

Tools and training for agrigenomics at ELIXIR Spain

B Contreras-Moreira¹, I Yruela¹, N Ksouri¹, J Sàrria¹, L Ruiz-Torres^{1,2}

¹ Estación Experimental Aula Dei-CSIC, 50059, Zaragoza, Spain

² Universidad de Jaén, 23071, Jaén, Spain

ELIXIR All Hands, 10-12 June 2024, Uppsala, Sweden



Summary: The Computational and Structural Biology Group at EEAD-CSIC is part of the ELIXIR Plant Sciences Community and develops software for the analysis of pangenomes and regulatory sequences of plants. The tools currently included in the Service Delivery Plan of ELIXIR-ES are **GET_HOMOLOGUES** and **RSAT::Plants**, which we benchmark regularly in pre-breeding studies with crops such as barley and peach, and in evolutionary studies with populations of *Sinapis alba* and *Brachypodium*.

GET_HOMOLOGUES

GitHub, bioconda, Docker

Originally developed for the construction of bacterial **gene-based** pangenomes and then updated to handle **plant** transcriptomes and genomes (**figure 1**). Comprehensive resources, including code, documentation and step-by-step tutorials are available at https://github.com/eead-csic-compbio/get_homologues.

This software also helped build [GET_PANGENES](#), a collaborative effort with the EMBL-EBI and the University of Liverpool, which computes genome alignments to call **collinear pangenomes**.



GitHub, Docker

Analysis of regulatory sequences and motif discovery in 100+ plant genomes mostly sourced from **Ensembl Plants** (**figure 2**). Integrates several collections of DNA motifs to help understand and annotate discovered *cis* sequences. These include sets of curated motifs and transcription factors with annotated interface residues at **footprintDB**. Available at <http://plants.rsat.eu>.



[plant pangenome protocol]



[motif discovery tutorial]



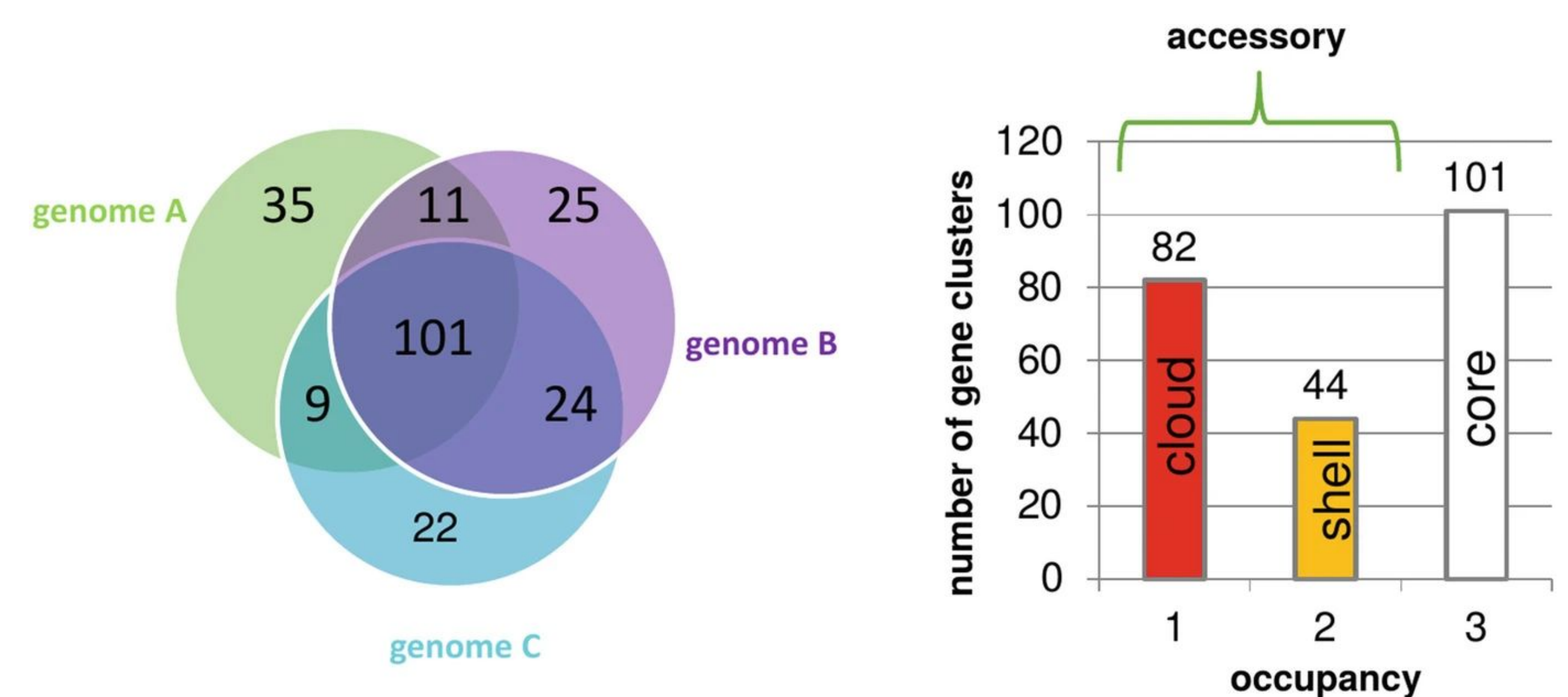
[GEA tutorial]

