

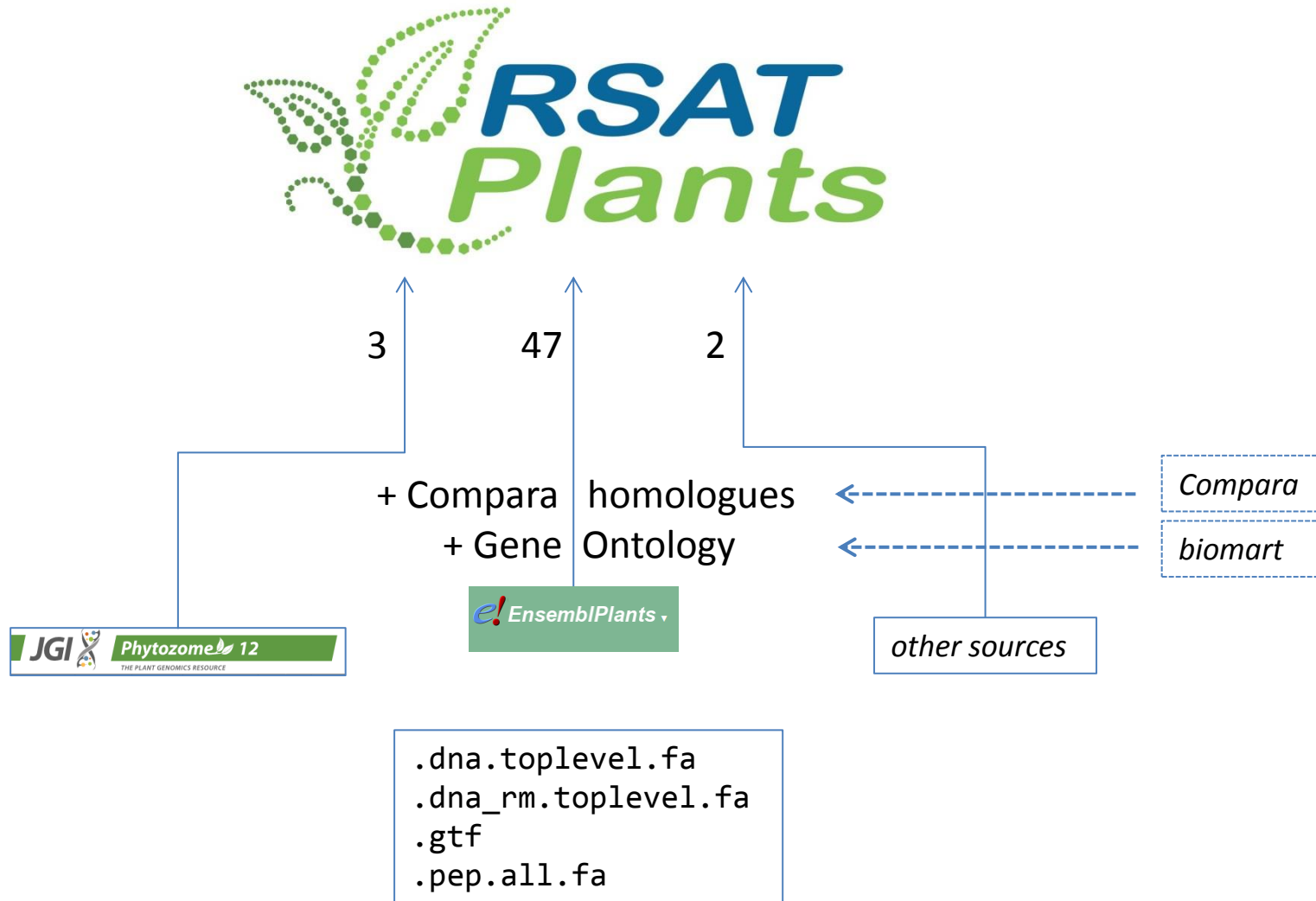
footprintDB & RSAT::Plants

RSAT mirror for plants integrated with DB of transcription factors of known DNA motifs

