Genome-wide analysis of a wheat transcription factor family:

The power of bioinformatics resources

27th May 2020

Susanne Schilling & Rainer Melzer
School of Biology and Environmental Science
University College Dublin

@UCDflowerpower https://ucdflowerpower.org/

What drives domestication?

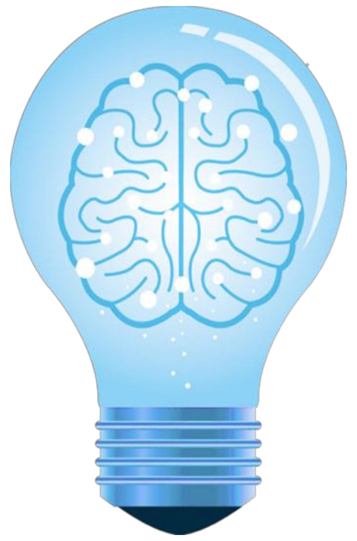
Domestication genes: genes with allelic versions that contributed to cultivating plants for human needs.

3 main groups of genes/proteins:

1. enzymes or structural proteins “superheros“

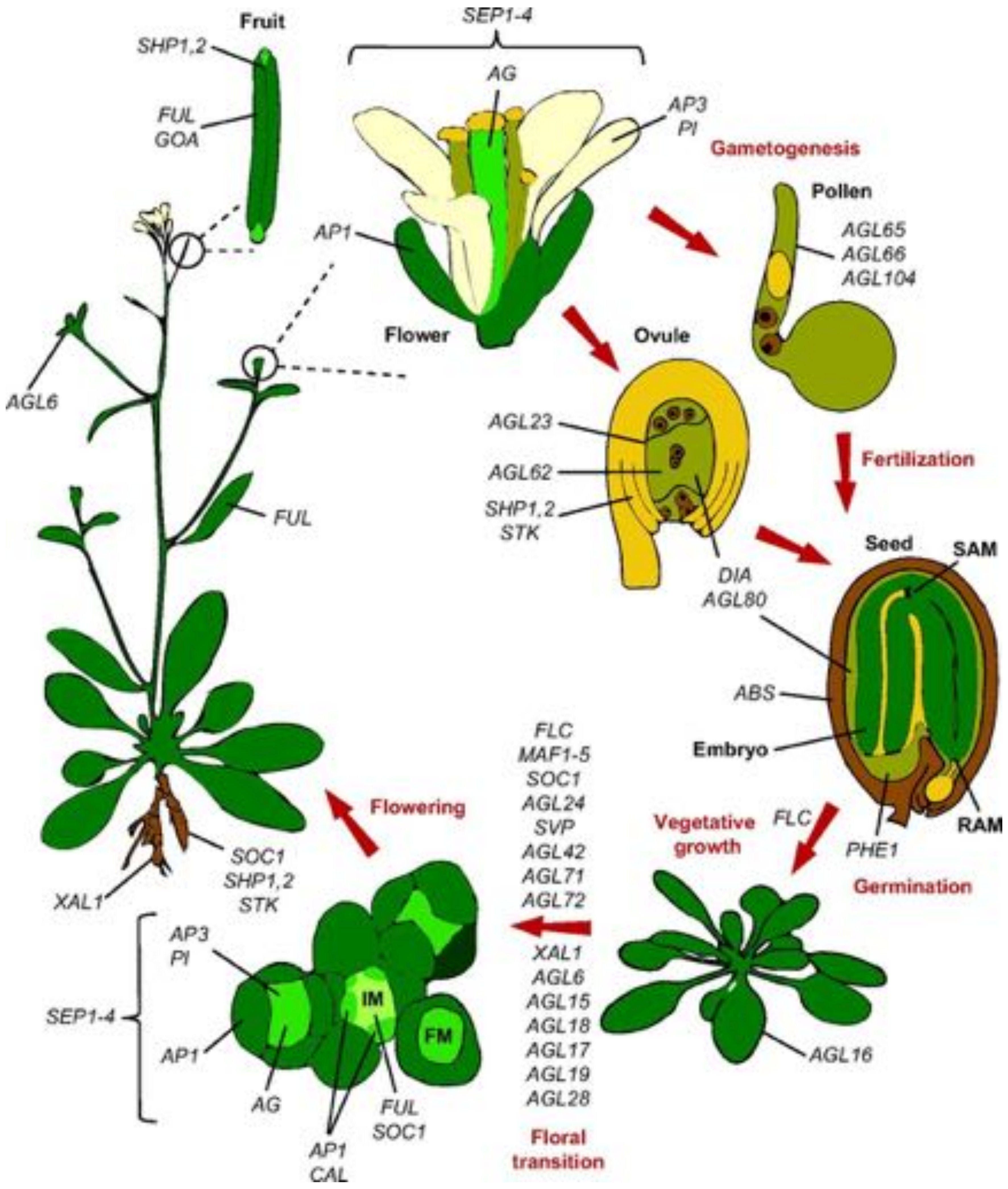
2. numerous genes collectively “minions“

3. transcription factors “masterminds“



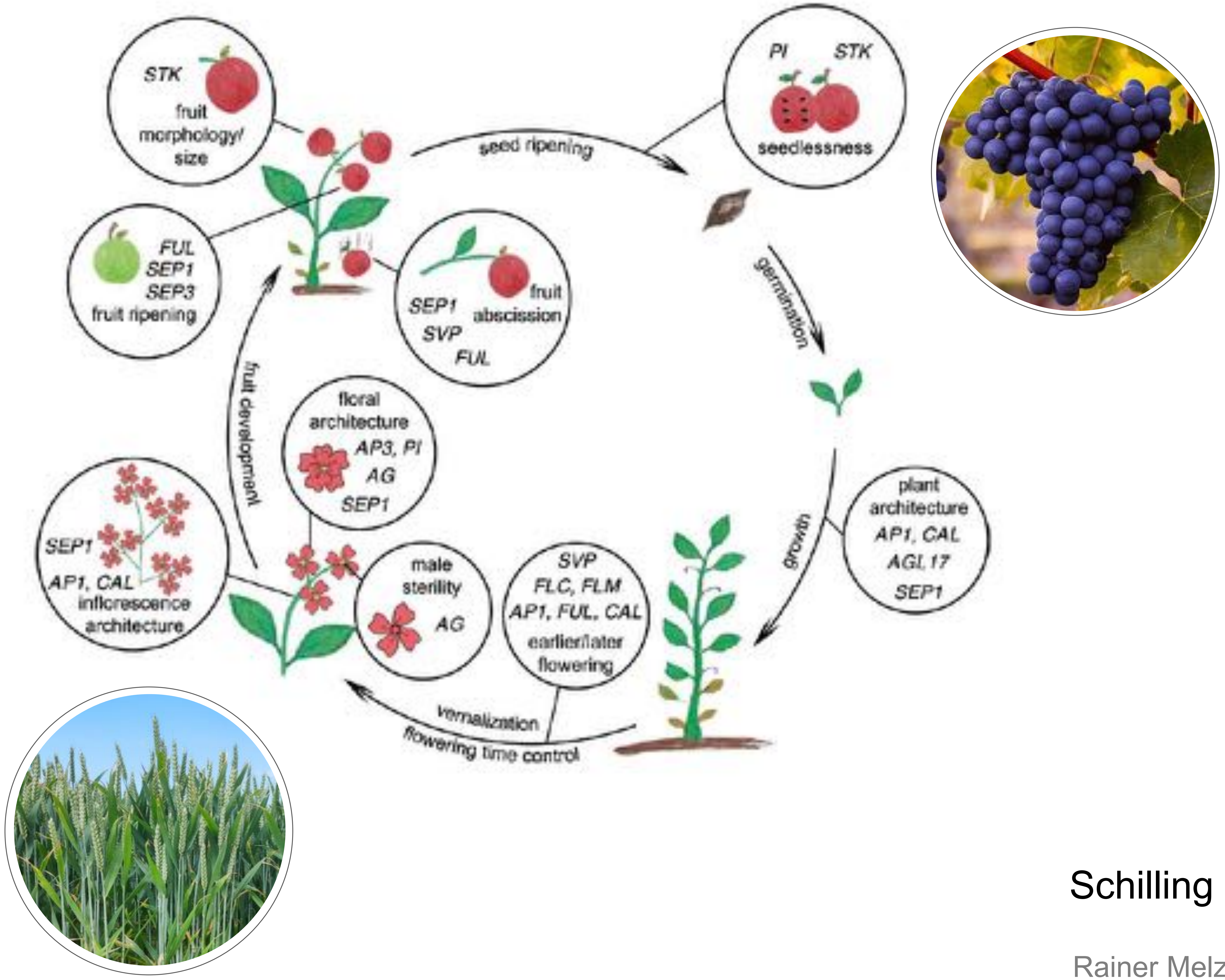
Martínez-Ainsworth & Tenaillon, 2016
Schilling et al., 2018

MADS-box genes are key regulators of plant development



Smaczniak et al., 2012

MADS-box genes and plant domestication



Schilling et al., 2018

Genome-wide characterization of MADS-box genes in wheat

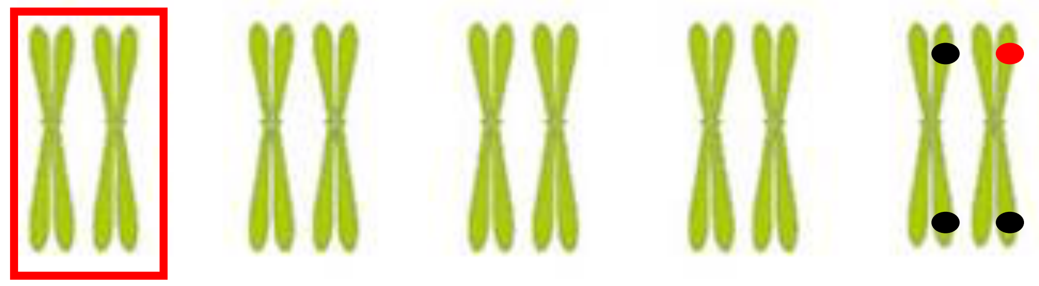
Rainer Melzer & Susanne Schilling, UCD, Ireland

Why wheat is not a model plant

diploid (*Arabidopsis* and rice)

2n

homologs

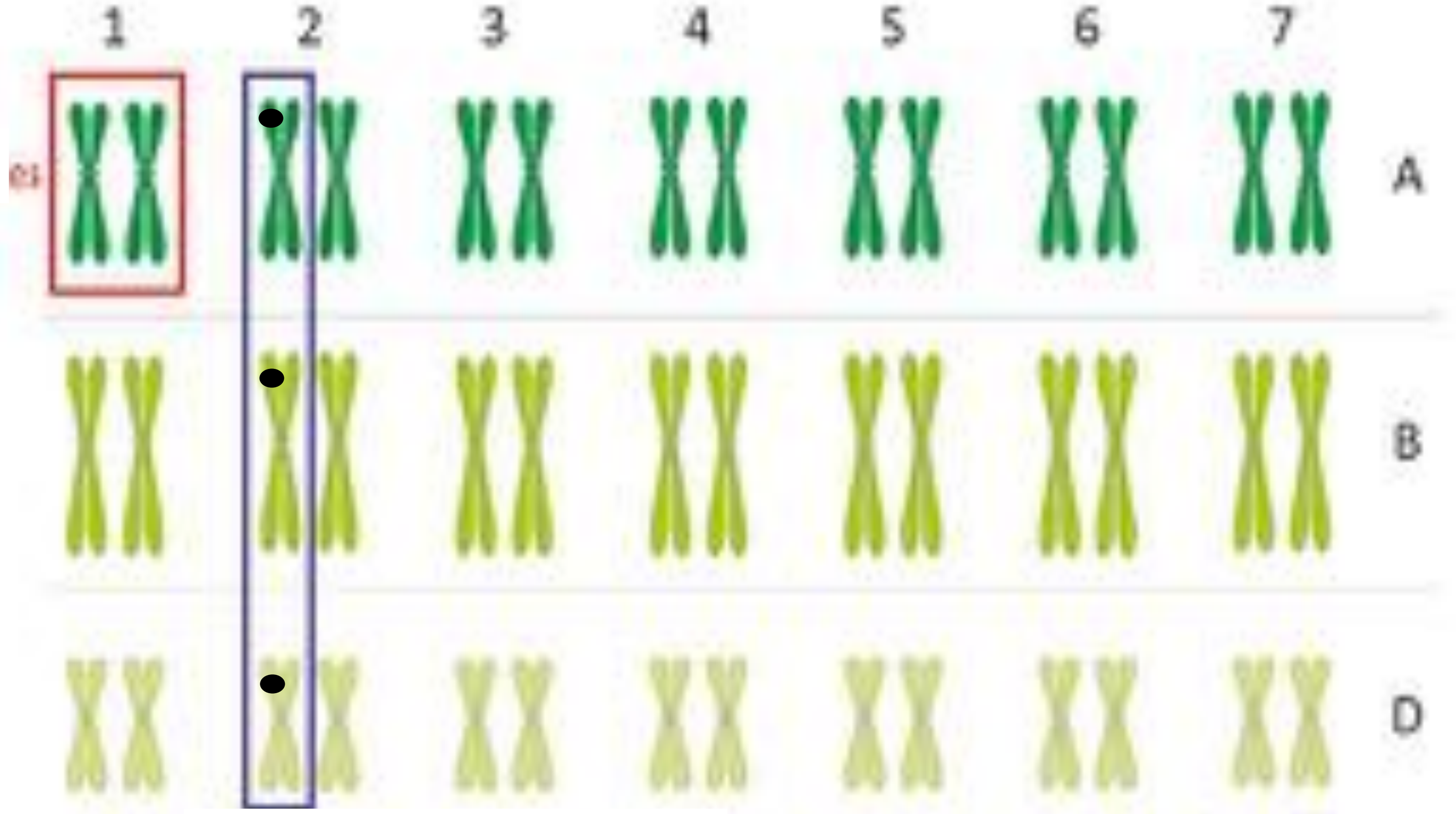


1 gene
different alleles

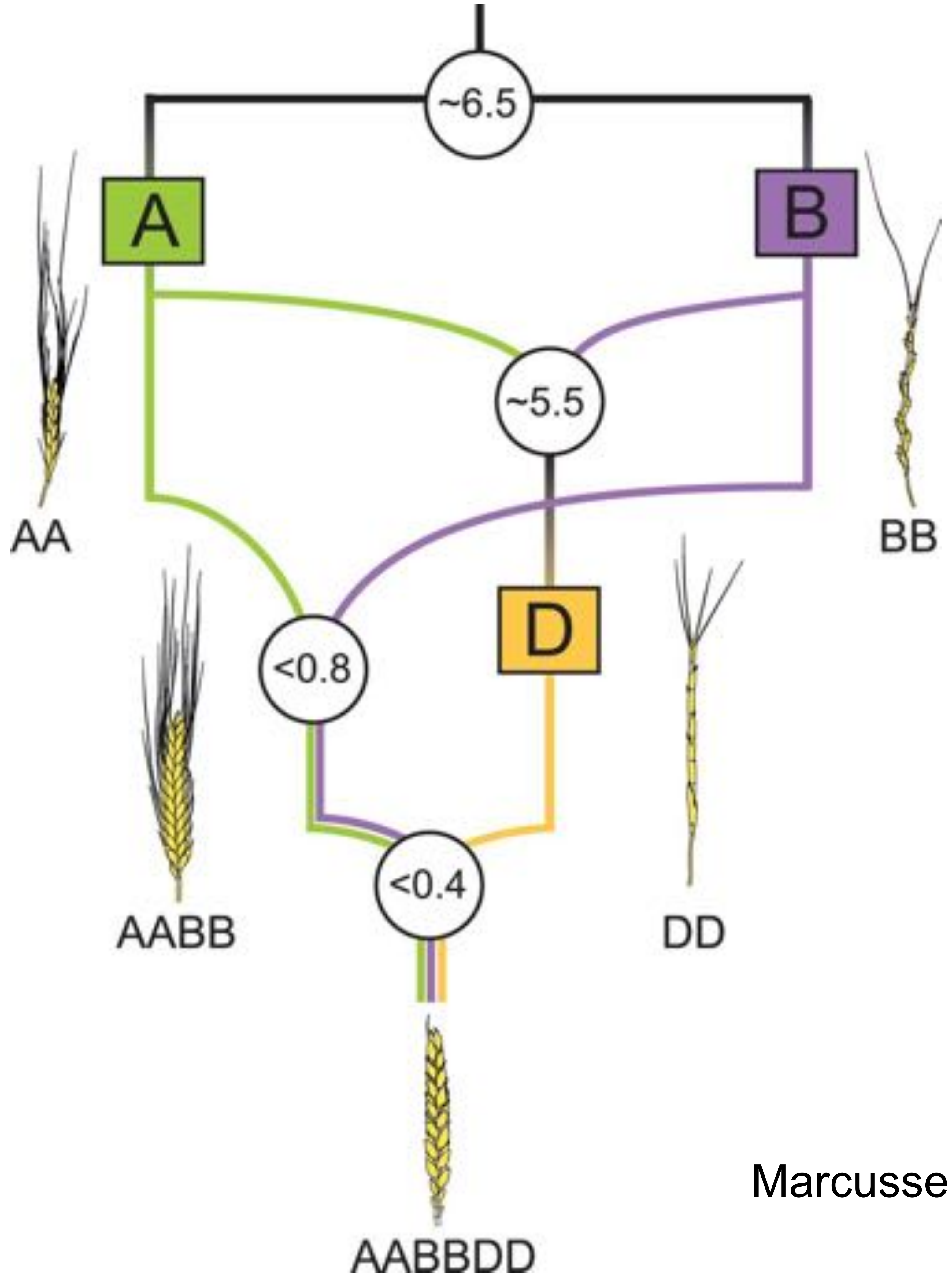
hexaploid (wheat)