Teaching and outreach

International Master in Plant Genetics, Genomics and Breeding

Members of the lab, together with colleagues from **Ensembl Outreach**, teach scripting skills and genomic crop analysis every two years (<http://www.masterplantbreeding.com>).

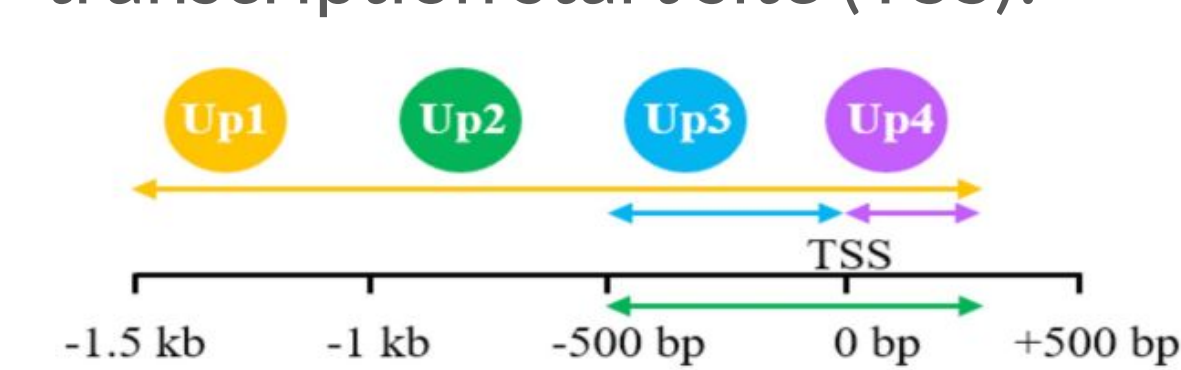
We also host visiting students who are interested in learning how to analyze their own crop genome and environmental data (GEA).