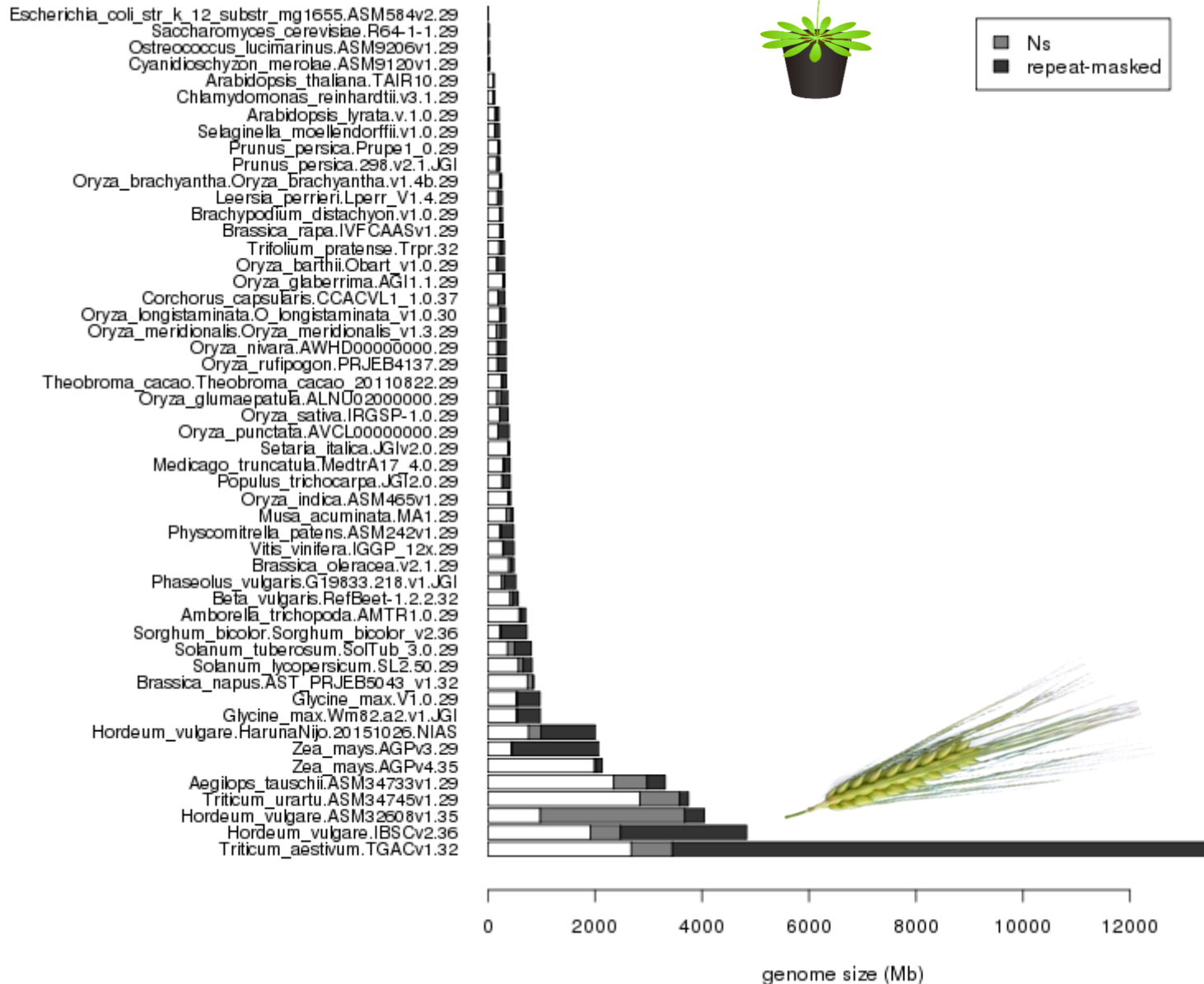
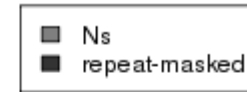
B Contreras-Moreira, A Sebastián, **J van Helden**

Estación Experimental de Aula Dei-CSIC ,Zaragoza, Spain
Université d'Aix-Marseille, France