homologs

6n



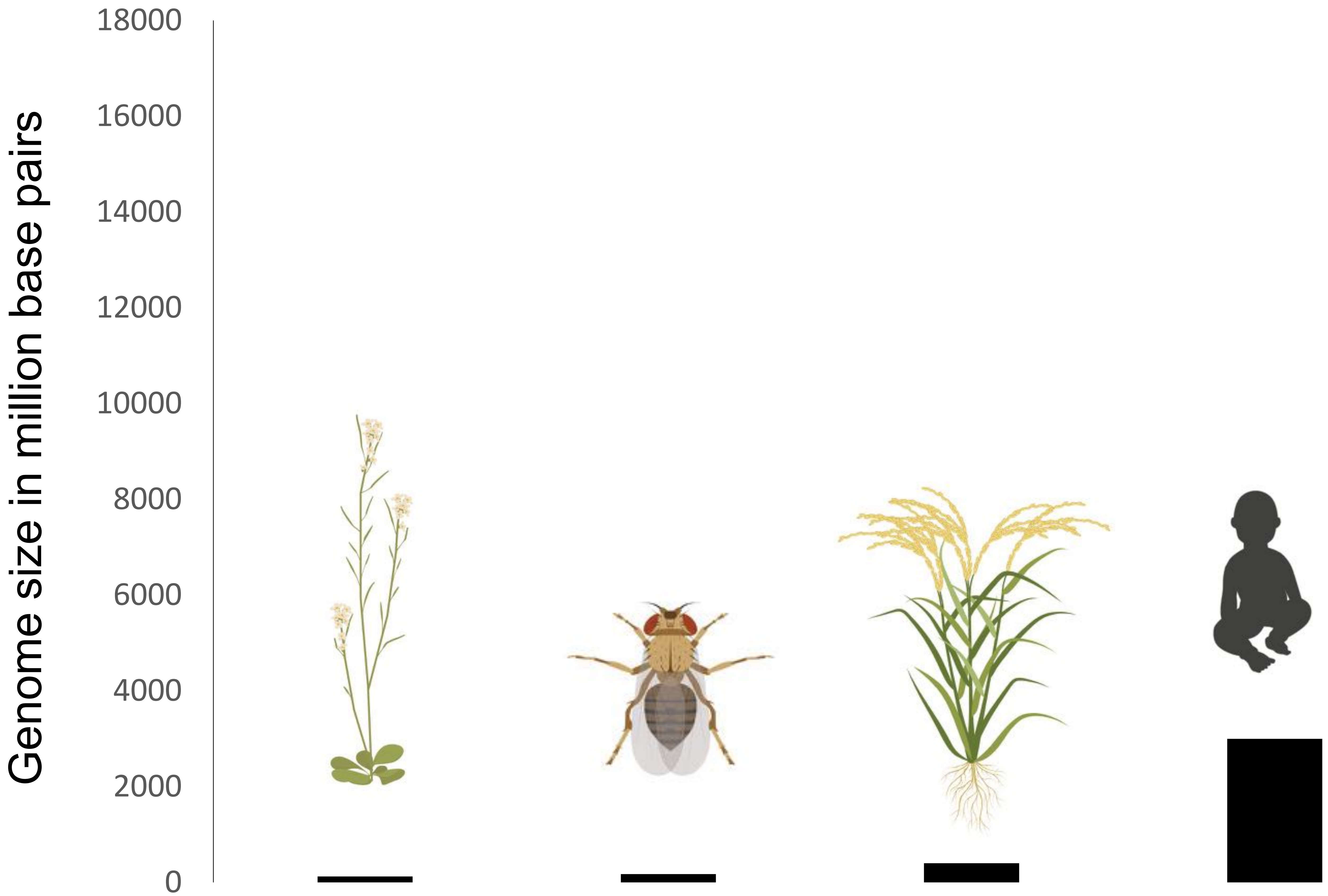
homoeologs



Marcussen et al., 2014

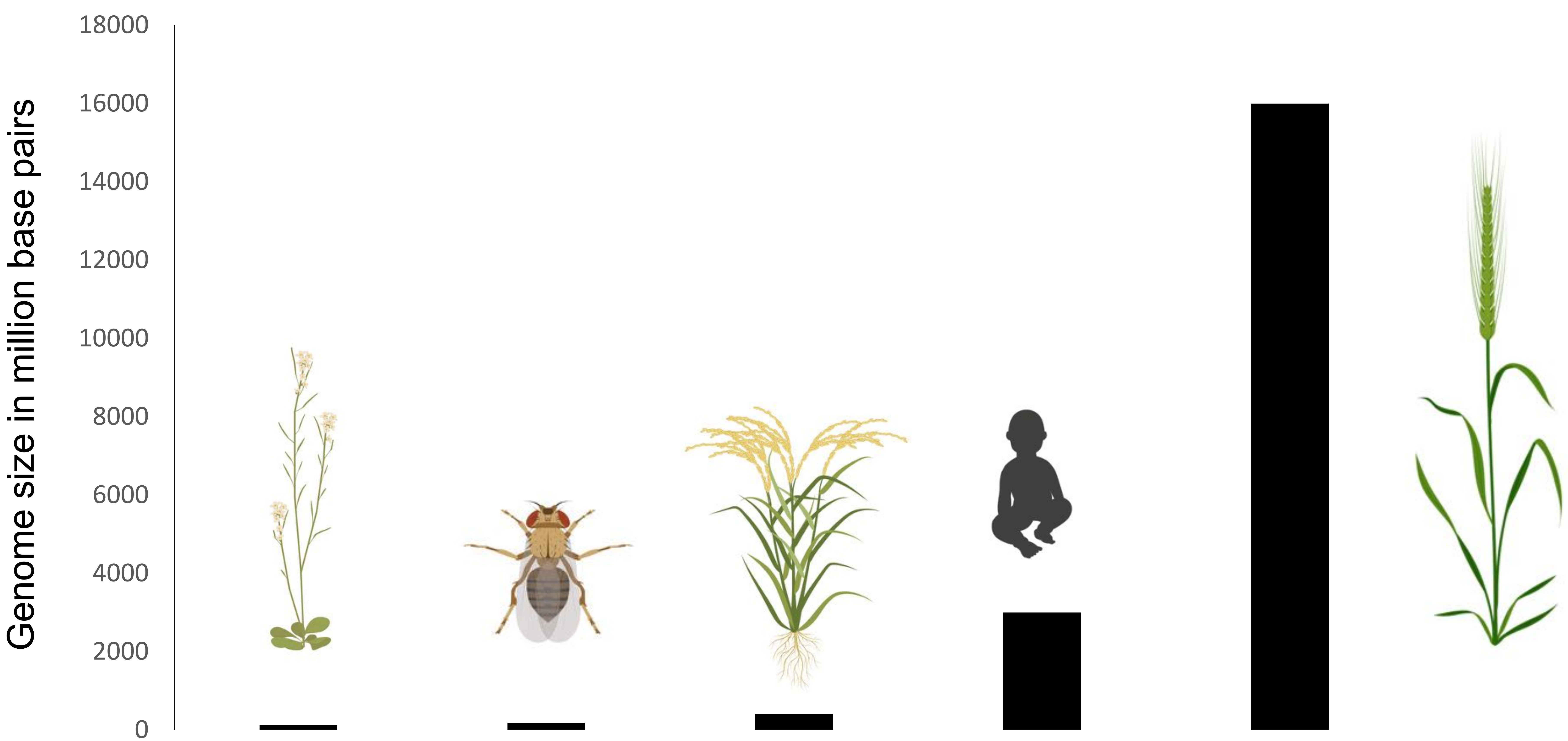
www.jic.ac.uk/

Why wheat is not a model plant



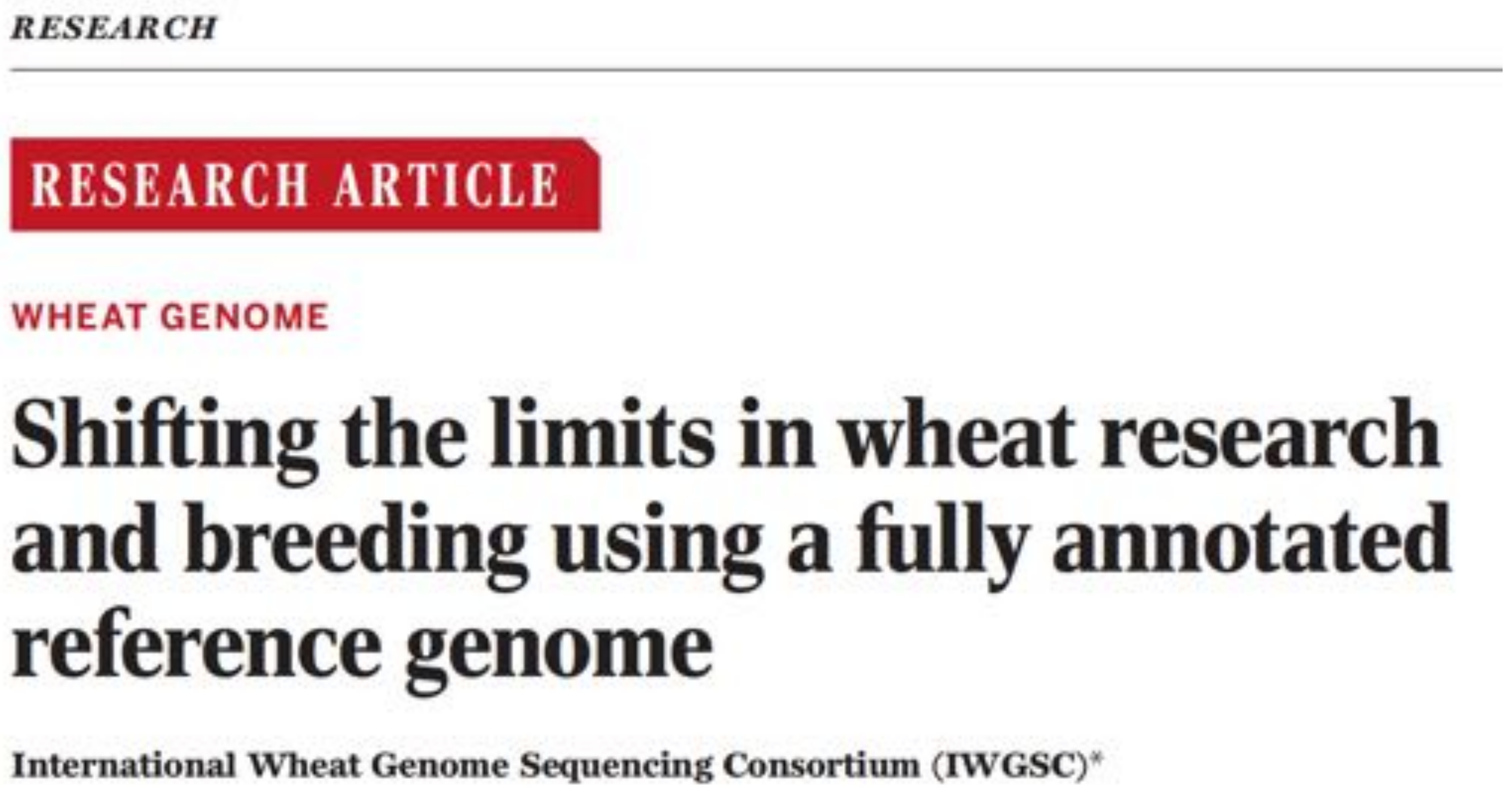
Rainer Melzer & Susanne Schilling, UCD, Ireland

Why wheat is not a model plant



Rainer Melzer & Susanne Schilling, UCD, Ireland

With the help of a high quality wheat genome



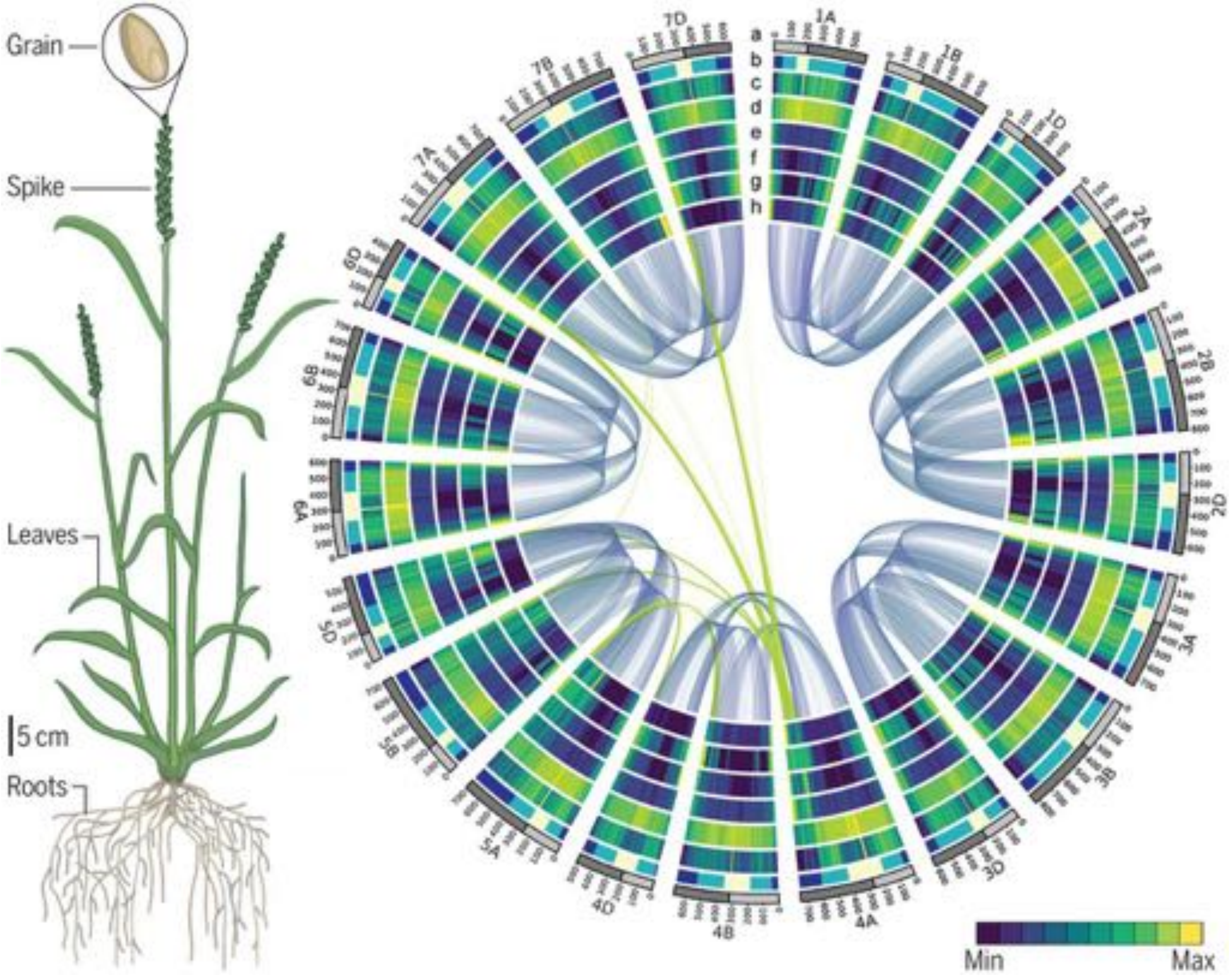
RESEARCH

RESEARCH ARTICLE

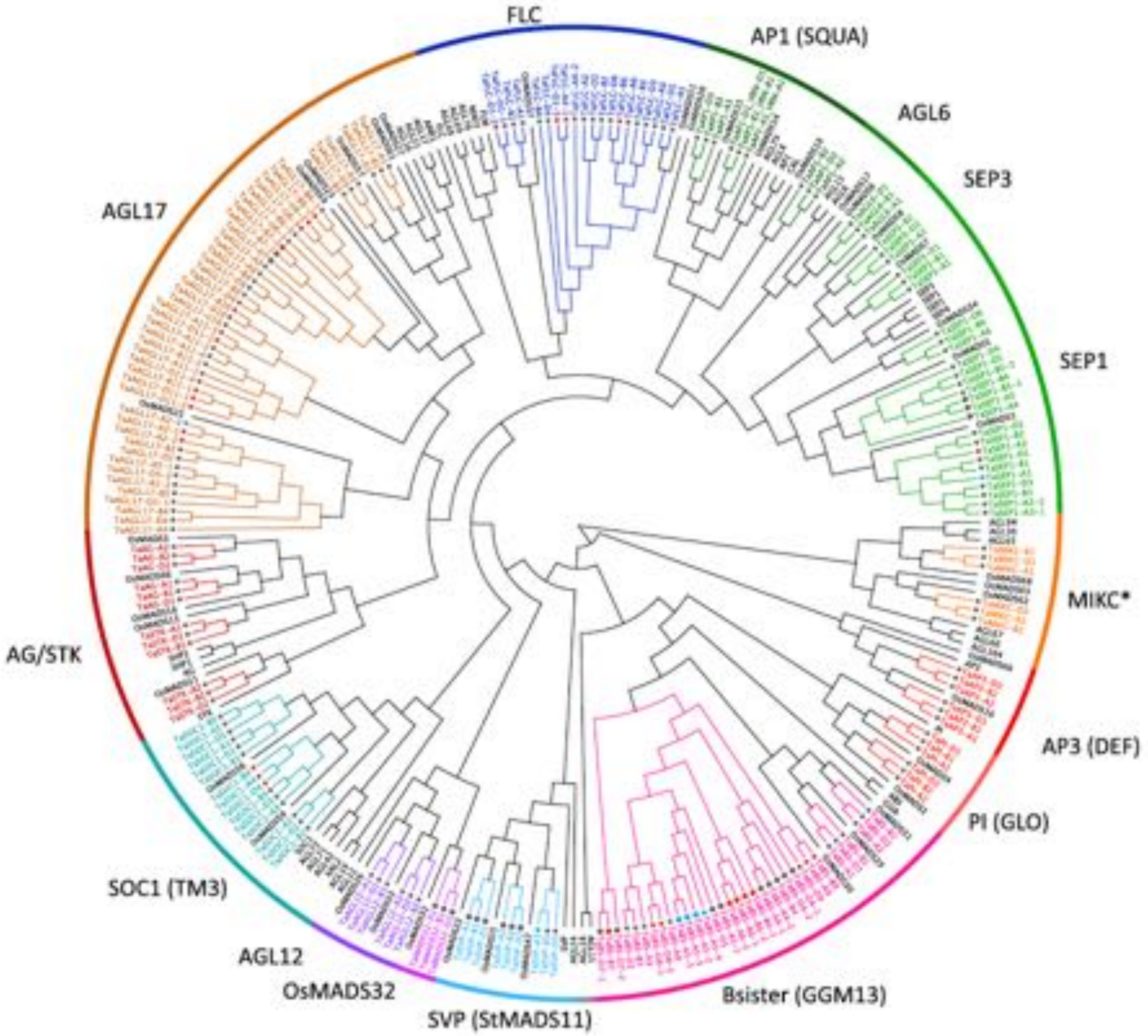
WHEAT GENOME

Shifting the limits in wheat research and breeding using a fully annotated reference genome

