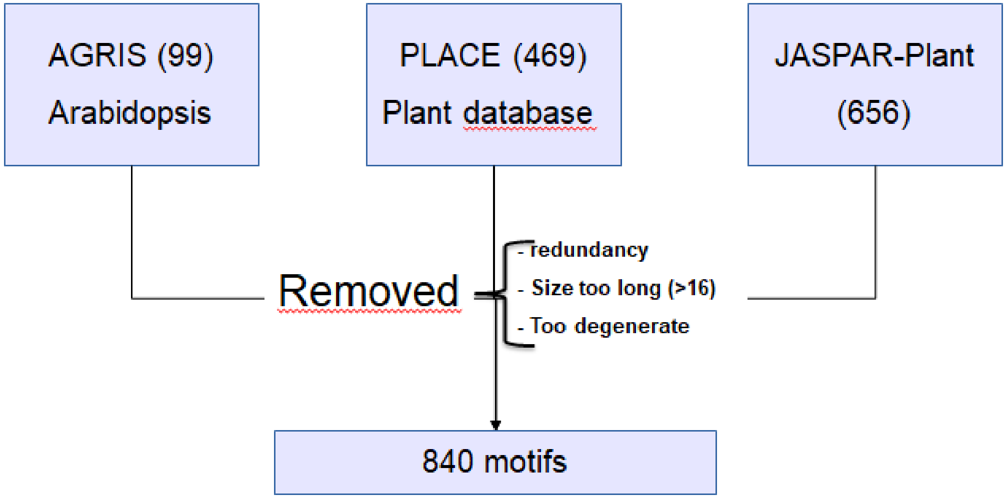
Sequence : TTGACCGAGYY

Name : WRKY1

Description : JASPAR: id_jaspard=MA0589.1, class=GCM domain factors, family=WRKY

Other names: WRKY1

Species list: Arabidopsis thalia



AGRIS (99)
Arabidopsis

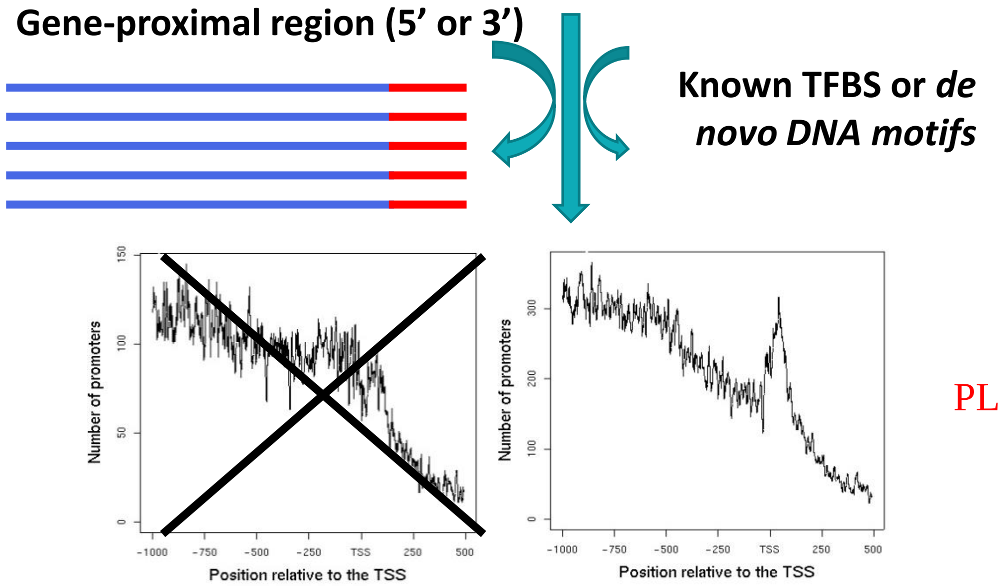
PLACE (469)
Plant database

JASPAR-Plant
(656)

Removed
- redundancy
- Size too long (>16)
- Too degenerate

840 motifs

PLMdetect - Usage


Gene-proximal region (5’ or 3’)