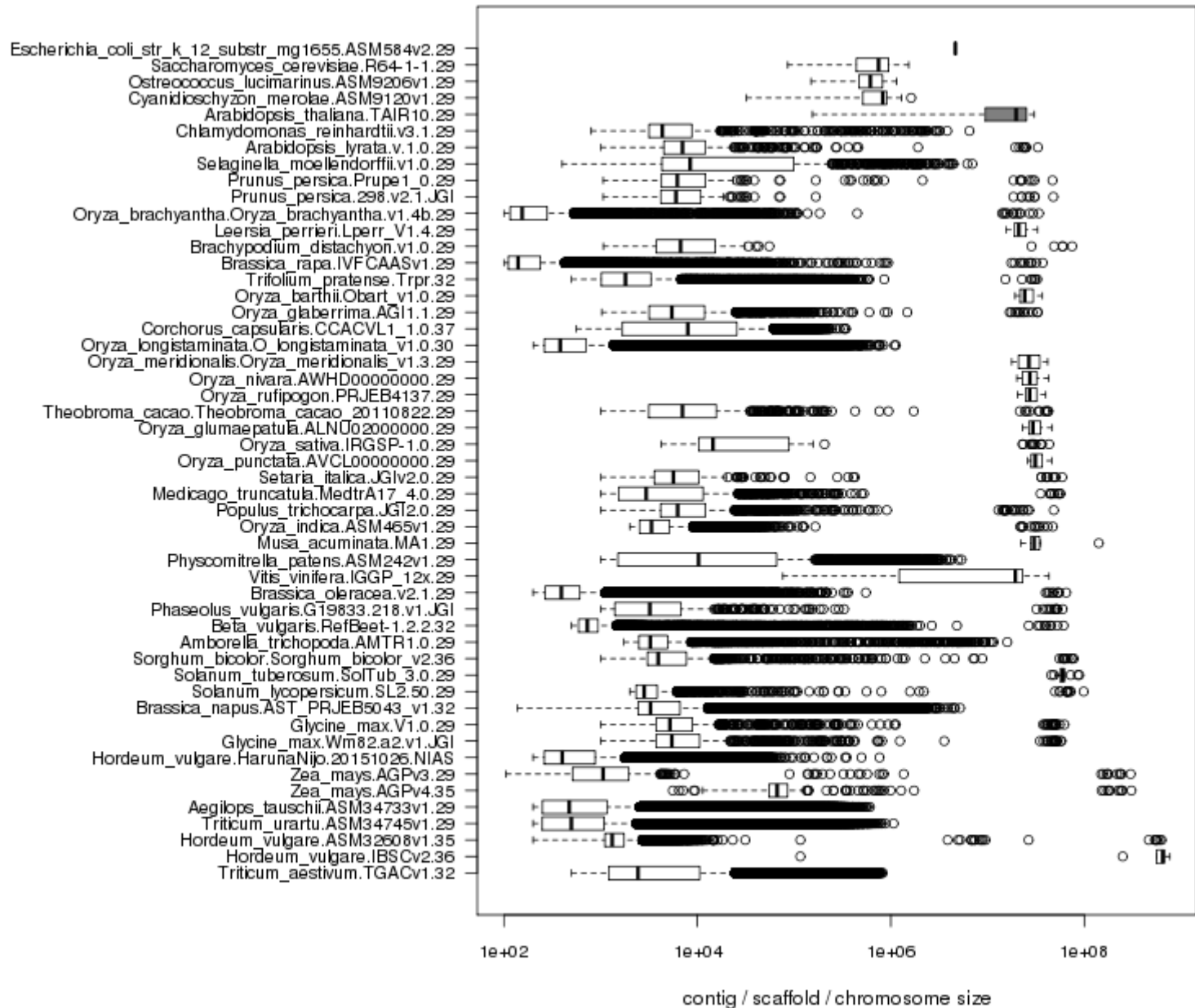
RSAT::Plants data sources



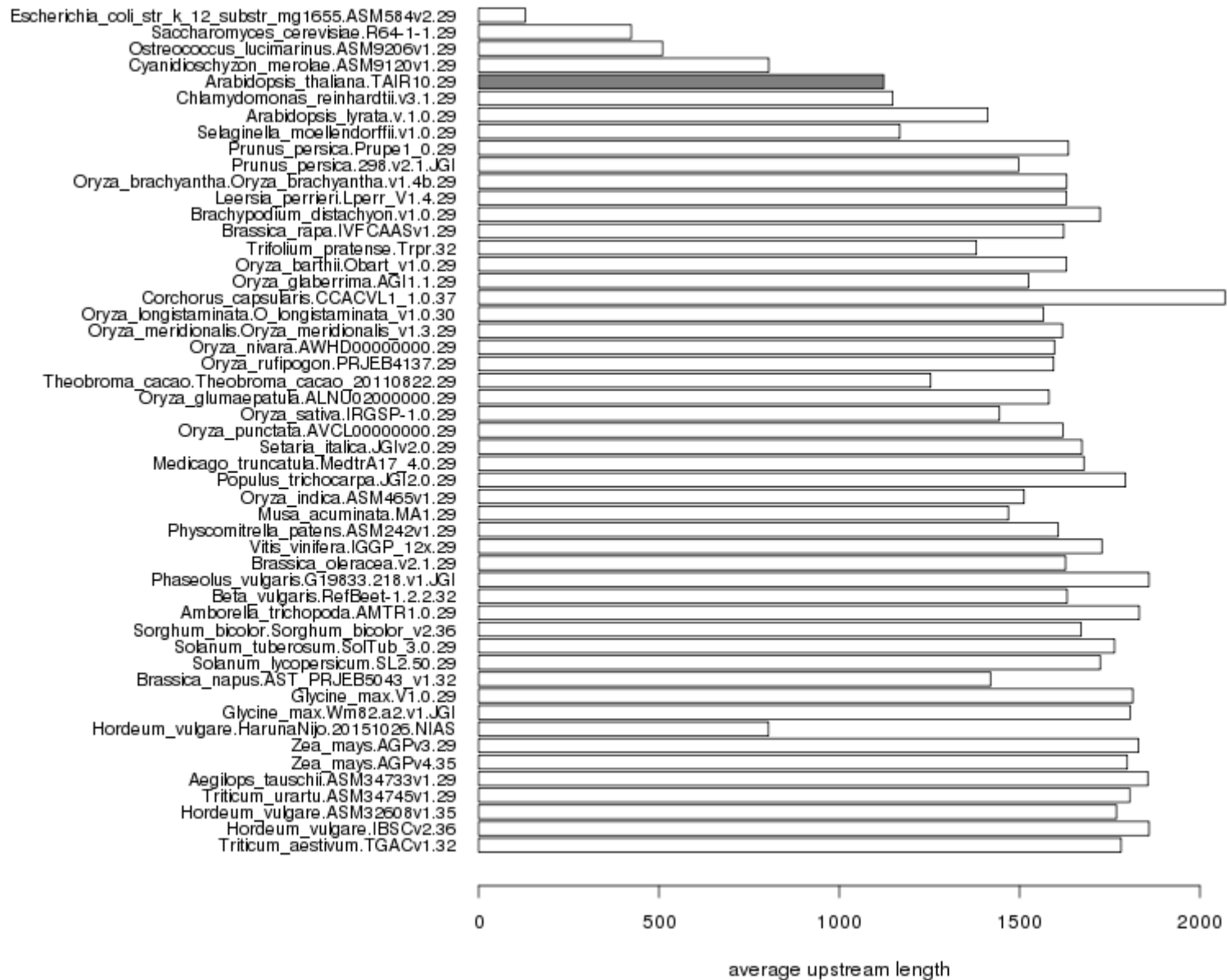
RSAT::Plants: assembled genome size



RSAT::Plants stats: scaffold/contig size



RSAT::Plants stats: upstream regions <1-2kb>



Menu

- Home
- Databases
- **Search**
 - Keywords
 - Sequences
- Credits

Sign In

User:

Password:

(Recover Account Info)

Help

- Documentation

Links

- Laboratory of Computational Biology
- TFcompare
- 3Dfootprint
- #lperl/bioinfo Blog

Welcome to footprintDB

Current version of **footprintDB** includes:

- **3796 Transcription Factors** (TFs, 3016 unique)
- **4413 Position Specific Scoring Matrices** (PSSMs, PWMs or DBMs, 4381 unique)
- **21855 DNA Binding Sites** (DBSs, 18683 unique)

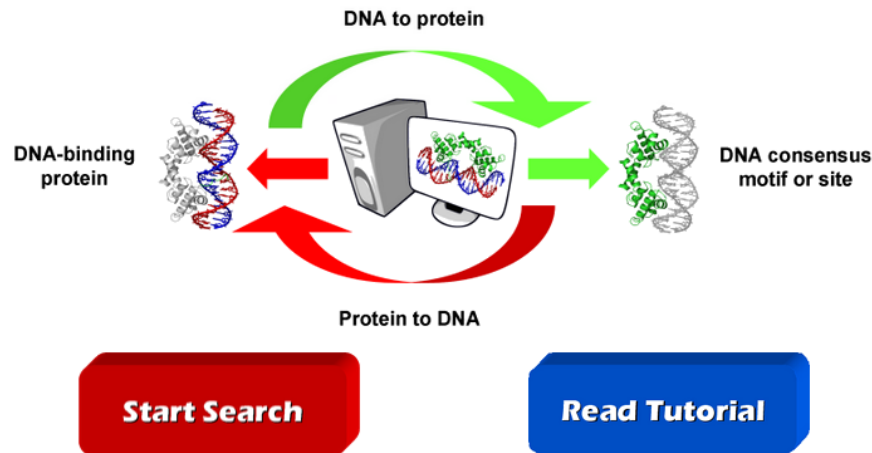
extracted from the [literature](#) and other repositories.