International Wheat Genome Sequencing Consortium (IWGSC)*

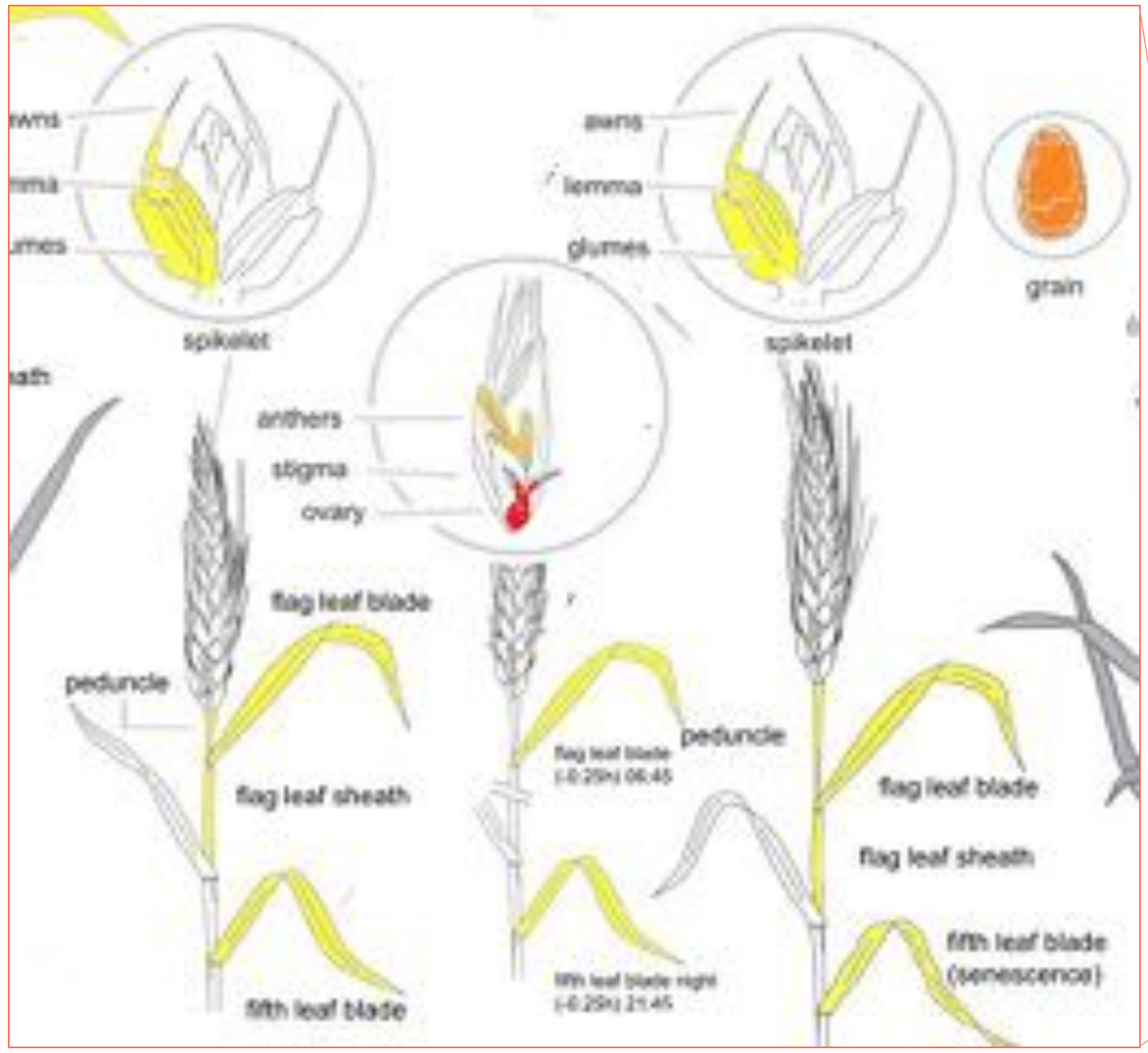


MADS-box genes are highly conserved in wheat

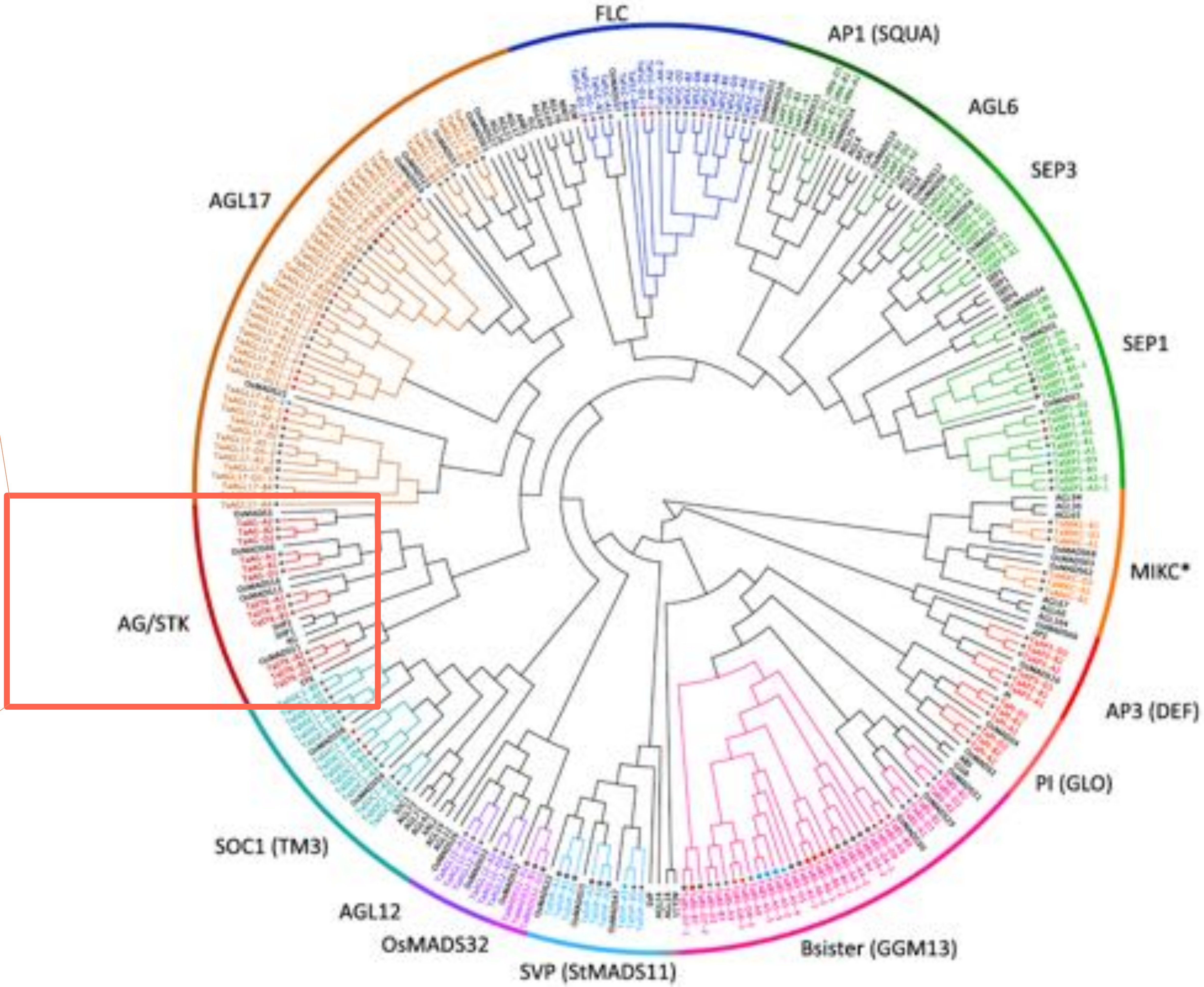


Schilling et al., 2020

MADS-box genes are highly conserved in wheat



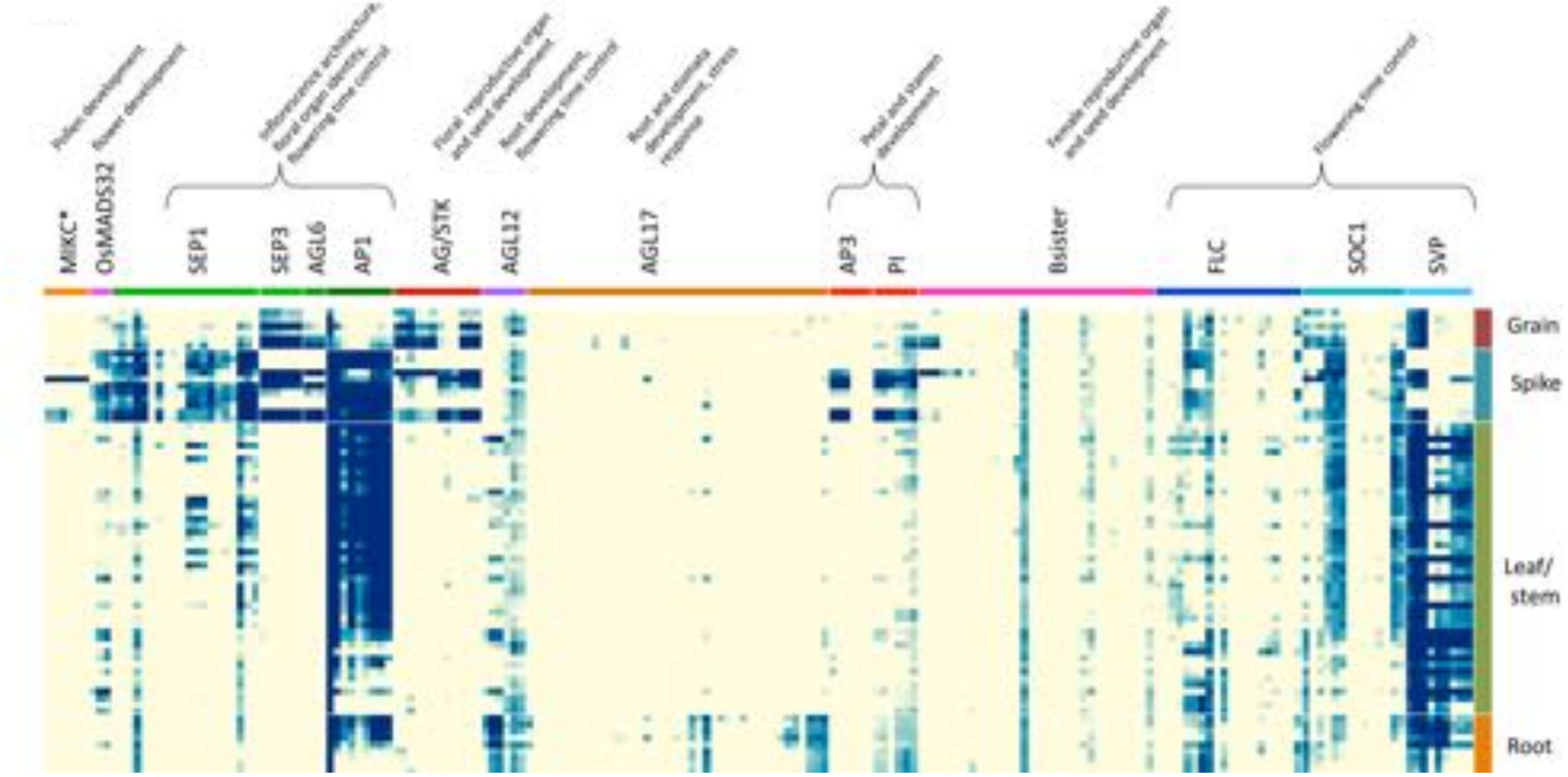
http://bar.utoronto.ca/eplant_wheat/



Schilling et al. 2020

MADS-box genes are highly conserved in wheat

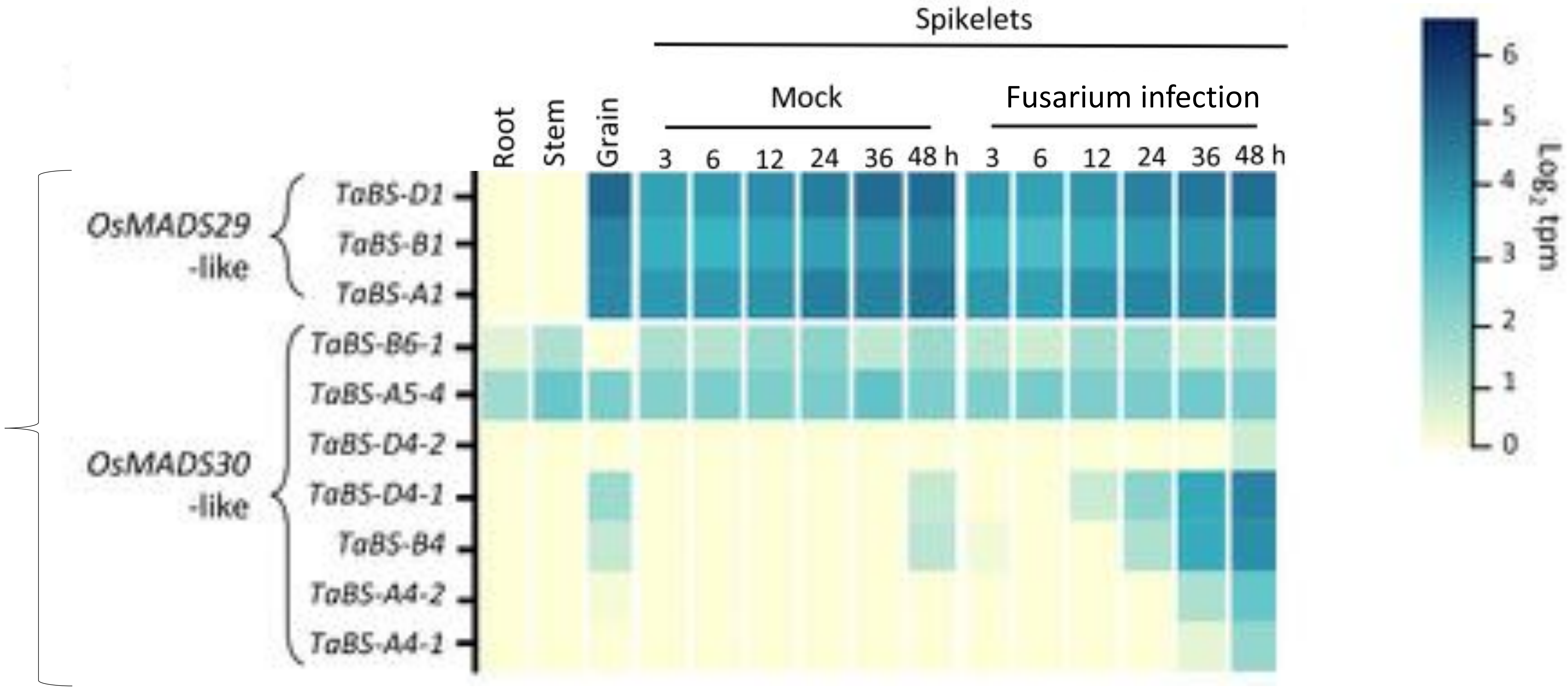
Schilling et al. 2020



Rainer Melzer & Susanne Schilling, UCD, Ireland

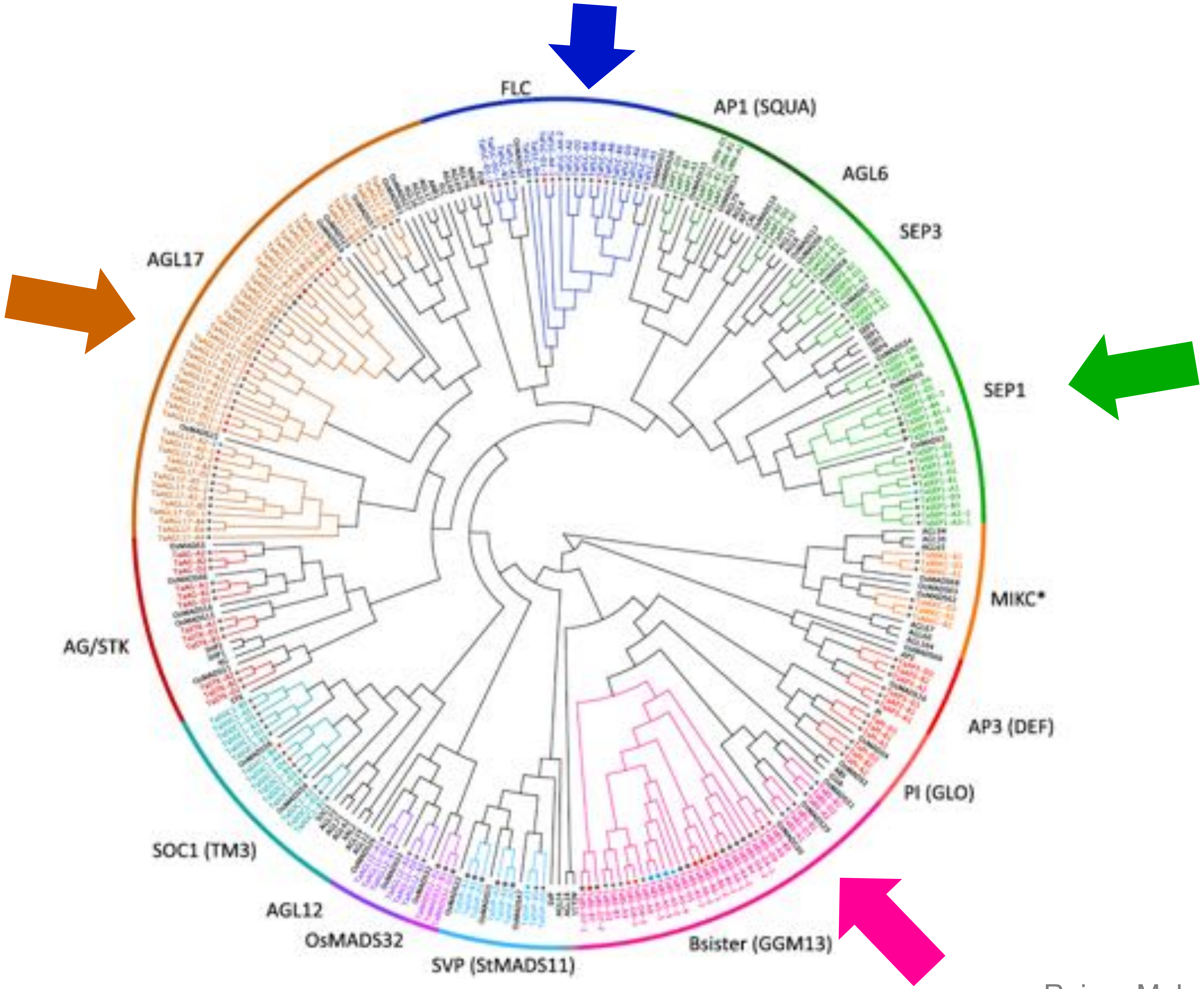
Putative neofunctionalization of MADS-box genes?