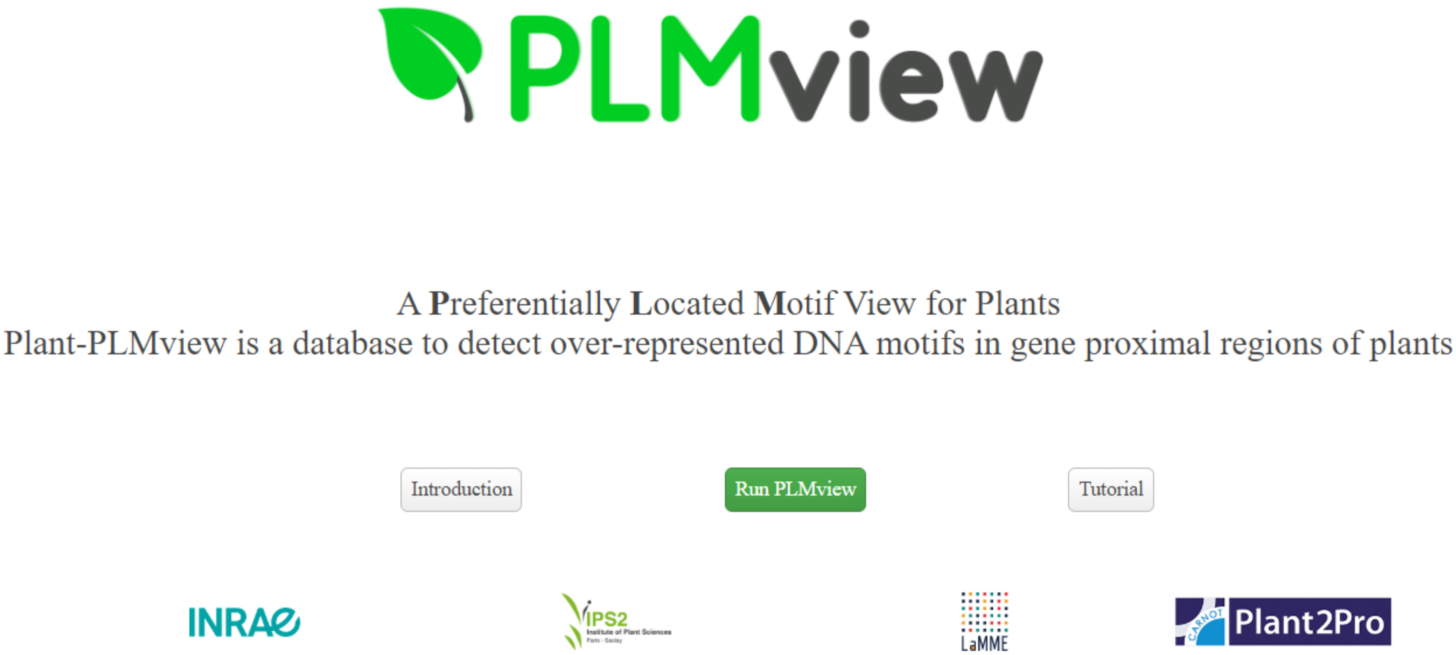
Known TFBS or *de novo DNA motifs*

PLM ?

- ✓ Apply at genome-wide in order to list and characterize PLM in the proximal regions of the genes (Roziere et al. 2022) – applied in Maize and Arabidopsis genomes
- ✓ Apply to a list of differentially expressed genes to identify TFs that regulate these genes (5 publications since 2014)