The binding interfaces of (most) proteins in the database are inferred from the collection of protein-DNA complexes described in [3D-footprint](#).

footprintDB predicts:

1. **Transcription factors** which bind a specific DNA site or motif
2. **DNA motifs or sites** likely to be recognized by a specific DNA-binding protein

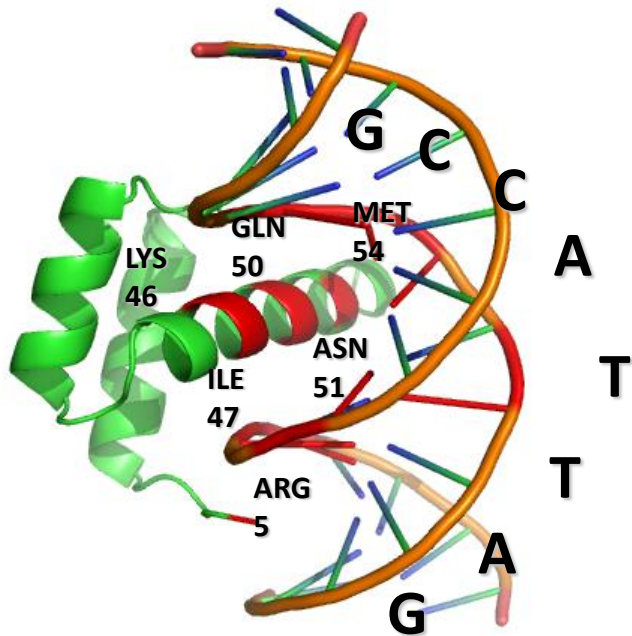
As summarized in the schema:



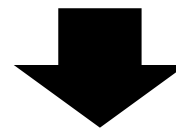
Disclaimer

These data are available AS IS and at your own risk. The EEA/CSIC do not give any representation or warranty nor assume any liability or responsibility for the data nor the results posted (whether as to their accuracy, completeness, quality or otherwise). Access to these data is available free of charge for ordinary use in the course of research.