Bsister



Schilling et al. 2020

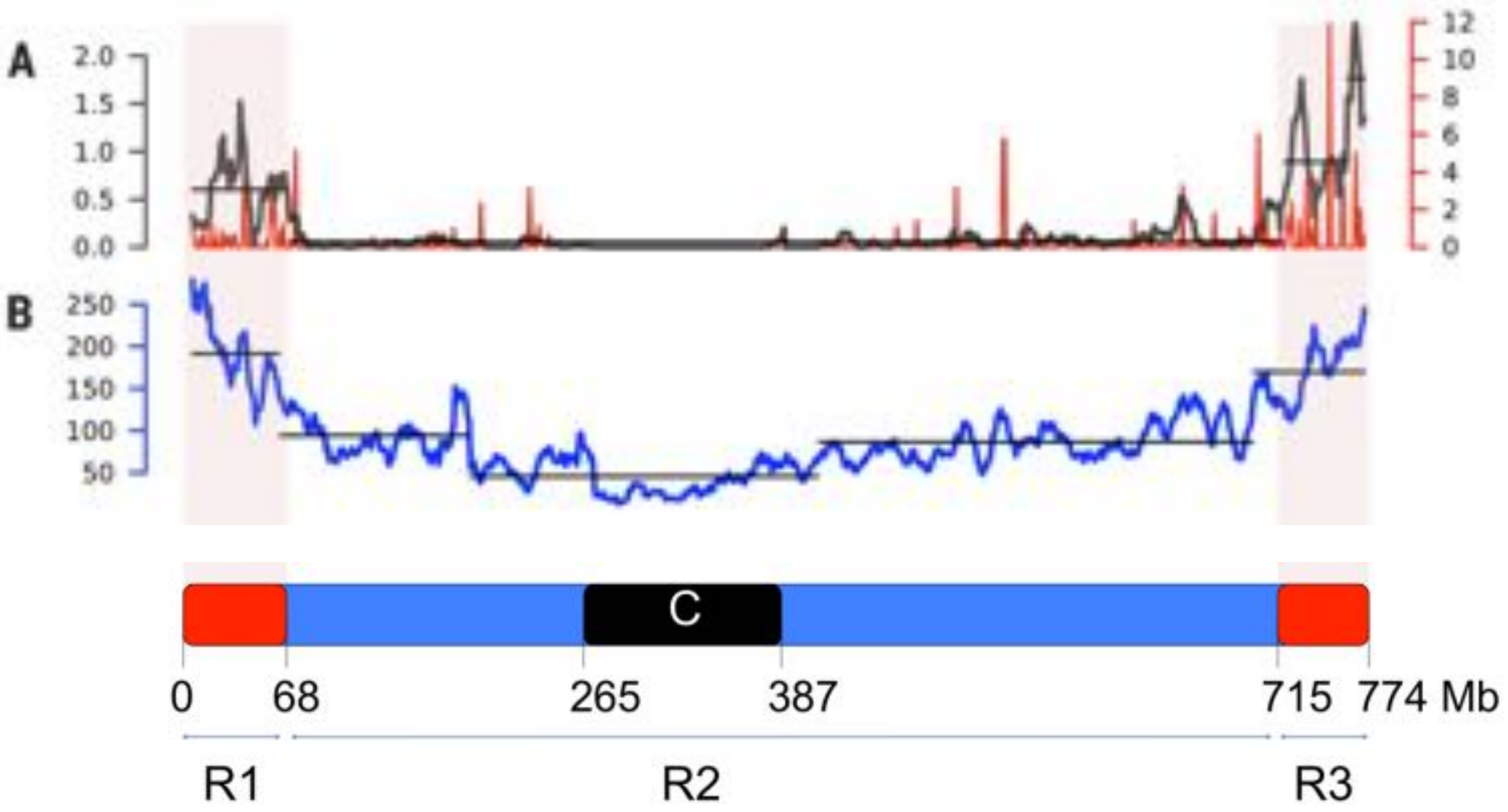
Some MADS-box gene subfamilies underwent expansion in wheat



Schilling et al. 2020

Rainer Melzer & Susanne Schilling, UCD, Ireland

Some MADS-box gene clades have undergone expansion in wheat – how?

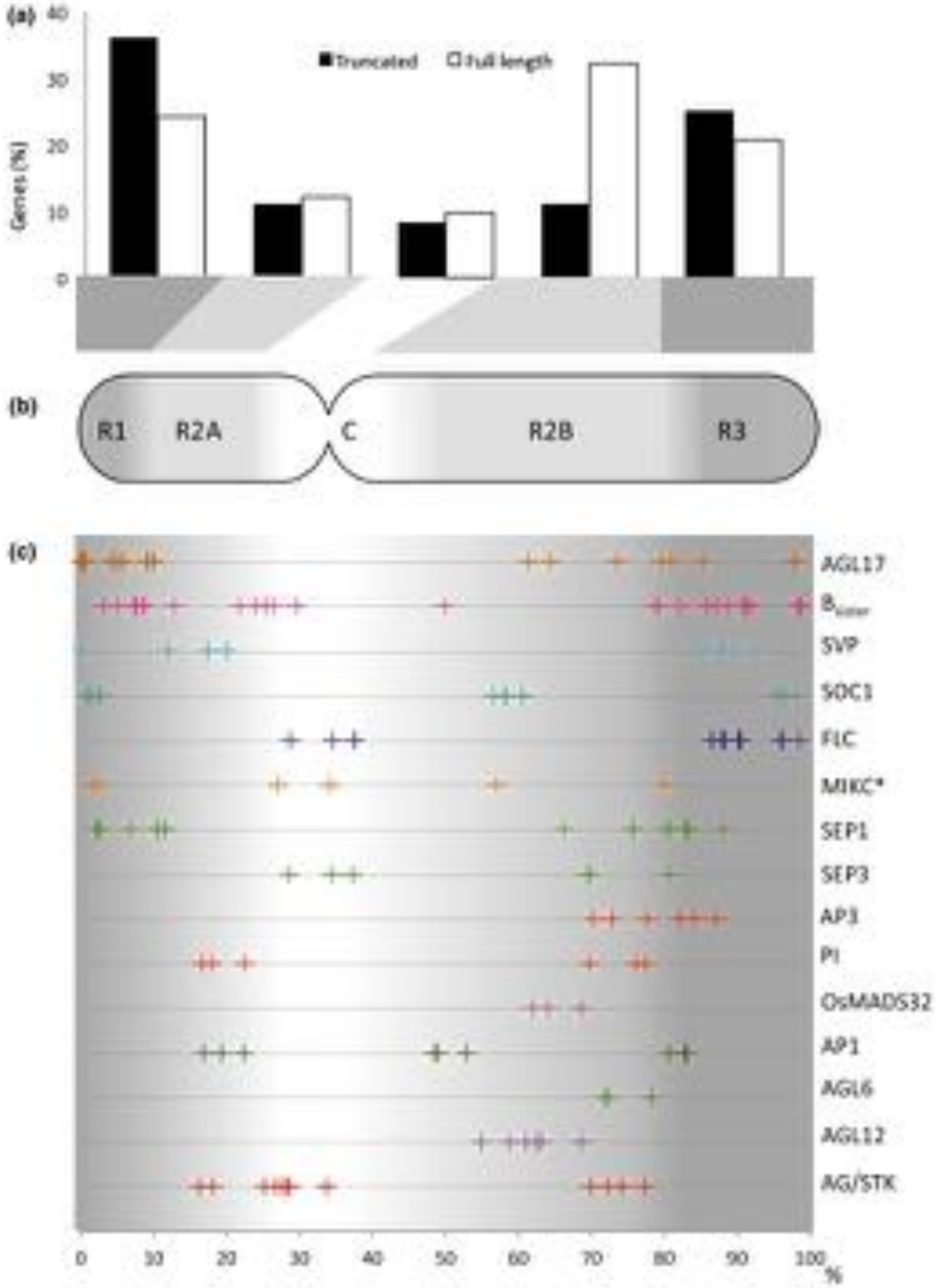


Recombination rate

Gene density

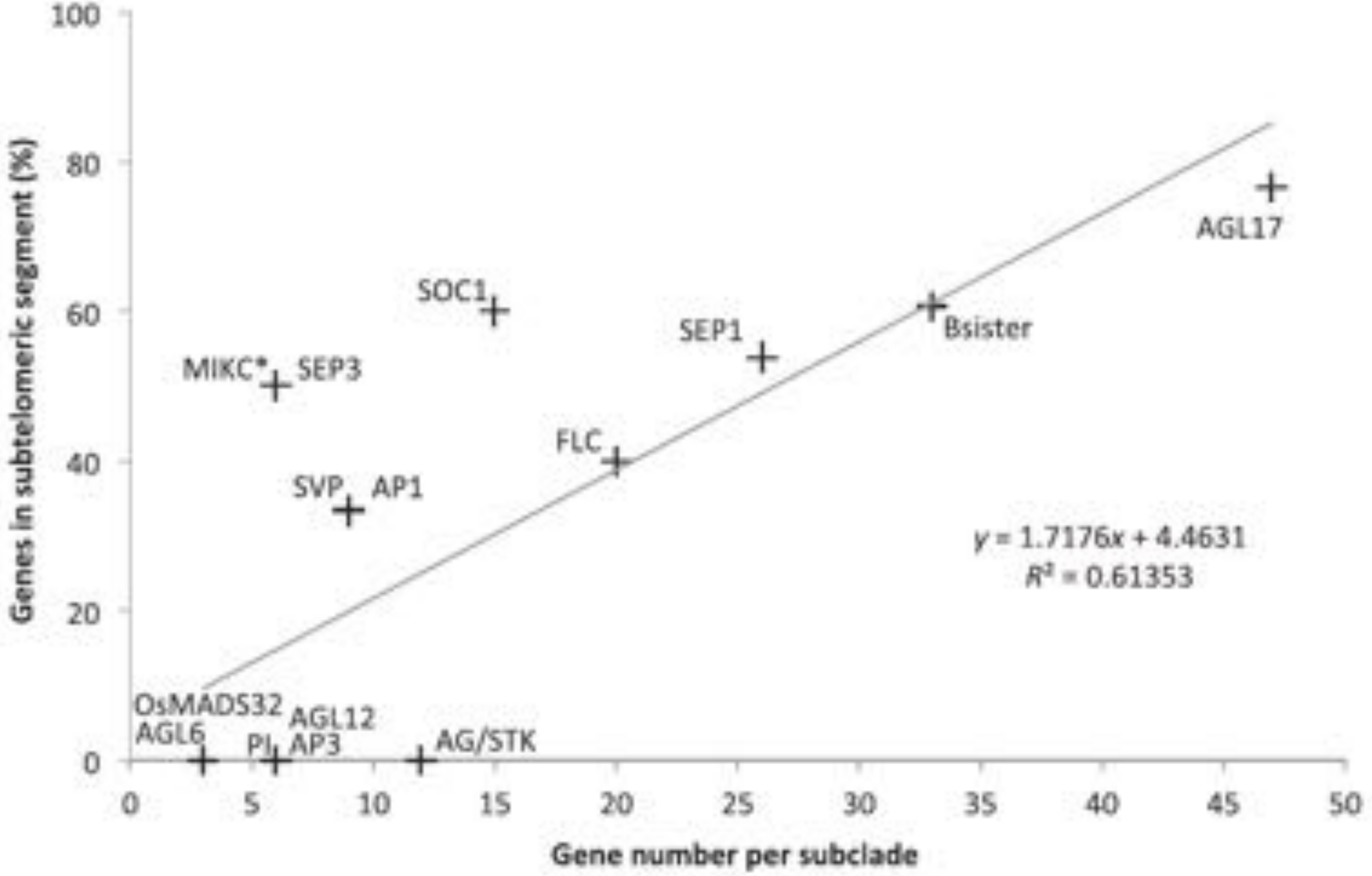
Choulet et al., 2014

Chromosomal distribution of wheat MADS-box genes



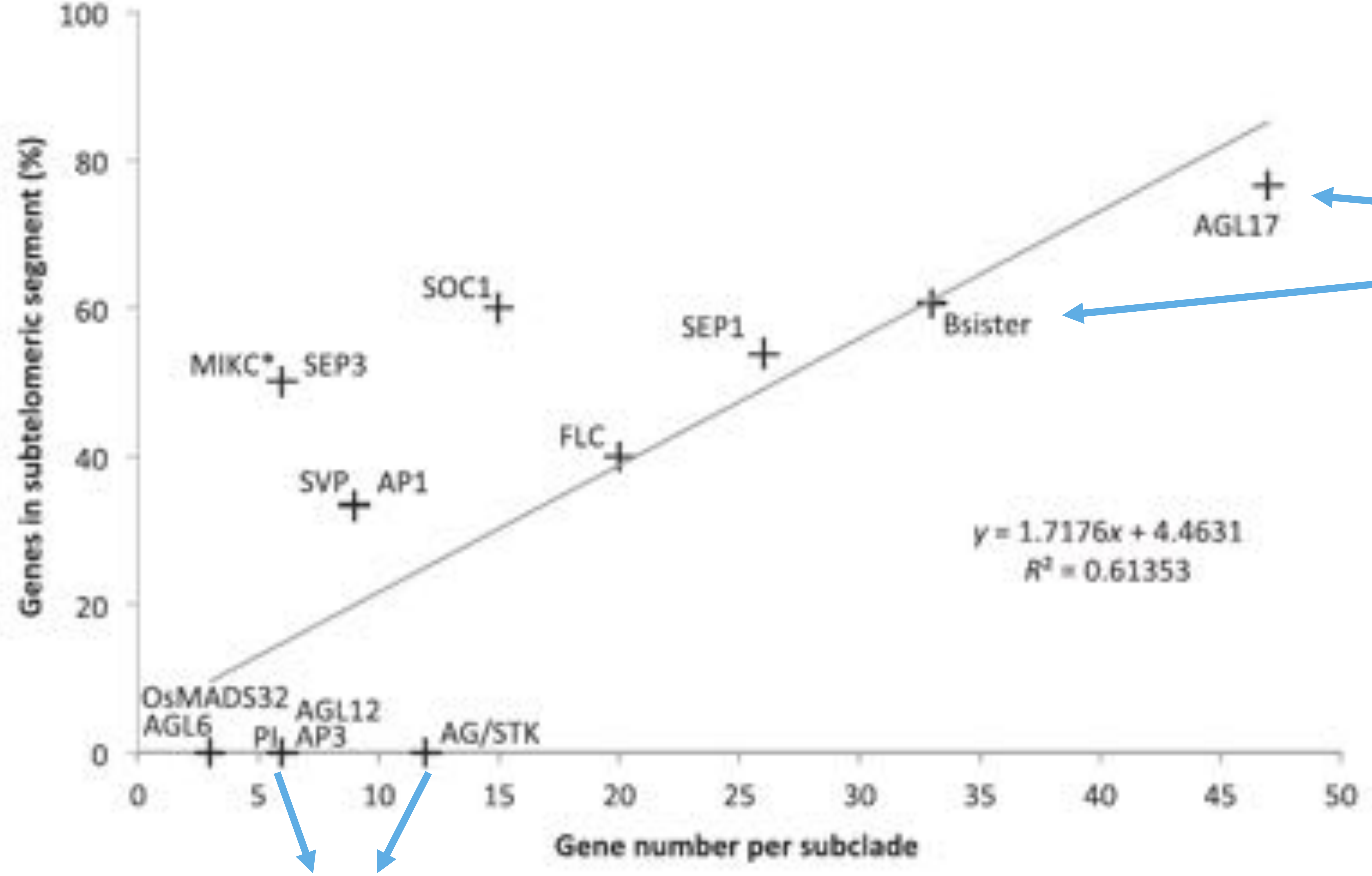
Schilling et al. 2020

Subtelomeric segments as hot spots for MADS-box gene evolution?