⇨ PLMdetect not easy to use and analyze the results → PLMView

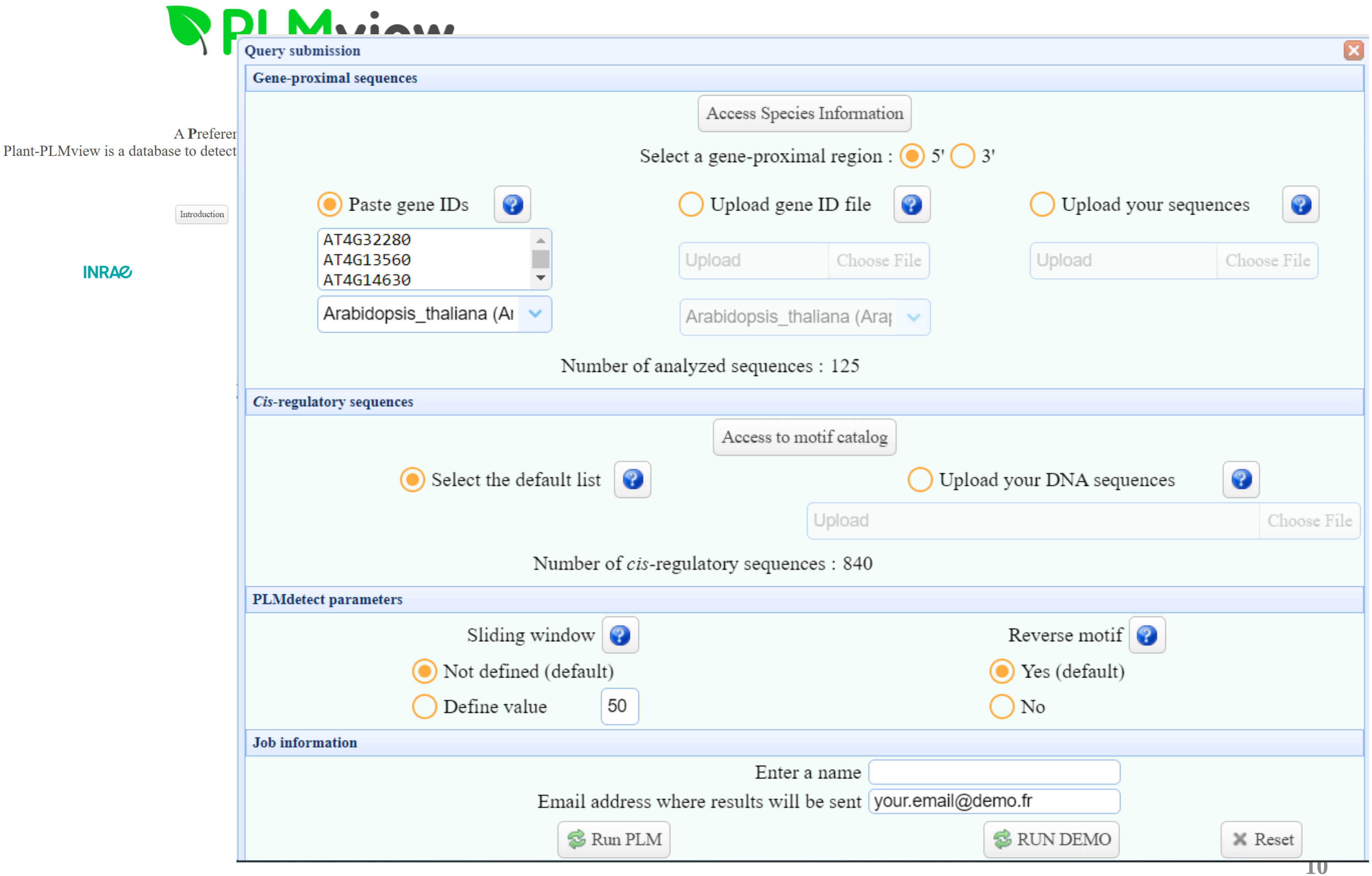
PLMview a web application for PLMdetect
http://plmview.ips2.universite-paris-saclay.fr/



A **P**referentially **L**ocated **M**otif View for Plants

Plant-PLMview is a database to detect over-represented DNA motifs in gene proximal regions of plants

PLM Inputs: GeneID by species



Query submission

Gene-proximal sequences

Select a gene-proximal region : ● 5' ○ 3'

● Paste gene IDs

AT4G32280
AT4G13560
AT4G14630

Arabidopsis_thaliana (Ar

○ Upload gene ID file

Arabidopsis_thaliana (Ara

○ Upload your sequences

Number of analyzed sequences : 125

***Cis*-regulatory sequences**

● Select the default list

○ Upload your DNA sequences

Number of *cis*-regulatory sequences : 840

PLMdetect parameters

Sliding window

● Not defined (default)