Protein-DNA complexes in the PDB help annotate TFs



	G	A	T	T	A	C	C	G
R	■							
K								■
I				■				
Q					■			
N			■					
M						■		



...**R**qtytryq...lslterqi**KI**wf**QN**rr**M**...

Sample footprintDB entry

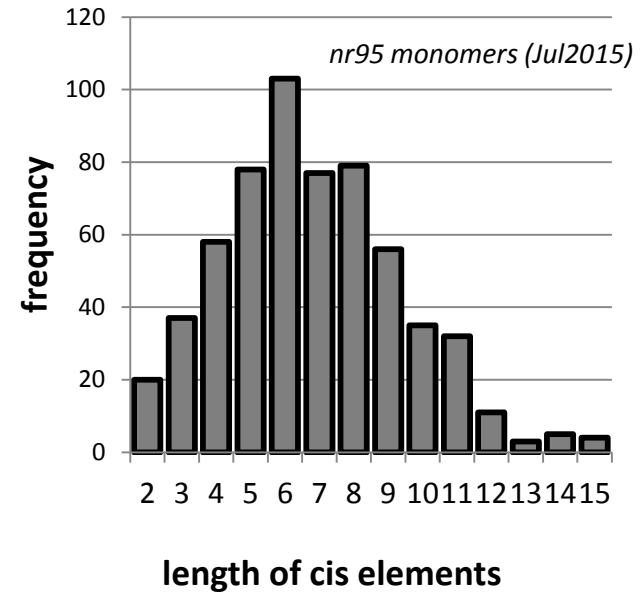
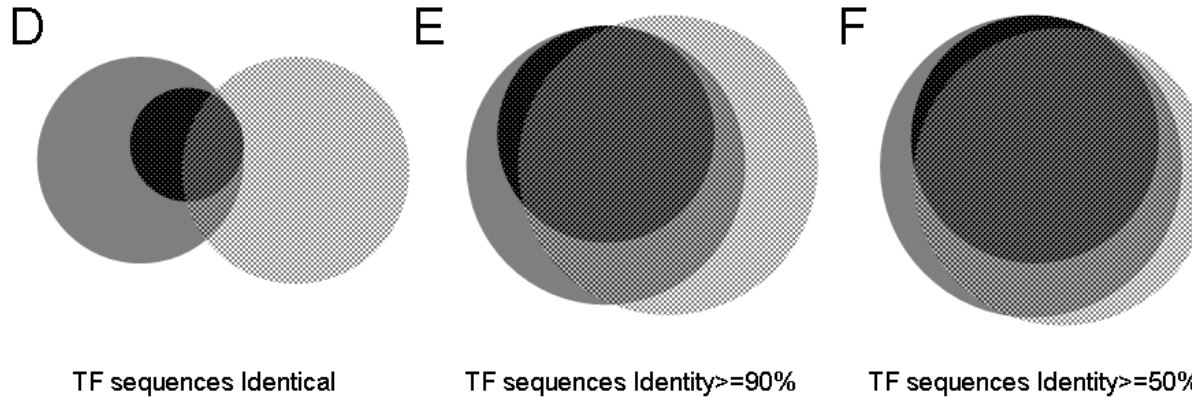
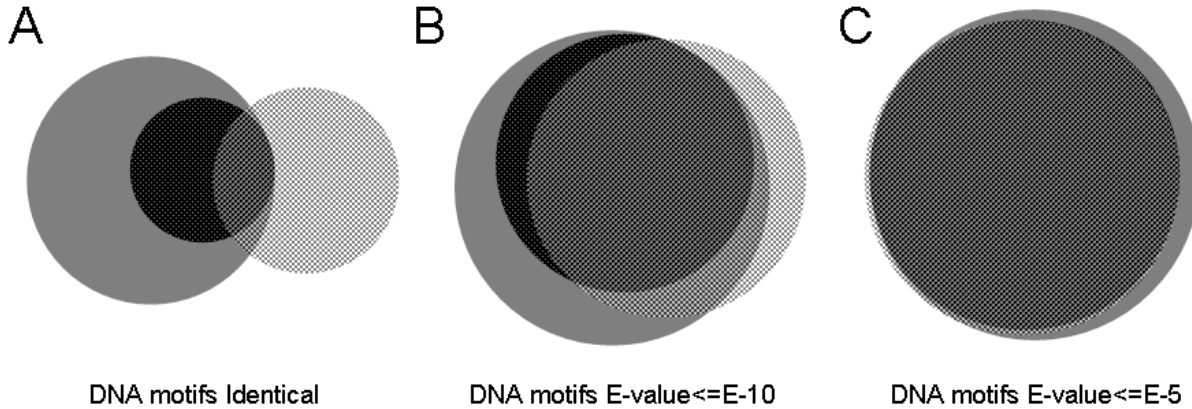