Schilling et al. 2020

Subtelomeric segments as hot spots for MADS-box gene evolution?



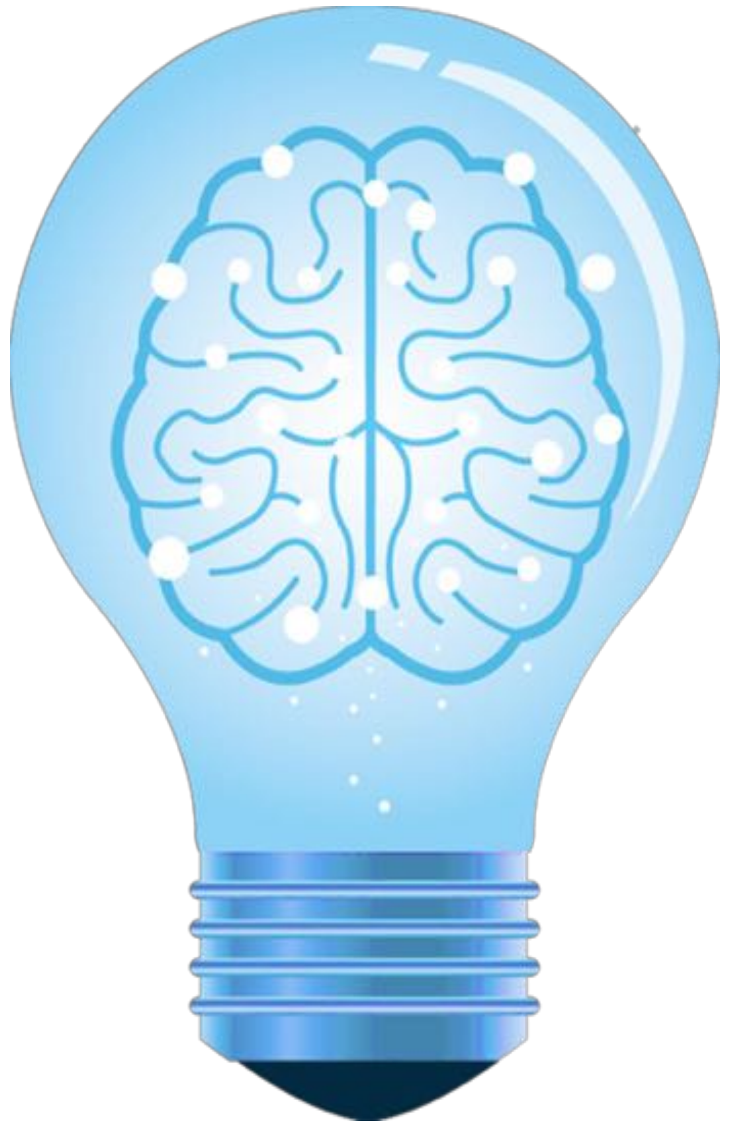
Putative neo-functionalization

Conserved core functions

Schilling et al. 2020

Summary I

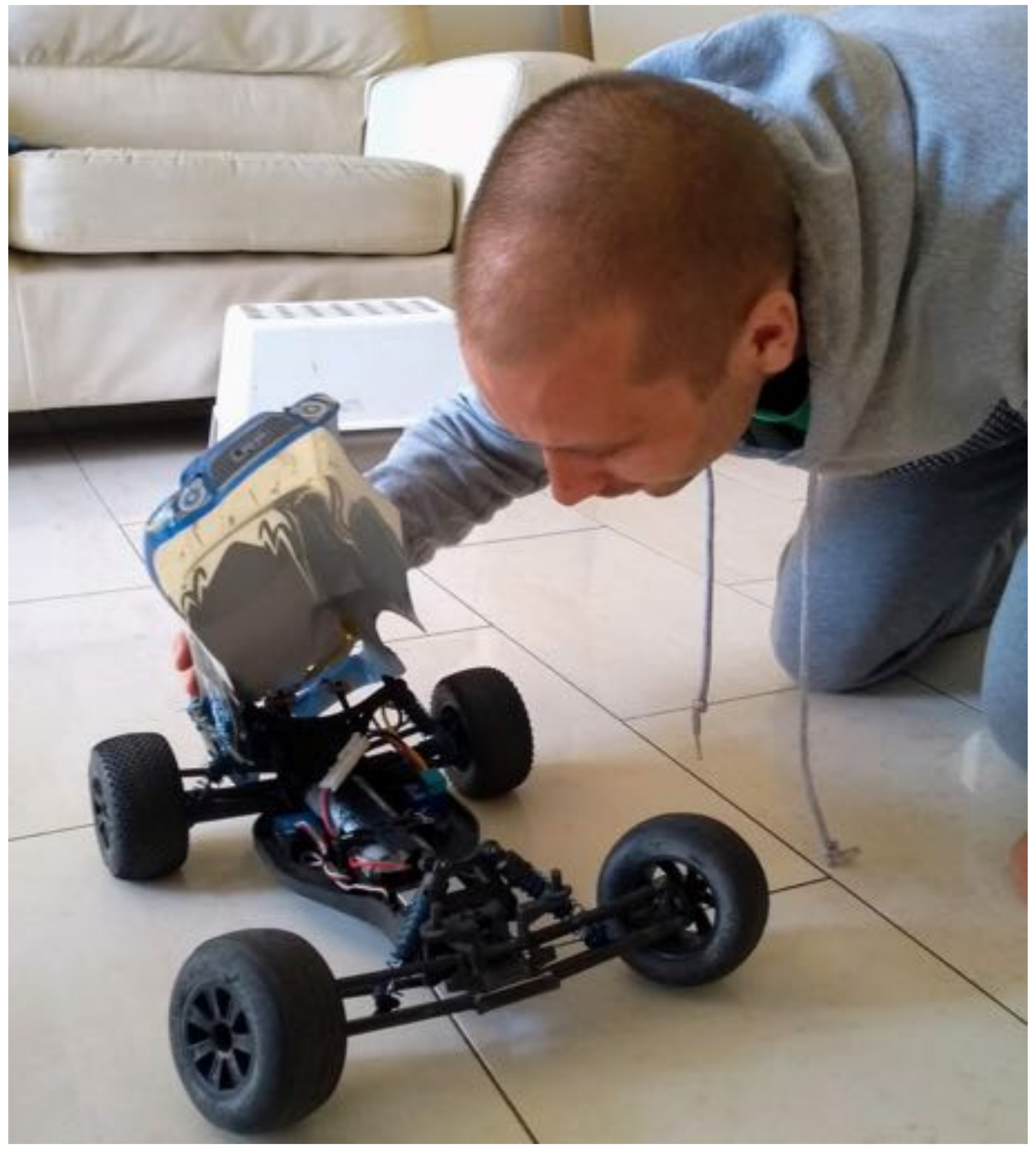
- MADS-box genes are key players in wheat development
- Conserved sequence and expression pattern
- MADS-box genes might have contributed to the success of wheat by neo- and subfunctionalization
- Good candidates for crop improvement



Genome-wide analysis of a wheat transcription factor family:

The power of bioinformatics resources

Part II: a look under the hood



bioinformatics
???
Wheat
resources

From Genes to Phylogenies

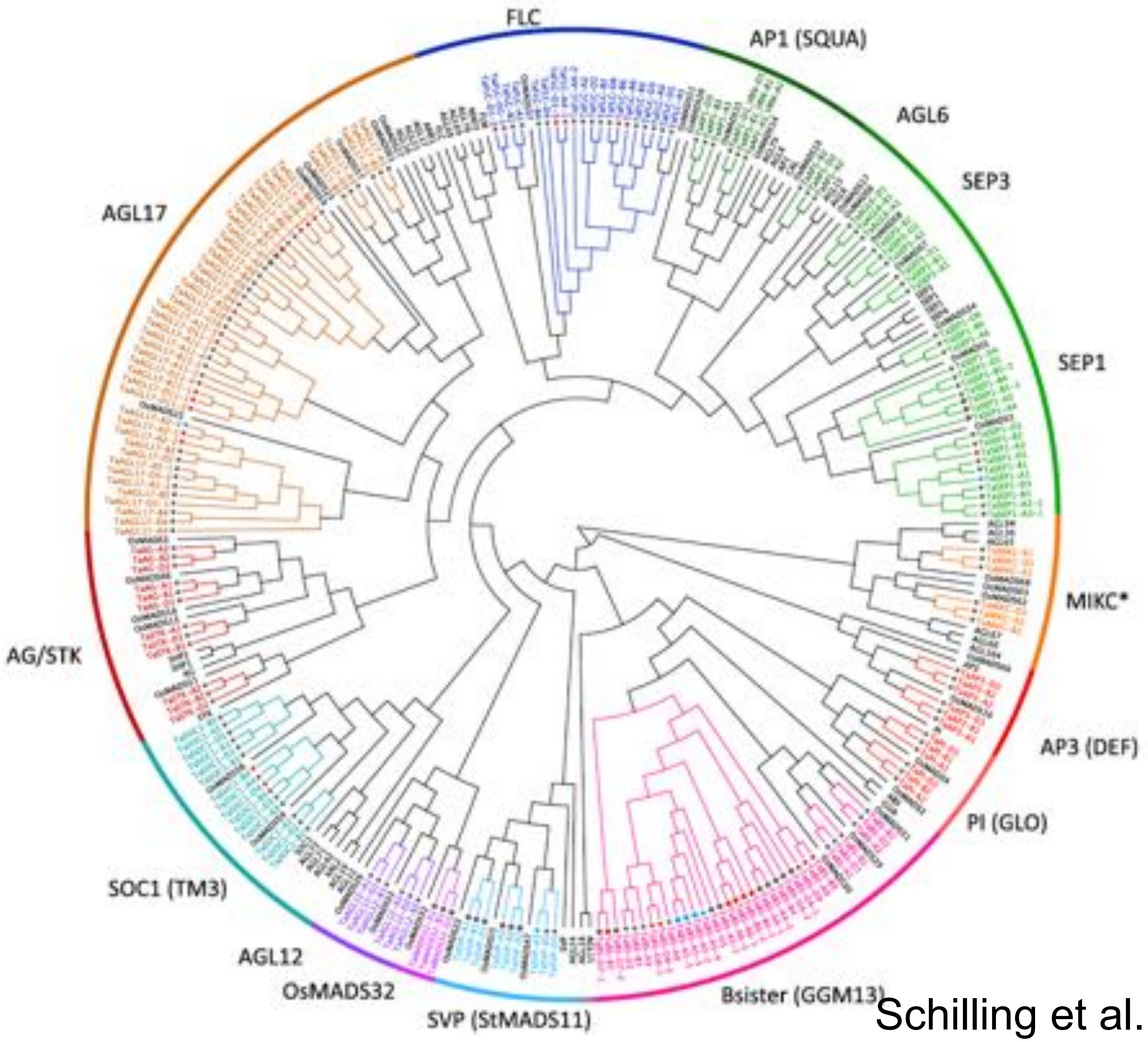
gene mining and identification

↓

filtering and sorting

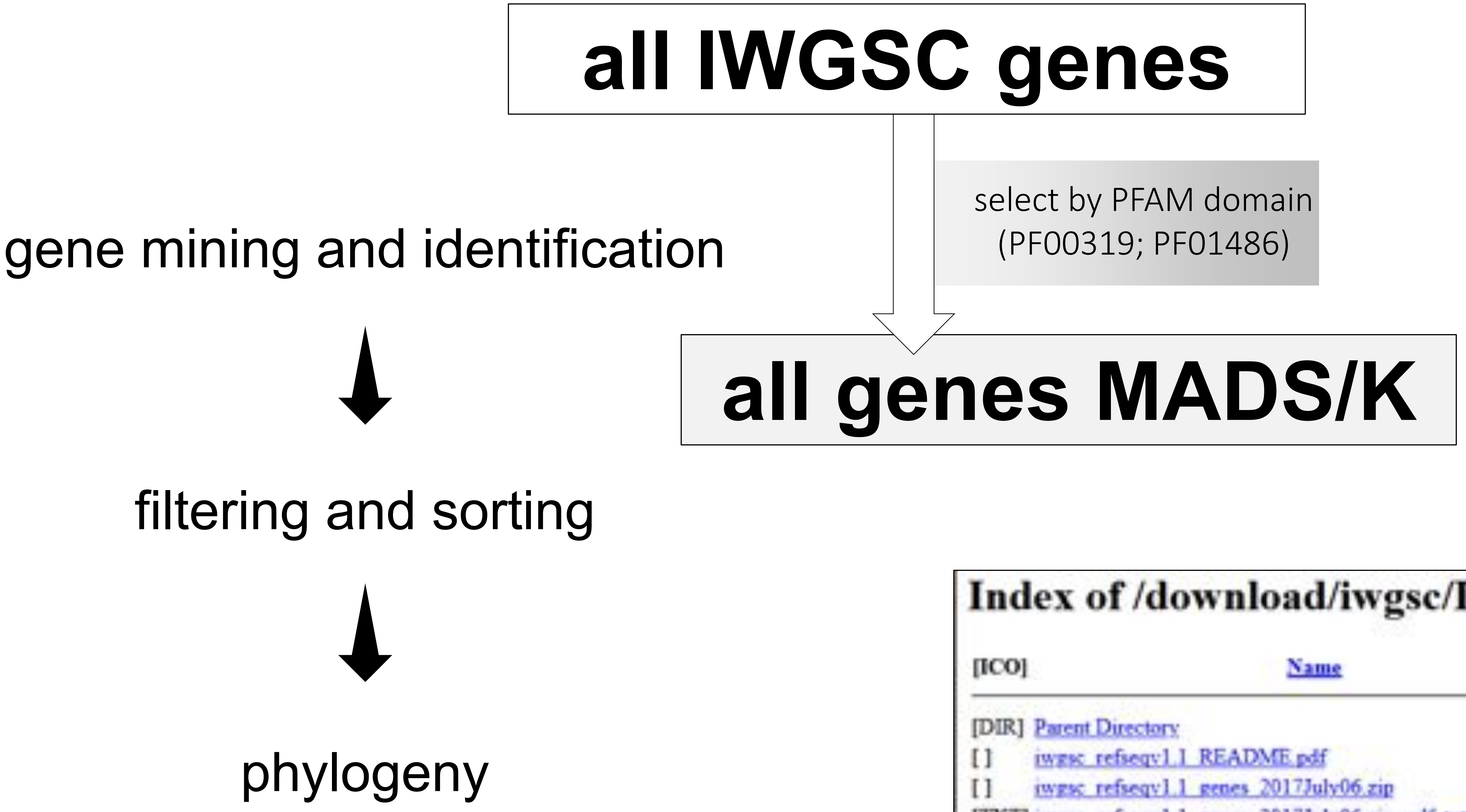
↓

phylogeny



Schilling et al. 2020

From Genes to Phylogenies



all IWGSC genes

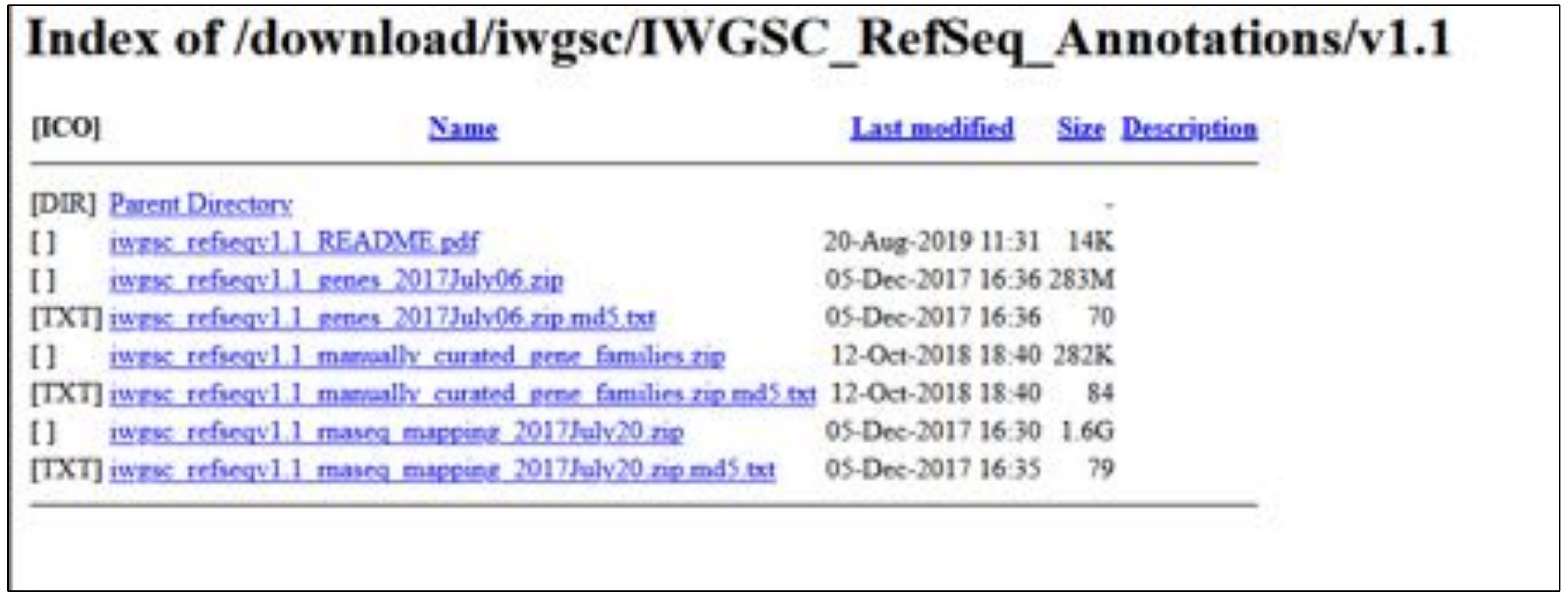
select by PFAM domain
(PF00319; PF01486)