○ Define value 50

Reverse motif

● Yes (default)

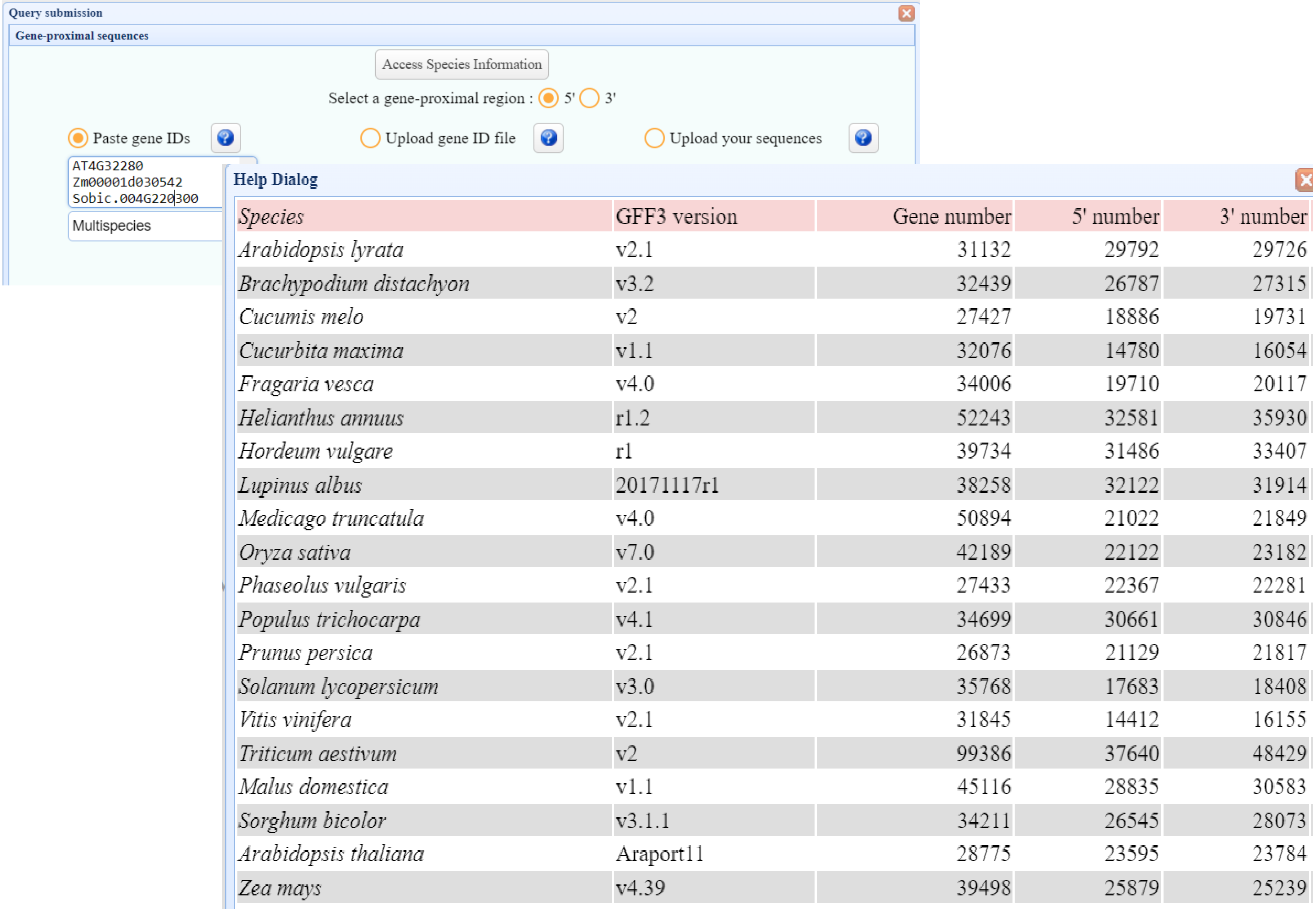
○ No

Job information

Enter a name

Email address where results will be sent your.email@demo.fr

PLM Inputs: GeneID by species or multispecies (20 species)