Accessions:	VRN1 (EEADannot 20150505)
Names:	BM5A, HvBM5A, Vernalization1, VRN-H1, VRN1
Organisms:	Hordeum vulgare cv Strider
Libraries:	EEADannot 20150505 ¹
Notes:	family:MADS-box
Length:	244
Pfam Domains:	9-59 SRF-type transcription factor (DNA-binding and dimerisation domain) 78-173 K-box region
Sequence:	<pre> 1 MGRGKVQLKRIENKINRQVTFSKRRSGLLKKKAHEISVLCDAEVGLIIFSTKGKLYEFSTE 60 61 SCMDKILERYERYSYAEKVLVSSSESEIQGNWCHEYRKLKAKVETIQKCQKHLMGEDLESL 120 121 NLKELQQLEQQLESSLKHIRARKNQLMHESISSELQKKERSLQEENKVLQKELVEKQKAQA 180 181 AQQDQTQPQTSSSSSSSFMMRDAPPVADTSNHPAAAGERAEDVAVQPQVPLRTALPLWMVS 240 241 HING</pre>
Interface Residues:	3, 4, 15, 18, 19, 23, 159, 163
3D-footprint homologues:	1n6j_A , 1egw_B , 1hbx_D , 1mnm_A , 3a5t_B

Database	Version	TFs	Motifs	Sites	Description	Authors	Data
JASPAR ¹	2018	1576	2404	12534	Curated, non-redundant set of profiles, derived from published collections of experimentally defined transcription factor binding sites for eukaryotes	Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, Zhang AW, Parcy F, Lenhard B, Sandelin A, Wasserman WW.	Transcription factors Download DNA Binding Motifs DNA Binding Sites
CISBP ²	1.02	1878	1892	0	Subset of directly determined DNA motifs within the CIS-BP Database.	Weirauch MT, Yang A, Albu M et al.	Transcription factors Download DNA Binding Motifs
3D-footprint ³	20161221	1423	1177	1545	Database of DNA-binding protein structures that is updated weekly with Protein Data Bank complexes	Contreras-Moreira,B.	Transcription factors Download DNA Binding Motifs Download DNA Binding Sites Download
HT-SELEX2 ⁴	May2017	533	1761	0	SELEX motifs of human TFs bound to methylated and unmethylated ligands	Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, Nitta KR, Taipale M, Popov A, Ginno PA, Domcke S, Yan J, Schubeler D, Vinson C and Taipale J	Transcription factors Download DNA Binding Motifs
UniPROBE ⁵	20160601	697	1042	4503	Repository of experimental data from universal protein binding microarray (PBM) experiments	Hume MA, Barrera LA, Gisselbrecht SS, Bulyk ML	Transcription factors Download DNA Binding Motifs DNA Binding Sites
HOCOMOCO ⁶	10	535	1066	0	Homo sapiens comprehensive model collection containing carefully hand-curated TFBS models constructed by integration of binding sequences obtained by both low and high-throughput methods	Kulakovskiy I.V., Vorontsov I.E., Yevshin I.S., Soboleva A.V., Kasianov A.S., Ashoor H., Ba-Alawi W., Bajic V.B., Medvedeva Y.A., Kolpakov F.A., Makeev V.J.	Transcription factors Download DNA Binding Motifs
AthalianaCistrome ⁷	v4_May2016	532	838	0	Raw and unmethylated genomic DNA DAP-seq motifs for Arabidopsis thaliana in-vitro-expressed TFs.	OMalley RC, Huang SC, Song L et al.	Transcription factors Download DNA Binding Motifs

footprintDB benchmark

● footprintDB ● TRANSFAC ● JASPAR CORE



Contreras-Moreira B & Sebastian A (2016) *Methods in Molecular Biology*, 1482:259-77 [URL](#)

Sebastian A & Contreras-Moreira B (2014) *Bioinformatics* 30(2):258-65 [URL](#)

1) interrogating footprintDB: web interface

1.1) DNA Motif Search (register to save results)

1.2) Proteome-Specific DNA Search

Query: **bZIP910_(JASPAR_CORE)** m**TGACGT**

footprintDB template	Source	Organisms	STAMP e-value	Motif similarity	footprinDB PWM / Consensus	Binding proteins	Interface sequences	Pfam domains
MA0096.1: bZIP910	JASPAR 2016	Antirrhinum majus	1.0e-12	7.00 / 7	ACGTC Ak ACGTC Ak	Show proteins	Show interfaces	Show domains
MA1068.1: TGA2	JASPAR 2016	Arabidopsis thaliana	1.7e-11	6.97 / 7	ACGTC Ak - ACGTC AkC	Show proteins	Show interfaces	Show domains
M0270_1.02: OBF5 / T023874_1.02	CISBP 1.02	Arabidopsis thaliana	1.5e-10	6.76 / 7	- ACGTC Ak k ACGTC Ay	Show proteins	Show interfaces	Show domains
MA1047.1: TGA5	JASPAR 2016	Arabidopsis thaliana	1.5e-10	6.76 / 7	- ACGTC Ak k ACGTC Ay	Show proteins	Show interfaces	Show domains
TGA2_2: TGA2	ArabidopsisPBM 20140210	Arabidopsis thaliana	3.0e-10	6.97 / 7	- ACGTC Ak - - k ACGTC AkCa	Show proteins	Show interfaces	Show domains