all genes MADS/K

gene mining and identification

filtering and sorting

phylogeny

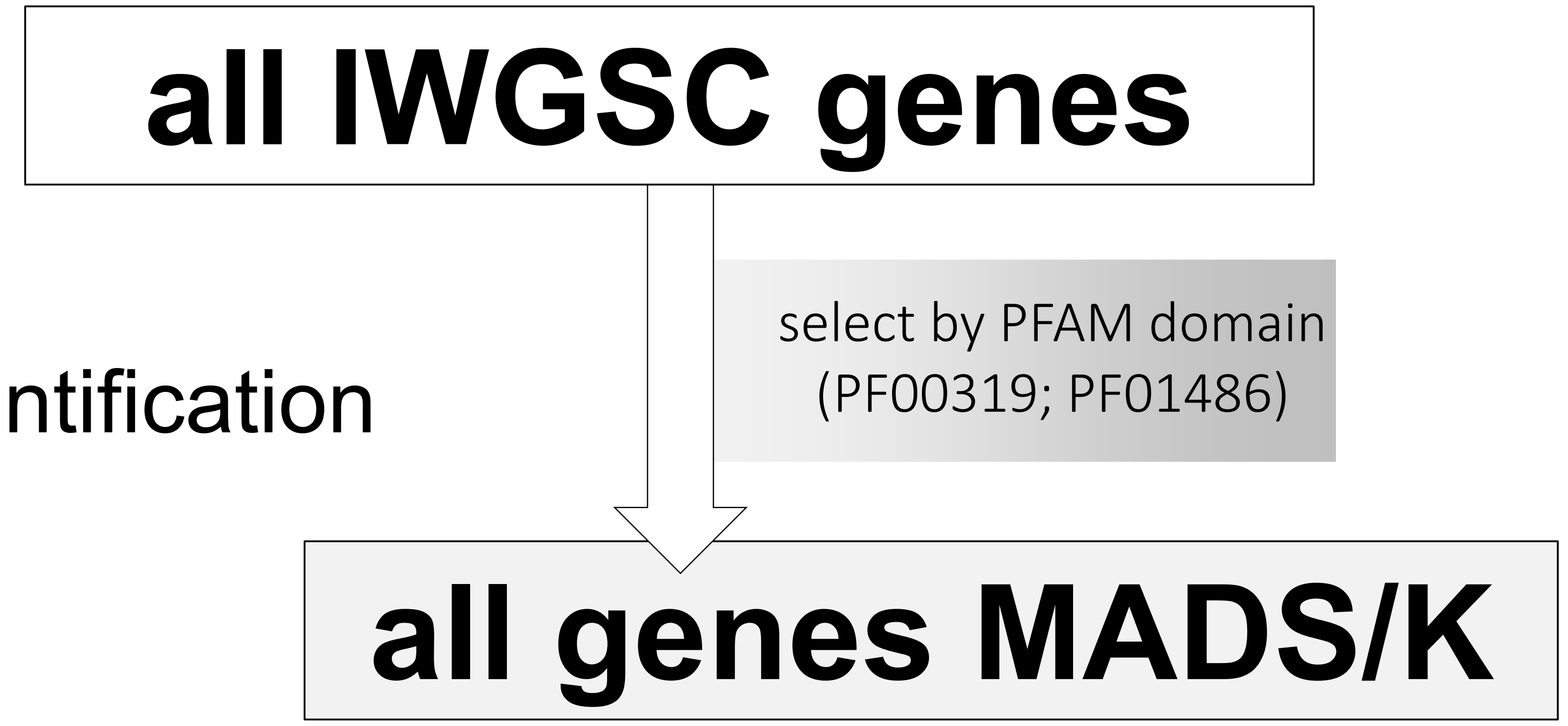


Index of /download/iwgsc/IWGSC_RefSeq_Annotations/v1.1

```
[ICO]                      Name                                  Last modified     Size  Description
[DIR] Parent Directory                                                                -
[ ]   iwgsc_refseqv1.1_README.pdf                              20-Aug-2019 11:31   14K
[ ]   iwgsc_refseqv1.1_genes_2017July06.zip                    05-Dec-2017 16:36  283M
[TXT] iwgsc_refseqv1.1_genes_2017July06.zip.md5.txt            05-Dec-2017 16:36    70
[ ]   iwgsc_refseqv1.1_manually_curated_gene_families.zip      12-Oct-2018 18:40  282K
[TXT] iwgsc_refseqv1.1_manually_curated_gene_families.zip.md5.txt 12-Oct-2018 18:40    84
[ ]   iwgsc_refseqv1.1_rnaseq_mapping_2017July20.zip           05-Dec-2017 16:30  1.6G
[TXT] iwgsc_refseqv1.1_rnaseq_mapping_2017July20.zip.md5.txt   05-Dec-2017 16:35    79
```

https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Annotations/v1.1/

From Genes to Phylogenies



all IWGSC genes

select by PFAM domain (PF00319; PF01486)

all genes MADS/K

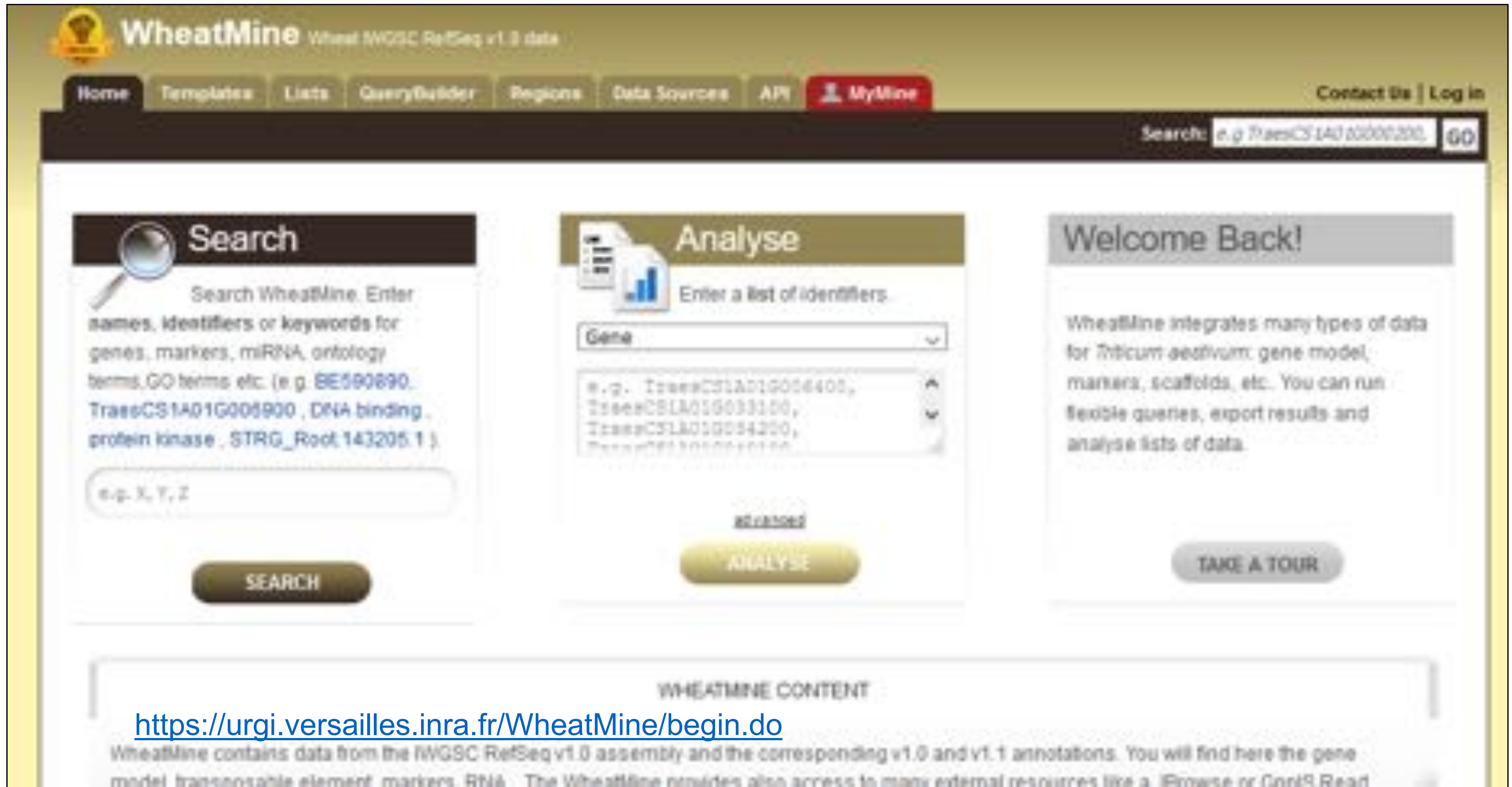
gene mining and identification

↓

filtering and sorting

↓

phylogeny



Search WheatMine. Enter **names**, **identifiers** or **keywords** for genes, markers, miRNA, ontology terms,GO terms etc. (e.g. BE590890, TraesCS1A01G006900 , DNA binding , protein kinase , STRG_Root.143205.1).

e.g. X, Y, Z

Enter a list of identifiers.

e.g. TraesCS1A01G006400, TraesCS1A01G033100, TraesCS1A01G054200,

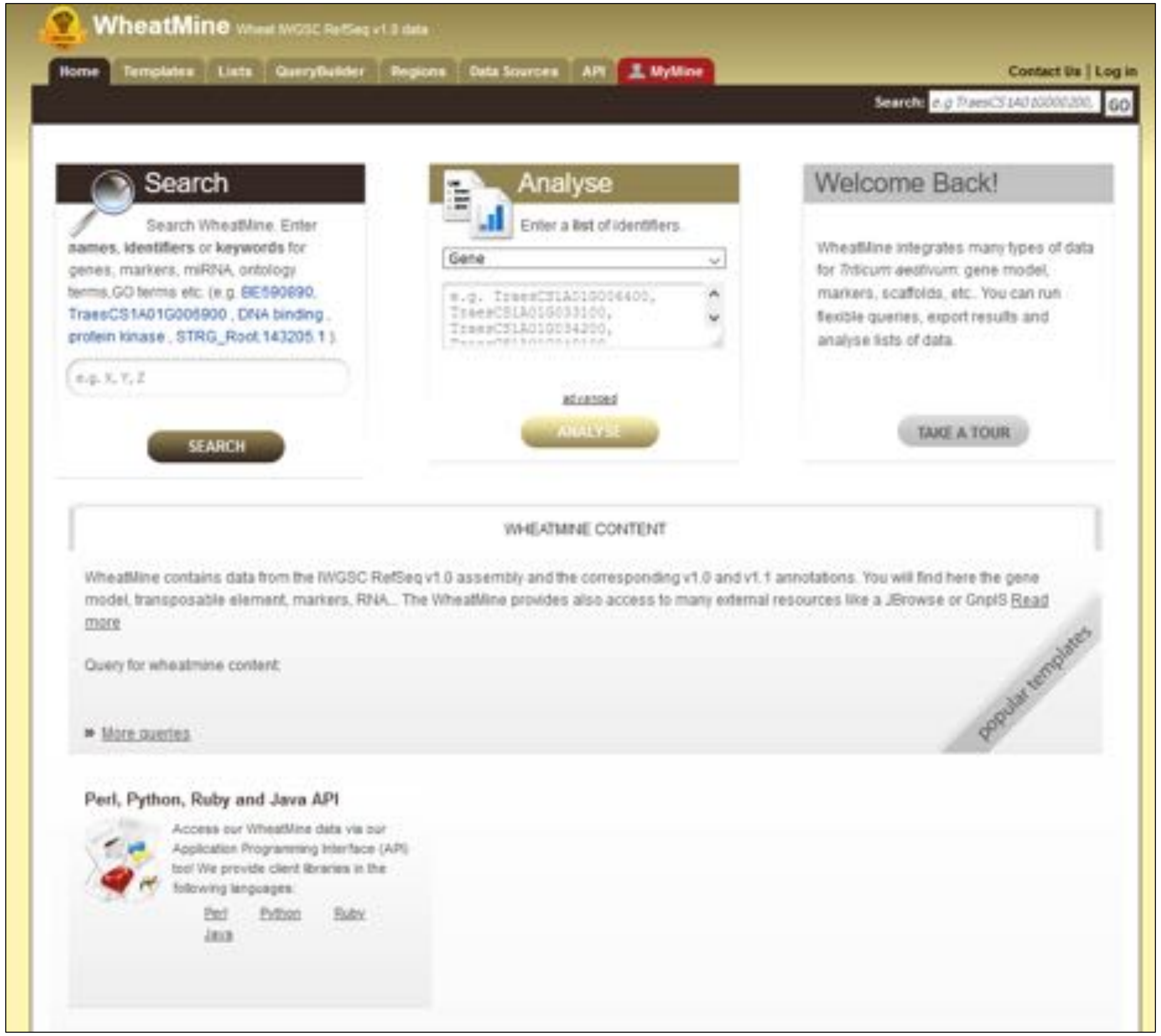
WheatMine integrates many types of data for *Triticum aestivum*: gene model, markers, scaffolds, etc.. You can run flexible queries, export results and analyse lists of data.

WHEATMINE CONTENT

https://urgi.versailles.inra.fr/WheatMine/begin.do

WheatMine contains data from the IWGSC RefSeq v1.0 assembly and the corresponding v1.0 and v1.1 annotations. You will find here the gene model, transposable element, markers, RNA... The WheatMine provides also access to many external resources like a JBrowse or GnpIS Read

IWGSC Wheat mine



Search WheatMine. Enter **names, identifiers** or **keywords** for genes, markers, miRNA, ontology terms,GO terms etc. (e.g. BE590890, TraesCS1A01G006900 , DNA binding , protein kinase , STRG_Root.143205.1)

Enter a list of identifiers.

WheatMine integrates many types of data for *Triticum aestivum*: gene model, markers, scaffolds, etc.. You can run flexible queries, export results and analyse lists of data.