Query submission

Gene-proximal sequences

Select a gene-proximal region : ● 5' ○ 3'

● Paste gene IDs ○ Upload gene ID file ○ Upload your sequences

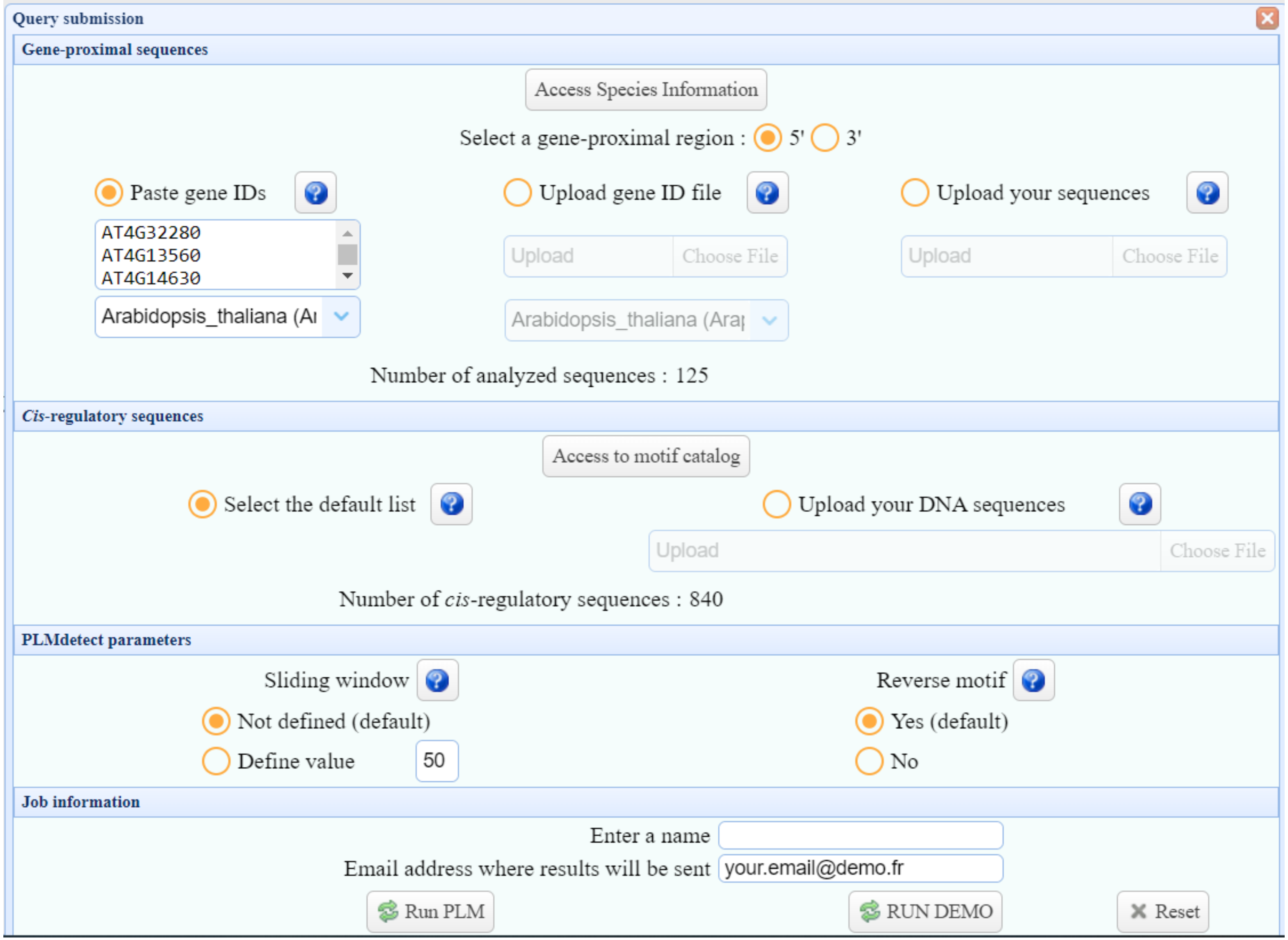
```
AT4G32280
Zm00001d030542
Sobic.004G220300
```