1) interrogating footprintDB: web interface

1.3) Protein Sequence Search

Query: bZIP910 (JASPAR CORE)

Sequence: MASQQRSTSPGIDDDERKRKRKLSNRESARRSRMRKQQRLELIAQESQMQEDNKKLRDTINGATQLYLNFA SDNNVLR AQLAELTDR LHSLNSVLQIASEVSGLVLDIPDIPDALLEPWQLPCPIQADIFQC

footprintDB template	Source	Organisms	Blast e-value	Interface similarity	Interface residues	Interface sequence	footprinDB PWM / Consensus	Pfam domains
bZIP910 / T028334_1.02 / O22676: bZIP910 / O22676_ANTMA / T028334_1.02;	Athamap 20091028 CISBP 1.02 JASPAR 2018	Antirrhinum majus	6e-94	6 / 6	21, 25, 28, 29, 32, 33	... (16) erkrkRklsNreSarrSRmrkq Q(38) ...	bZIP910(1): GATGACGTGGCm bZIP910(2): GgrTGCTGACGT MA0096.1: mTGACGT M2679_1.02: GATGACGTGGCm	*13-61: PF07716 (Basic region leucine zipper)
bZIP911 / T028335_1.02 / O22677: bZIP911 / O22677_ANTMA / T028335_1.02;	Athamap 20091028 CISBP 1.02 JASPAR 2018	Antirrhinum majus	7e-64	9 / 9	23, 26, 27, 28, 30, 31, 32, 34, 35	... (18) erkrkRkqSNreSARRSRmrkq Q(40) ...	bZIP911(1): GrTGACGTGGCC bZIP911(2): GrTGACGTGTAC MA0097.1: GrTGACGTGkmC M2681_1.02: GrTGACGTGGCC	*17-67: PF07716 (Basic region leucine zipper)
T11942 / Q9LZP8: AT3G62420 / bZIP53 / T11942; / AtbZIP53 / bZIP transcription factor 53 / BZP53_ARATH	AthalianaCistrome v4_May2016 JASPAR 2018	Arabidopsis thaliana Arabidopsis thaliana Col-0	1e-46	9 / 9	25, 29, 32, 33, 34, 36, 37, 40, 41	... (20) tvtdeRkrkRmiSNreSarrSR mrkqk(46) ...	M0196: GTGGCa M0188: GTGKCa : dwwGmTGACGTGGCa MA1341.1: dwwGmTGACGTGGCa	*22-72: PF07716 (Basic region leucine zipper)

2) interrogating footprintDB: SOAP web service

Source: <http://floresta.eead.csic.es/footprintdb/index.php?documentation>
<https://bioinfoperl.blogspot.com.es/2017/10/soap-interface-of-footprintdb.html>

3) interrogating footprintDB: aligning to FASTA peptides, BLASTP

3.1) all nr TFs

3.2) metazoa

3.3) plants

```
>1:ACE2 [Saccharomyces cerevisiae] libs:JASPAR;CISBP;  
motif:vTGCTGGtym;mCCAGCa;  
http://floresta.eead.csic.es/footprintdb/index.php?motif=cb6f6b343b89  
5dfa1c3776c99fbedda7  
MDNVVDPWYINPSGFAKDTQDEEYVQHHDNVNPTIPPPDNYILNNENDDGLDNLLGMDYYN  
IDDLLTQELRDLDIPLVPSPKTGDGSSDKKNIDRTWNLGDENNKVSHYSKKSMSHHRGLS  
GTAI FGFLGHNKTLSSISLQQSILNMSKDPQPMELINE LGNHNTVKNNNDDFDHIRENDGE  
NSYLS...
```

Sources: <http://floresta.eead.csic.es/footprintdb/download>

<https://bioinfoperl.blogspot.com.es/2017/04/genome-annotation-with-footprintdb.html>