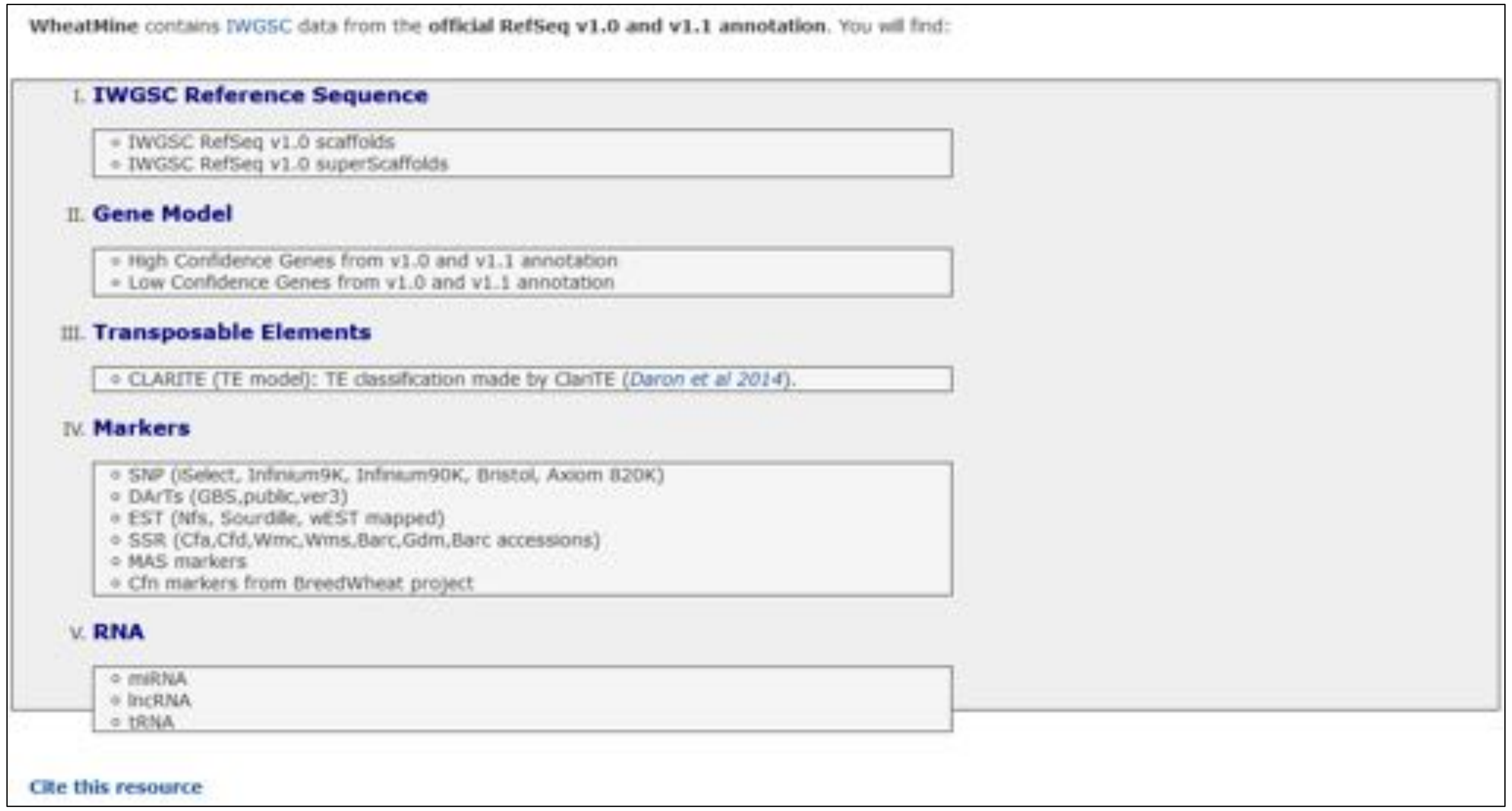
WheatMine contains data from the IWGSC RefSeq v1.0 assembly and the corresponding v1.0 and v1.1 annotations. You will find here the gene model, transposable element, markers, RNA... The WheatMine provides also access to many external resources like a JBrowse or GnpIS Read more

Query for wheatmine content:

Perl, Python, Ruby and Java API

Access our WheatMine data via our Application Programming Interface (API) too! We provide client libraries in the following languages:

https://urgi.versailles.inra.fr/WheatMine/begin.do



WheatMine contains IWGSC data from the **official RefSeq v1.0 and v1.1 annotation**. You will find:

I. **IWGSC Reference Sequence**
- IWGSC RefSeq v1.0 scaffolds
- IWGSC RefSeq v1.0 superScaffolds

II. **Gene Model**
- High Confidence Genes from v1.0 and v1.1 annotation
- Low Confidence Genes from v1.0 and v1.1 annotation

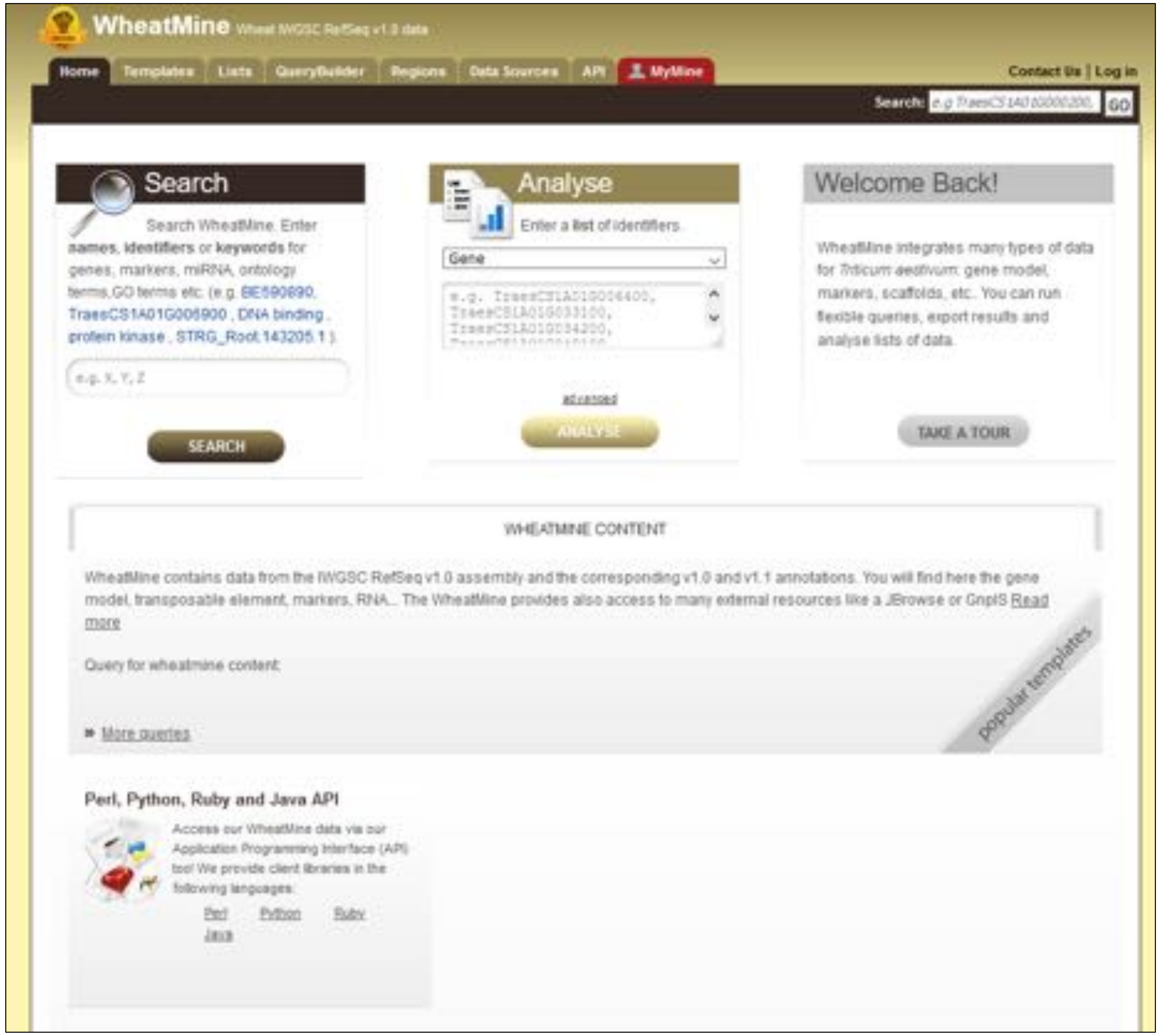
III. **Transposable Elements**
- CLARITE (TE model): TE classification made by ClariTE (*Daron et al 2014*).

IV. **Markers**
- SNP (iSelect, Infinium9K, Infinium90K, Bristol, Axiom 820K)
- DArTs (GBS,public,ver3)
- EST (Nfs, Sourdille, wEST mapped)
- SSR (Cfa,Cfd,Wmc,Wms,Barc,Gdm,Barc accessions)
- MAS markers
- Cfn markers from BreedWheat project

V. **RNA**
- miRNA
- lncRNA
- tRNA

Cite this resource

IWGSC Wheat mine



Search

Search WheatMine. Enter **names, identifiers** or **keywords** for genes, markers, miRNA, ontology terms,GO terms etc. (e.g. BE590890, TraesCS1A01G006900 , DNA binding , protein kinase , STRG_Root.143205.1)

Analyse

Enter a list of identifiers

Welcome Back!

WheatMine integrates many types of data for *Triticum aestivum*: gene model, markers, scaffolds, etc.. You can run flexible queries, export results and analyse lists of data.

WHEATMINE CONTENT

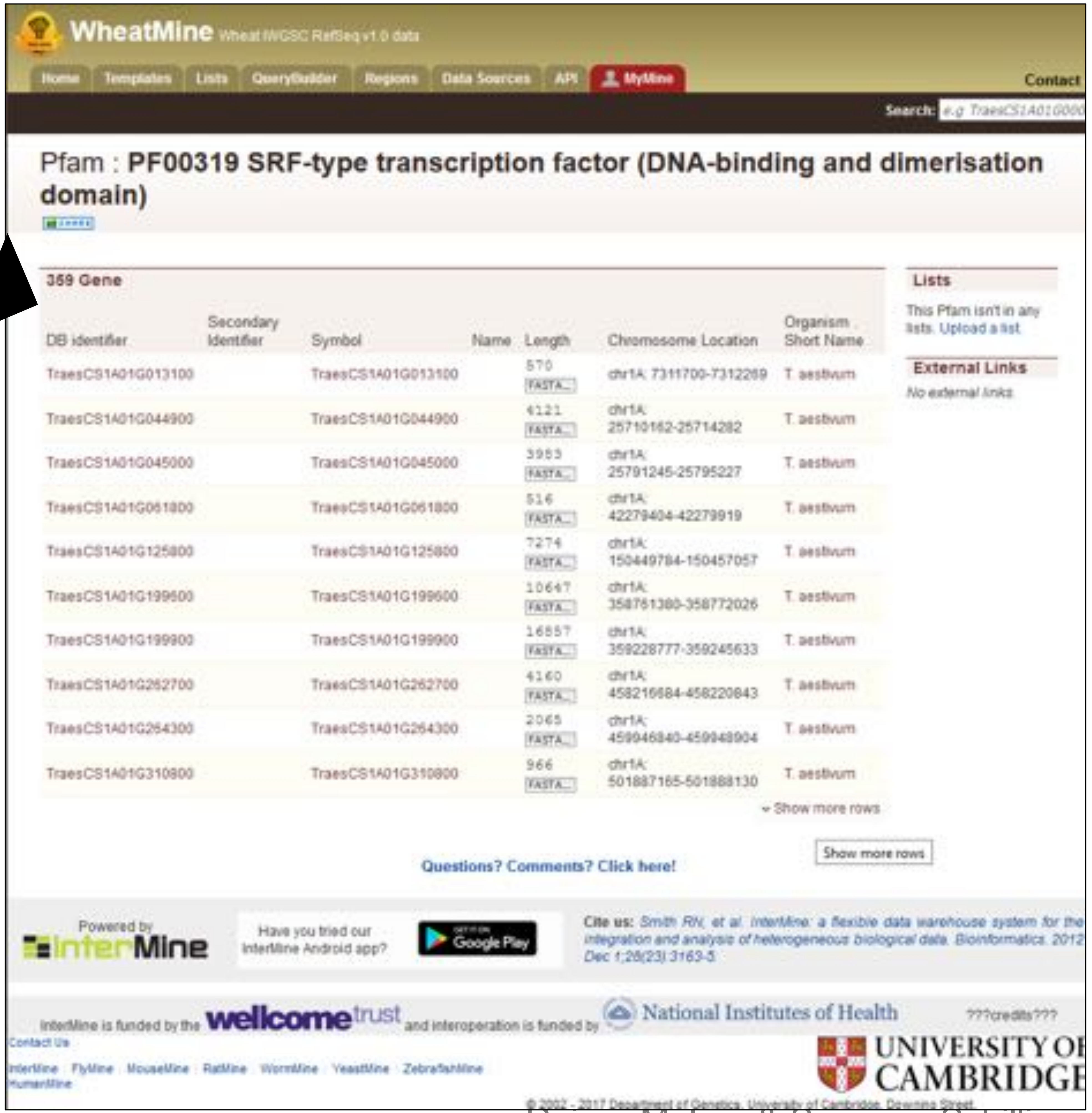
WheatMine contains data from the IWGSC RefSeq v1.0 assembly and the corresponding v1.0 and v1.1 annotations. You will find here the gene model, transposable element, markers, RNA... The WheatMine provides also access to many external resources like a JBrowse or GnpIS Read more

Query for wheatmine content:

Perl, Python, Ruby and Java API

Access our WheatMine data via our Application Programming Interface (API) too! We provide client libraries in the following languages:

https://urgi.versailles.inra.fr/WheatMine/begin.do


WheatMine Wheat IWGSC RefSeq v1.0 data
Home
Templates
Lists
QueryBuilder
Regions
Data Sources
API
MyMine
Contact
Search: e.g TraesCS1A01G000

Pfam : **PF00319 SRF-type transcription factor (DNA-binding and dimerisation domain)**

369 Gene

DB identifier	Secondary Identifier	Symbol	Name	Length	Chromosome Location	Organism . Short Name
TraesCS1A01G013100		TraesCS1A01G013100		570 FASTA...	chr1A: 7311700-7312269	T. aestivum
TraesCS1A01G044900		TraesCS1A01G044900		4121 FASTA...	chr1A: 25710162-25714282	T. aestivum
TraesCS1A01G045000		TraesCS1A01G045000		3983 FASTA...	chr1A: 25791245-25795227	T. aestivum
TraesCS1A01G061800		TraesCS1A01G061800		516 FASTA...	chr1A: 42279404-42279919	T. aestivum
TraesCS1A01G125800		TraesCS1A01G125800		7274 FASTA...	chr1A: 150449784-150457057	T. aestivum
TraesCS1A01G199600		TraesCS1A01G199600		10647 FASTA...	chr1A: 358761380-358772026	T. aestivum
TraesCS1A01G199900		TraesCS1A01G199900		16857 FASTA...	chr1A: 359228777-359245633	T. aestivum
TraesCS1A01G262700		TraesCS1A01G262700		4160 FASTA...	chr1A: 458216684-458220843	T. aestivum
TraesCS1A01G264300		TraesCS1A01G264300		2065 FASTA...	chr1A: 459946840-459948904	T. aestivum
TraesCS1A01G310800		TraesCS1A01G310800		966 FASTA...	chr1A: 501887165-501888130	T. aestivum

Lists

This Pfam isn't in any lists. Upload a list

External Links

No external links

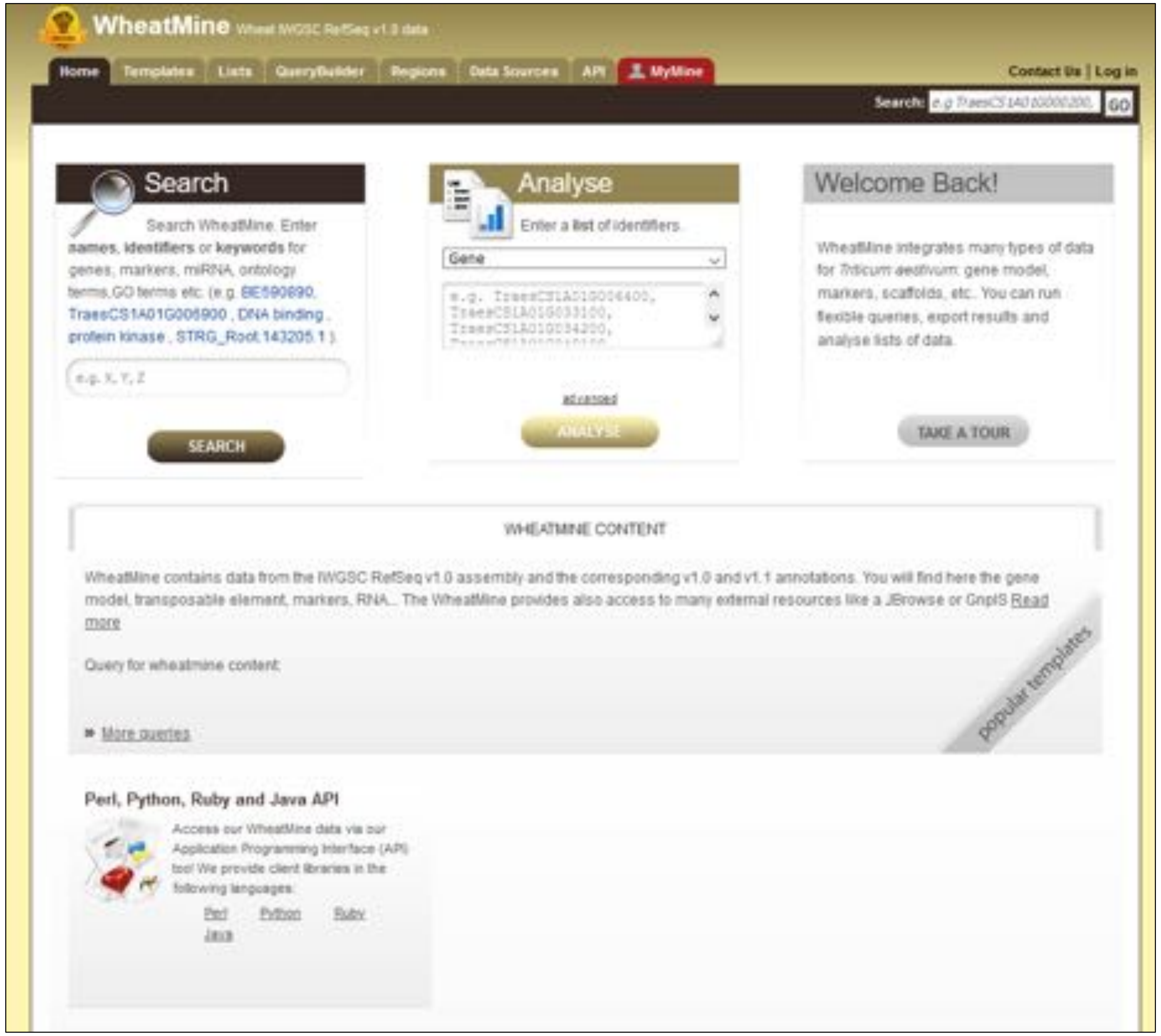
Questions? Comments? Click here!