Multispecies

Help Dialog

Species	GFF3 version	Gene number	5' number	3' number
Arabidopsis lyrata	v2.1	31132	29792	29726
Brachypodium distachyon	v3.2	32439	26787	27315
Cucumis melo	v2	27427	18886	19731
Cucurbita maxima	v1.1	32076	14780	16054
Fragaria vesca	v4.0	34006	19710	20117
Helianthus annuus	r1.2	52243	32581	35930
Hordeum vulgare	r1	39734	31486	33407
Lupinus albus	20171117r1	38258	32122	31914
Medicago truncatula	v4.0	50894	21022	21849
Oryza sativa	v7.0	42189	22122	23182
Phaseolus vulgaris	v2.1	27433	22367	22281
Populus trichocarpa	v4.1	34699	30661	30846
Prunus persica	v2.1	26873	21129	21817
Solanum lycopersicum	v3.0	35768	17683	18408
Vitis vinifera	v2.1	31845	14412	16155
Triticum aestivum	v2	99386	37640	48429
Malus domestica	v1.1	45116	28835	30583
Sorghum bicolor	v3.1.1	34211	26545	28073
Arabidopsis thaliana	Araport11	28775	23595	23784
Zea mays	v4.39	39498	25879	25239

PLM Inputs: Gene ID, Motif list (default list or custom list)



Query submission

Gene-proximal sequences

Access Species Information

Select a gene-proximal region : ● 5' ○ 3'