4) ways of interrogating footprintDB: motif collection in RSAT



RSAT NeAT

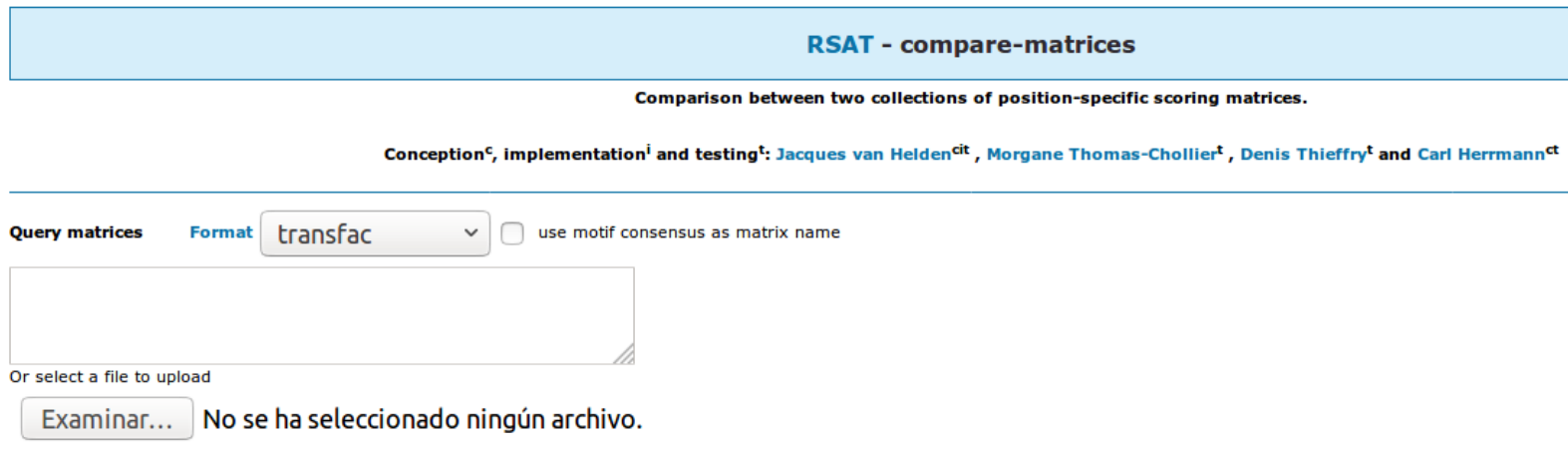
RSAT Plants
51 organisms

Search

New items

> view all tools

- Genomes and genes
- Sequence tools
- Matrix tools
 - convert matrix
 - compare matrices
 - matrix-clustering
 - matrix distrib
 - matrix quality
- Build control sets
- Motif discovery
- Pattern matching
- Comparative genomics
- NGS - ChIP-seq
- Genetic variations
- Conversion/Utilities
- Drawing
- SOAP Web services



RSAT - compare-matrices

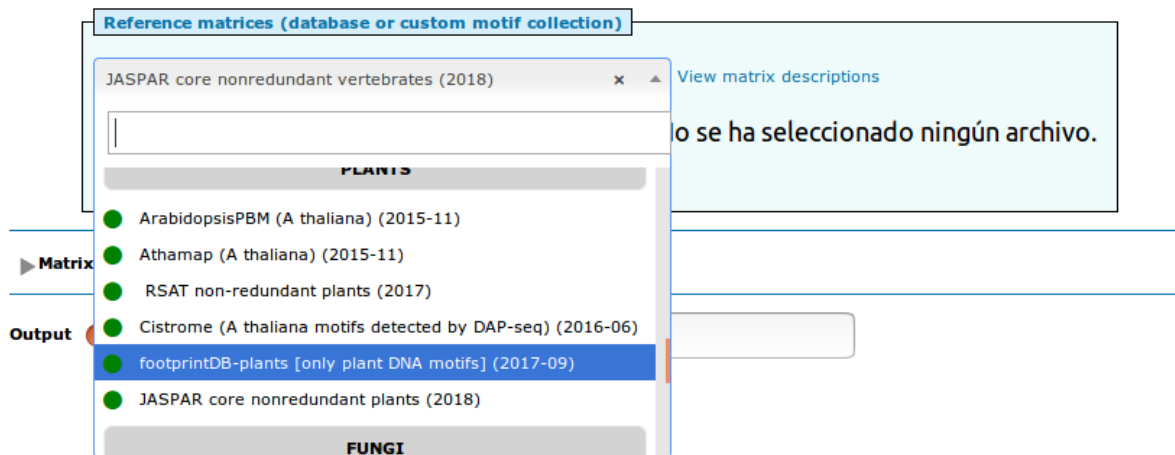
Comparison between two collections of position-specific scoring matrices.

Conception^c, implementationⁱ and testing^t: Jacques van Helden^{dt}, Morgane Thomas-Chollier^t, Denis Thieffry^t and Carl Herrmann^{ct}

Query matrices Format: use motif consensus as matrix name

Or select a file to upload

No se ha seleccionado ningún archivo.



Reference matrices (database or custom motif collection)

JASPAR core nonredundant vertebrates (2018) View matrix descriptions

No se ha seleccionado ningún archivo.

Matrix

Output

- ArabidopsisPBM (A thaliana) (2015-11)
- Athamap (A thaliana) (2015-11)
- RSAT non-redundant plants (2017)
- Cistrome (A thaliana motifs detected by DAP-seq) (2016-06)
- footprintDB-plants [only plant DNA motifs] (2017-09)
- JASPAR core nonredundant plants (2018)

PLANTS

FUNGI

Contreras-Moreira B, Castro-Mondragon J, Rioualen C, Cantalapiedra CP, van Helden J (2016) *Methods in Molecular Biology*, 1482:279-95 [URL](#)

Castro-Mondragon J, Rioualen C, Contreras-Moreira B, van Helden J (2016) *Methods in Molecular Biology*, 1482:297-322 [URL](#)

Thanks!

<http://eead.csic.es/compbio>