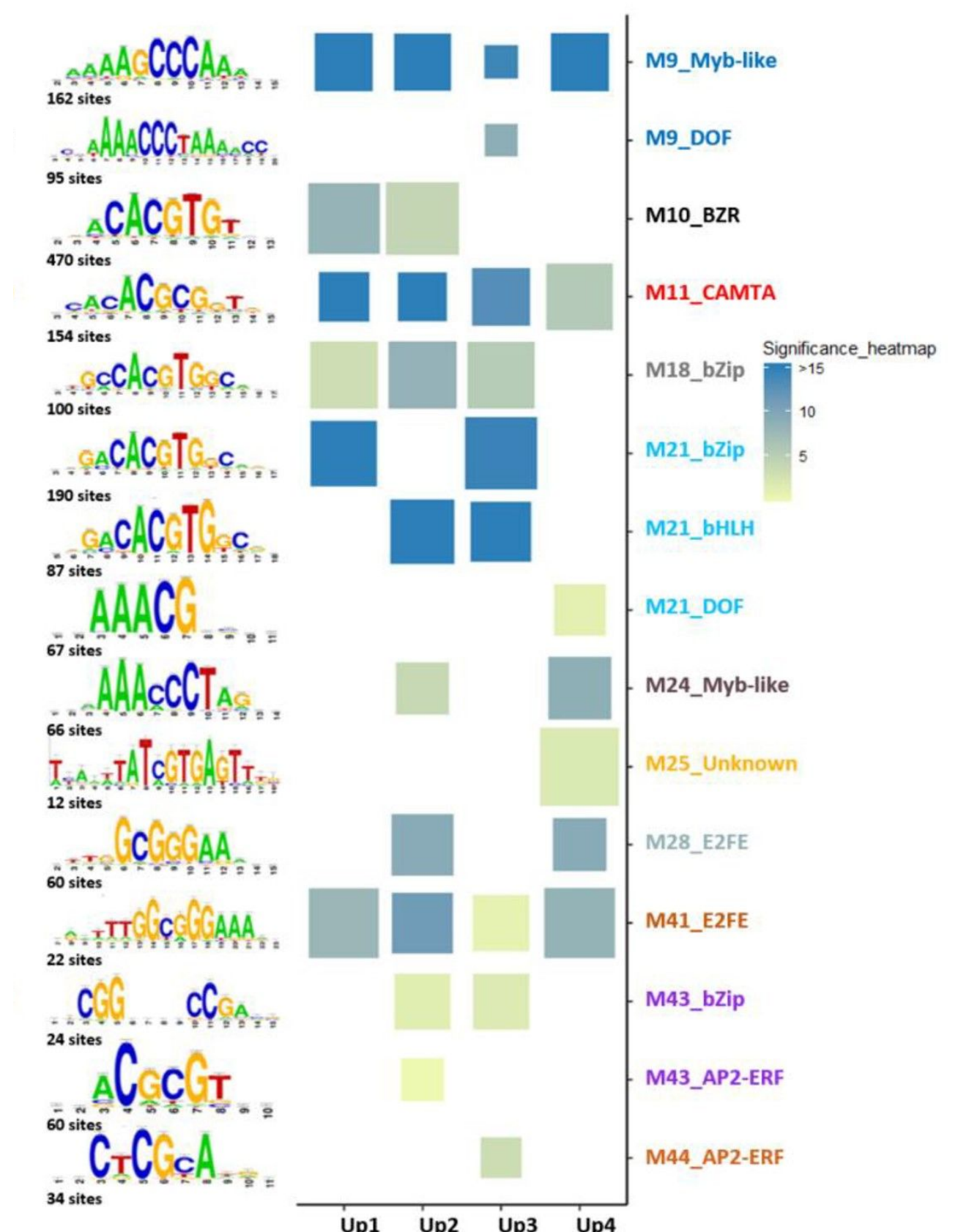
genome A	genome B	genome C	occupancy	class
gene 1A	gene 41B	-	2	shell
gene 2A	gene 42B	genes 114C, 120C	3 (inparalogues)	core
gene 3A	-	-	1	cloud
gene 4A	gene 112B	gene 122C	3 (single-copy)	core

Figure 1. Pangenome set of three genomes (A, B and C). In this example there are 101 core loci and 44 (11 + 9 + 24) shell loci are found in two accessions. Moreover, 82 (35 + 25 + 22) cloud loci are annotated only in one genome. The pangenome set size is 82 + 44 + 101 = 227 loci. The table below shows four example clusters and their occupancy.

Figure 2. Sequence logos of top motifs discovered motifs in modules of co-expressed peach genes along different upstream lengths. The x-axis corresponds to 4 intervals around the transcription start site (TSS):



Cell colors indicate the statistical significance of the identified motifs while cell sizes represent similarity to curated motifs.



Funding

[101091255.WG] by HORIZON-MISS-2021-SOIL-02
[PID2022-142116OB-I00] by MICIU/AEI/10.13039/501100011033



[A08_23R] funded by Aragón government
[FAS2022_052, conexión BCB] funded by CSIC.



Contact

Bruno Contreras Moreira
Estación Experimental de Aula Dei-CSIC
<https://eead.csic.es/compbio>
bcontreras@eead.csic.es



Funded by
the European Union



/company/elixir-europe