● Paste gene IDs ○ Upload gene ID file ○ Upload your sequences

AT4G32280
AT4G13560
AT4G14630

Arabidopsis_thaliana (Ar

Upload | Choose File

Upload | Choose File

Arabidopsis_thaliana (Arap

Number of analyzed sequences : 125

***Cis*-regulatory sequences**

Access to motif catalog

● Select the default list ○ Upload your DNA sequences

Upload | Choose File

Number of *cis*-regulatory sequences : 840

PLMdetect parameters

Sliding window

● Not defined (default)
○ Define value 50

Reverse motif

● Yes (default)
○ No

Job information

Enter a name

Email address where results will be sent your.email@demo.fr

Run PLM | RUN DEMO | Reset

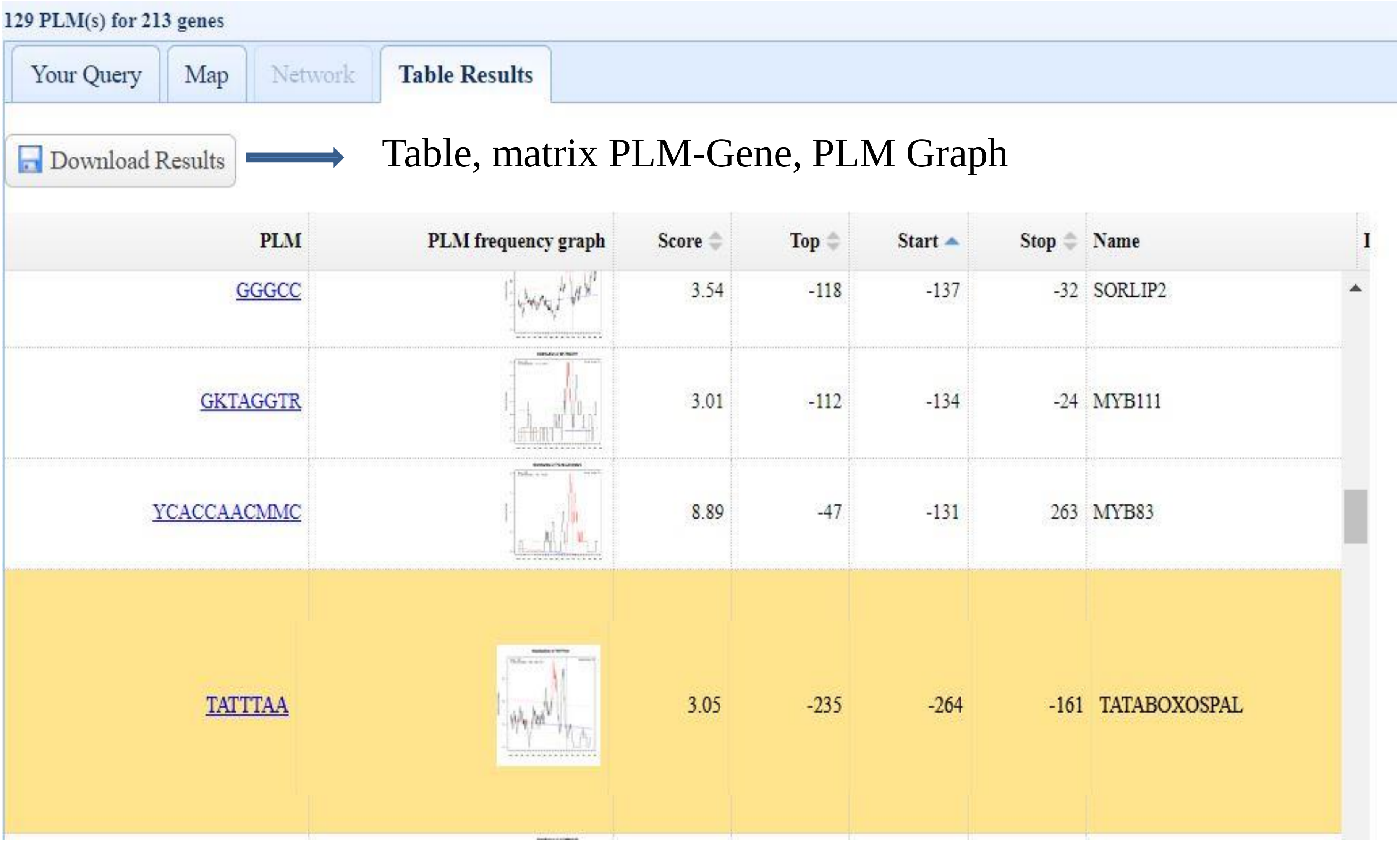
→ Mail with link to access and play with your results (kept one week)

→Table of results and map visualization

	TATAWA	7.91	-43	-27	-36	TATA-box	
	TATTTA	4.67	-151	24	-57	TATA-like	
	RMNATGGC	4.66	-43	407	50	TRANSINITMONOCOTS	Context sequence of transla initiation codon in monocot M=A/C
	TAAATA	4.2	-133	19	-41	TATA-like	
	TGACG	4.13	18	436	168	ASF1MOTIFCAMV	ASF-1 binding site in CaM promoter; ASF-1 binds to t motifs; See S000023 (AS1) HBP-1 binding site of whea gene; TGACG motifs are fo promoters and are involved transcriptional activation of by auxin and/or salicylic ac
	CCGTCG	3.88	1				

Your result of your PLM request, found 40 motif(s)

PLMview Table Results 129 PLM



129 PLM(s) for 213 genes

Table, matrix PLM-Gene, PLM Graph

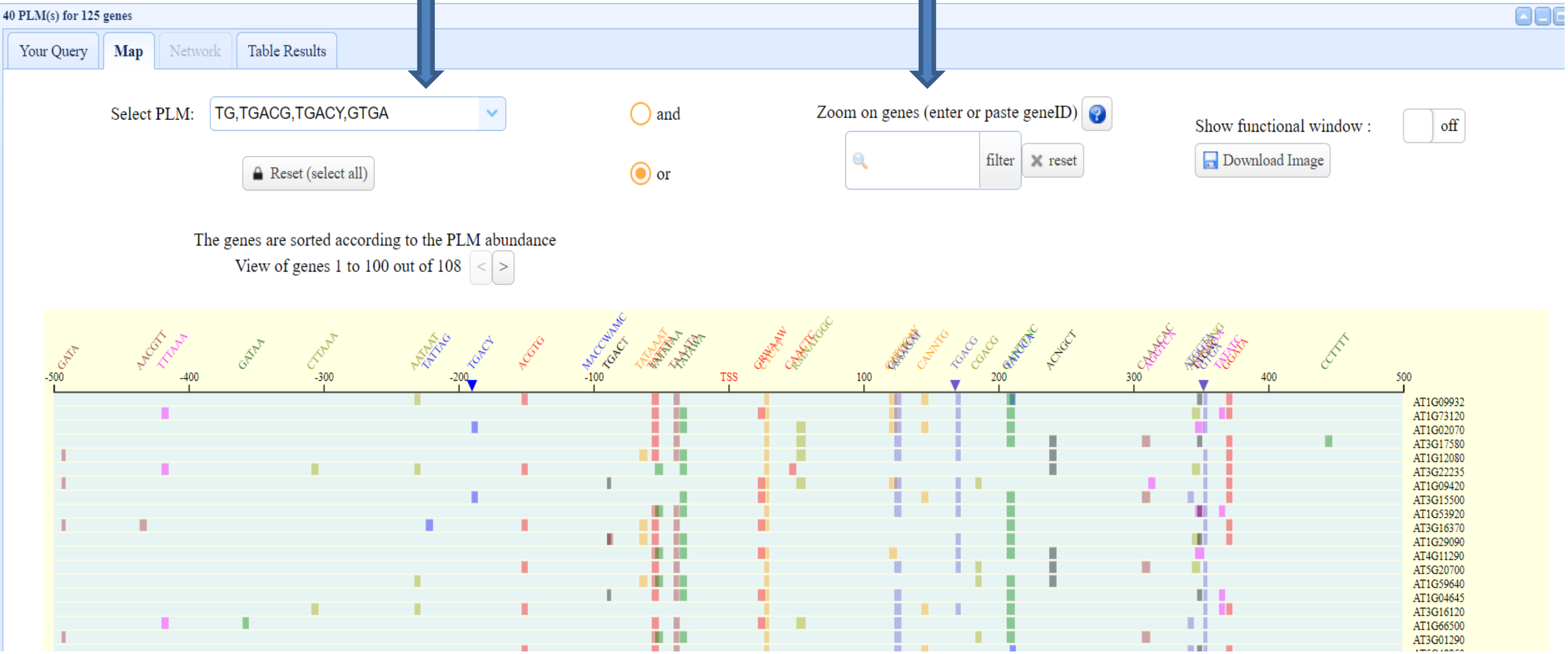
PLM	PLM frequency graph	Score	Top	Start	Stop	Name	I
GGGCC		3.54	-118	-137	-32	SORLIP2	
GKTAGGTR		3.01	-112	-134	-24	MYB111	
YCACCAACMMC		8.89	-47	-131	263	MYB83	
TATTTAA		3.05	-235	-264	-161	TATABOXOSPAL	

→ Resume of PLM characteristics and distributions

PLMview PLM map view

Select PLM (or, and)

Select GeneID



40 PLM(s) for 125 genes

Your Query | **Map** | Network | Table Results

Select PLM: TG,TGACG,TGACY,GTGA

○ and

● or

Zoom on genes (enter or paste geneID)

Show functional window : off

The genes are sorted according to the PLM abundance

View of genes 1 to 100 out of 108

PLMview - specificity

- ❑ Twenty Plant gene-proximal regions
  - Processed 3’ region
  - Select gene-proximal regions with UTR defined
- ❑ Multi-species access
- ❑ Download results with list of genes associated to each PLM
- ❑ PLM description (if information is present in our database)
- ❑ Map visualisation
- ❑ Web interface to play with your favorite genes!

Preferentially **L**ocated **M**otif View for Plants

GNet (IPS2) : Cécile Guichard, Julien Rozière (PhD), Véronique Brunaud

LaMMe (UEVE) : Margot Correa, Franck Samson

http://plmview.ips2.universite-paris-saclay.fr/

PLMview a database and web interface for PLMdetect tool