Cite us: *Smith RN, et al. InterMine: a flexible data warehouse system for the integration and analysis of heterogeneous biological data. Bioinformatics. 2012 Dec 1;28(23):3163-5.*

InterMine is funded by the wellcome trust and interoperation is funded by National Institutes of Health ???credits???

© 2002 - 2017 Department of Genetics, University of Cambridge, Downing Street,

IWGSC Wheat mine



WheatMine Wheat IWGSC RefSeq v1.0 data

Search

Search WheatMine. Enter **names, identifiers** or **keywords** for genes, markers, miRNA, ontology terms,GO terms etc. (e.g. BE590890, TraesCS1A01G006900 , DNA binding , protein kinase , STRG_Root.143205.1)

Analyse

Enter a list of identifiers.

Welcome Back!

WheatMine integrates many types of data for *Triticum aestivum*: gene model, markers, scaffolds, etc.. You can run flexible queries, export results and analyse lists of data.

WHEATMINE CONTENT

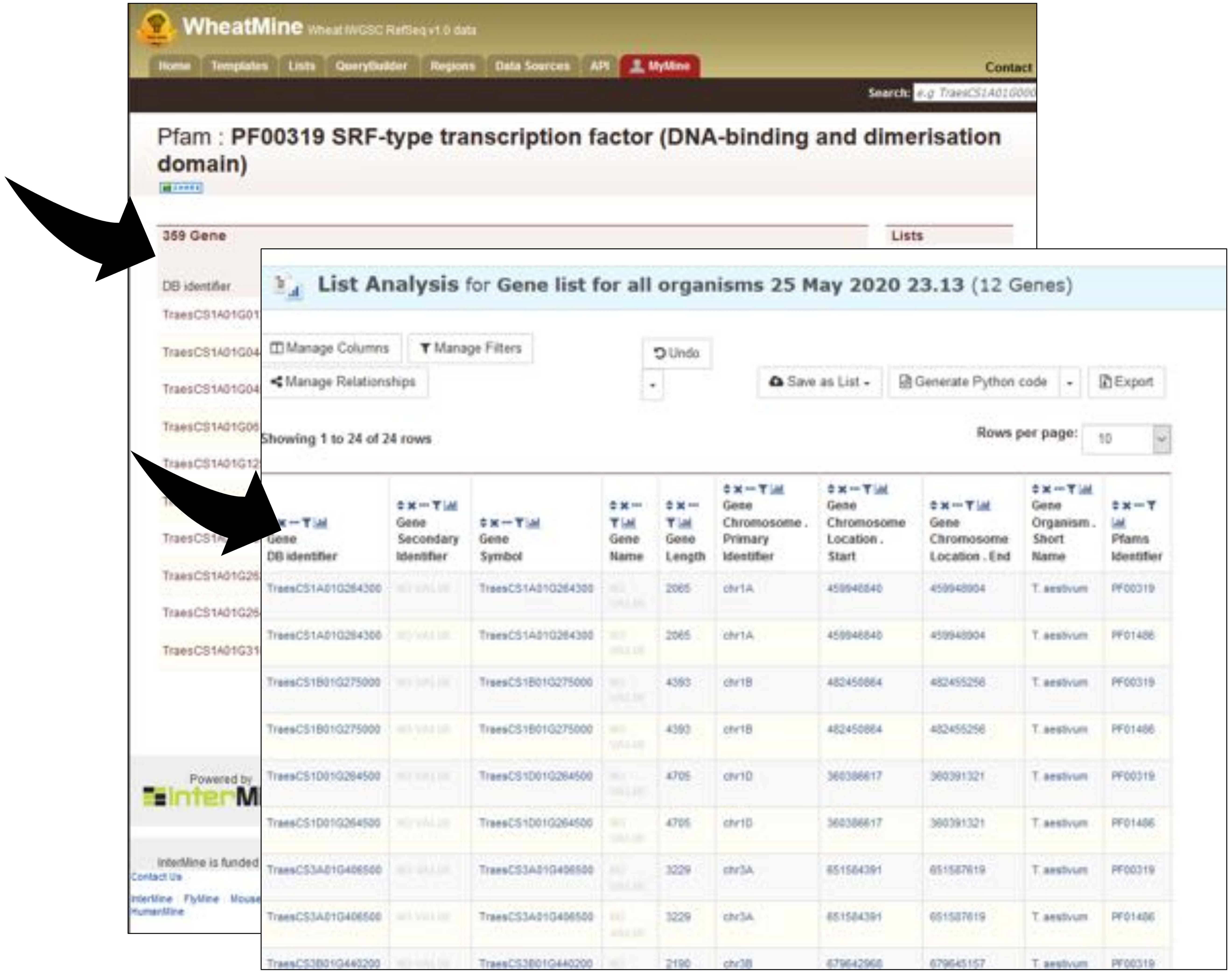
WheatMine contains data from the IWGSC RefSeq v1.0 assembly and the corresponding v1.0 and v1.1 annotations. You will find here the gene model, transposable element, markers, RNA... The WheatMine provides also access to many external resources like a JBrowse or GnpIS Read more

Query for wheatmine content:

Perl, Python, Ruby and Java API

Access our WheatMine data via our Application Programming Interface (API) too! We provide client libraries in the following languages:

https://urgi.versailles.inra.fr/WheatMine/begin.do



WheatMine Wheat IWGSC RefSeq v1.0 data

Pfam : **PF00319 SRF-type transcription factor (DNA-binding and dimerisation domain)**

359 Gene

Lists

DB identifier

TraesCS1A01G01
TraesCS1A01G04
TraesCS1A01G04
TraesCS1A01G06
TraesCS1A01G12
TraesCS1A
TraesCS1A01G26
TraesCS1A01G26
TraesCS1A01G31

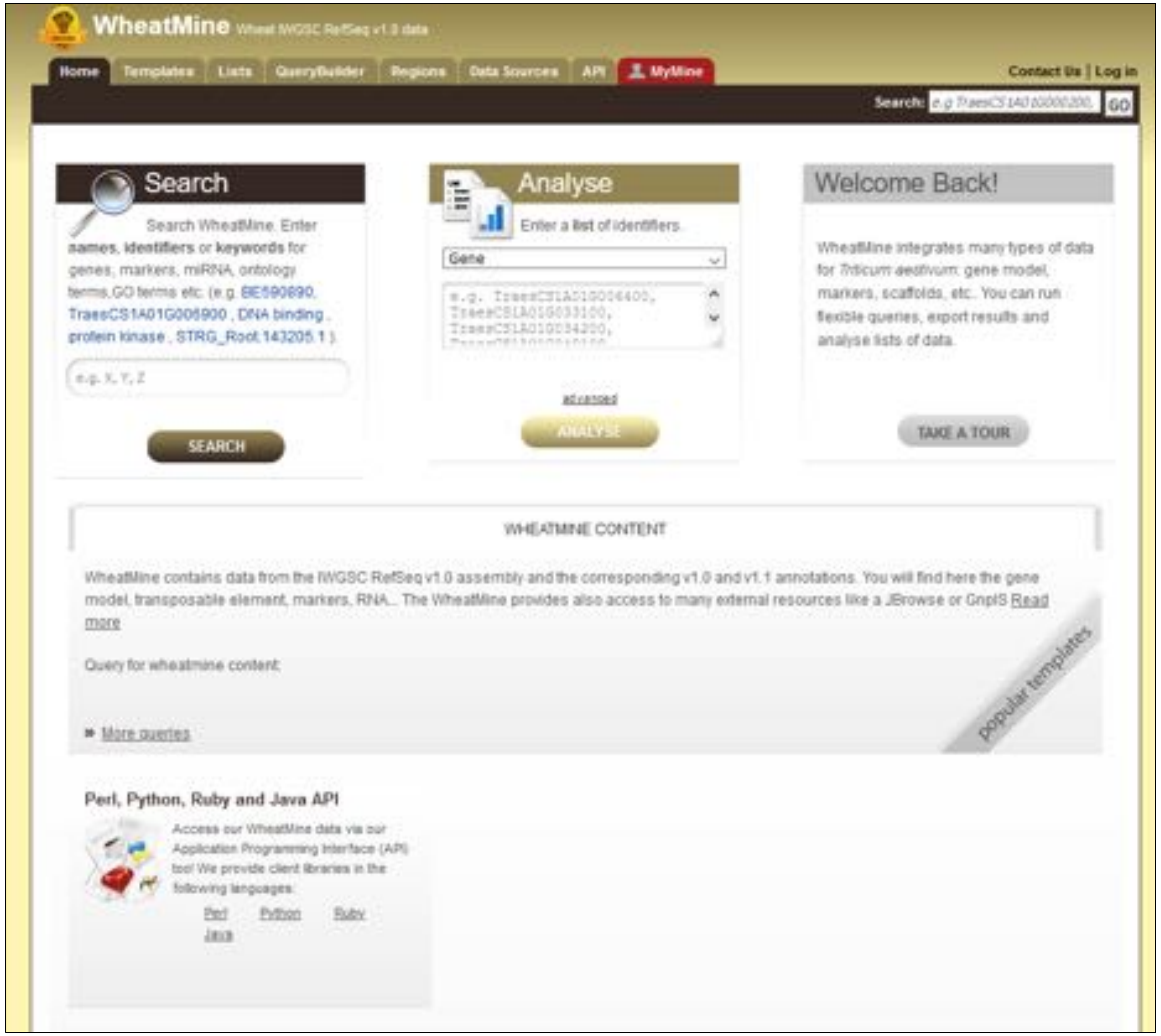
List Analysis for **Gene list for all organisms 25 May 2020 23.13** (12 Genes)

Showing 1 to 24 of 24 rows

Rows per page: 10

Gene DB identifier	Gene Secondary Identifier	Gene Symbol	Gene Name	Gene Length	Gene Chromosome . Primary Identifier	Gene Chromosome Location . Start	Gene Chromosome Location . End	Gene Organism . Short Name	Pfams Identifier
TraesCS1A01G264300	NO VALUE	TraesCS1A01G264300	NO VALUE	2065	chr1A	459946840	459948904	T. aestivum	PF00319
TraesCS1A01G264300	NO VALUE	TraesCS1A01G264300	NO VALUE	2065	chr1A	459946840	459948904	T. aestivum	PF01486
TraesCS1B01G275000	NO VALUE	TraesCS1B01G275000	NO VALUE	4393	chr1B	482450864	482455256	T. aestivum	PF00319
TraesCS1B01G275000	NO VALUE	TraesCS1B01G275000	NO VALUE	4393	chr1B	482450864	482455256	T. aestivum	PF01486
TraesCS1D01G264500	NO VALUE	TraesCS1D01G264500	NO VALUE	4705	chr1D	360386617	360391321	T. aestivum	PF00319
TraesCS1D01G264500	NO VALUE	TraesCS1D01G264500	NO VALUE	4705	chr1D	360386617	360391321	T. aestivum	PF01486
TraesCS3A01G406500	NO VALUE	TraesCS3A01G406500	NO VALUE	3229	chr3A	651584391	651587619	T. aestivum	PF00319
TraesCS3A01G406500	NO VALUE	TraesCS3A01G406500	NO VALUE	3229	chr3A	651584391	651587619	T. aestivum	PF01486
TraesCS3B01G440200	NO VALUE	TraesCS3B01G440200	NO	2190	chr3B	679642968	679645157	T. aestivum	PF00319

Data Mining - IWGSC Wheat mine



Search

Search WheatMine. Enter **names, identifiers** or **keywords** for genes, markers, miRNA, ontology terms,GO terms etc. (e.g. BE590890, TraesCS1A01G006900 , DNA binding , protein kinase , STRG_Root.143205.1).

e.g. X, Y, Z

SEARCH

Analyse

Enter a **list** of identifiers.

Gene

e.g. TraesCS1A01G006400, TraesCS1A01G033100, TraesCS1A01G034200,

advanced

ANALYSE

Welcome Back!

WheatMine integrates many types of data for *Triticum aestivum*: gene model, markers, scaffolds, etc.. You can run flexible queries, export results and analyse lists of data.

TAKE A TOUR

WHEATMINE CONTENT

WheatMine contains data from the IWGSC RefSeq v1.0 assembly and the corresponding v1.0 and v1.1 annotations. You will find here the gene model, transposable element, markers, RNA... The WheatMine provides also access to many external resources like a JBrowse or GnpIS Read more

Query for wheatmine content

» More queries

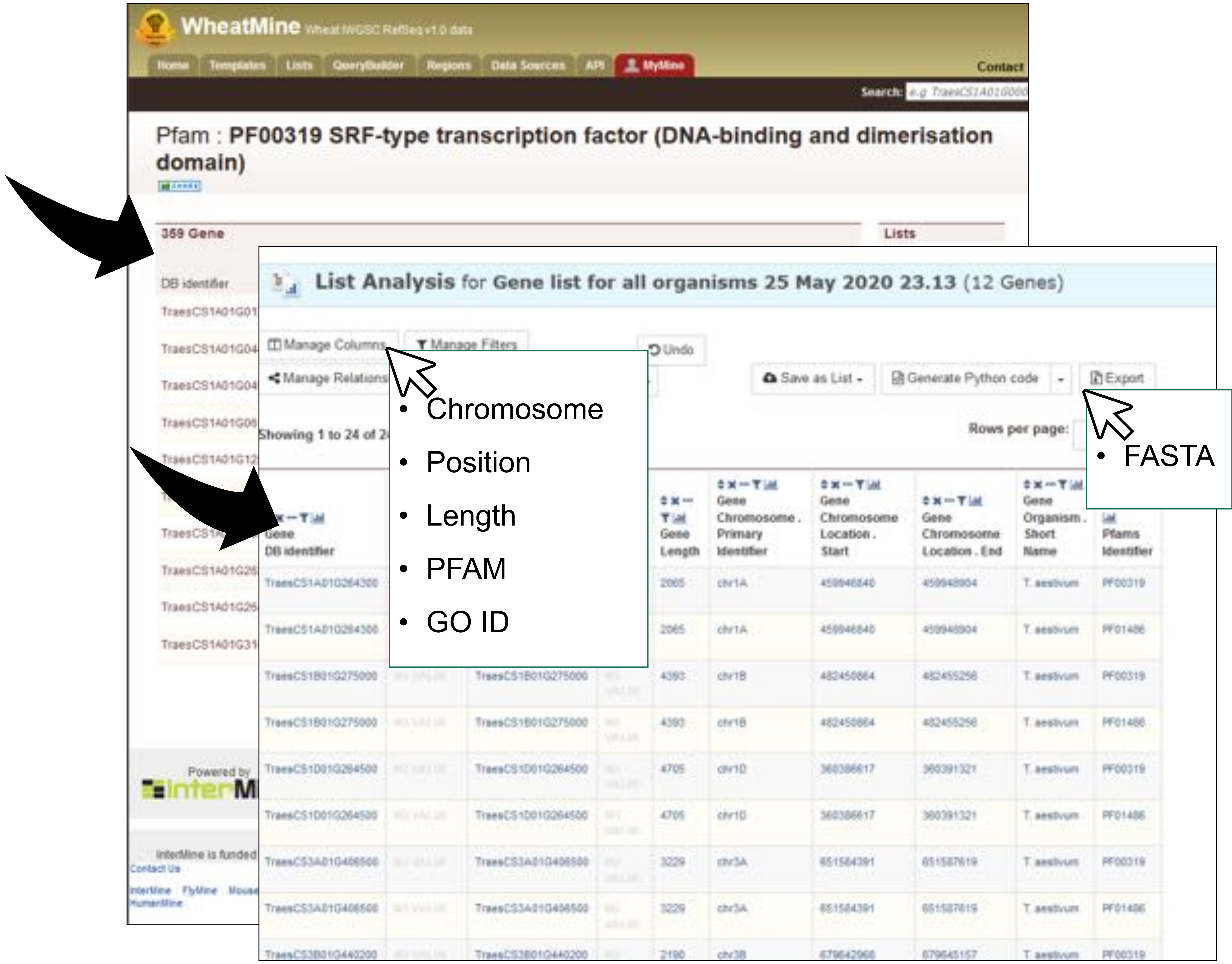
popular templates

Perl, Python, Ruby and Java API

Access our WheatMine data via our Application Programming Interface (API) too! We provide client libraries in the following languages:

Perl Python Ruby Java

https://urgi.versailles.inra.fr/WheatMine/begin.do



Pfam : **PF00319 SRF-type transcription factor (DNA-binding and dimerisation domain)**

369 Gene

Lists

DB identifier

TraesCS1A01G01

TraesCS1A01G04

TraesCS1A01G04

TraesCS1A01G06

TraesCS1A01G12

TraesCS1A

TraesCS1A01G26

TraesCS1A01G26

TraesCS1A01G31

List Analysis for **Gene list for all organisms 25 May 2020 23.13** (12 Genes)

Showing 1 to 24 of 2

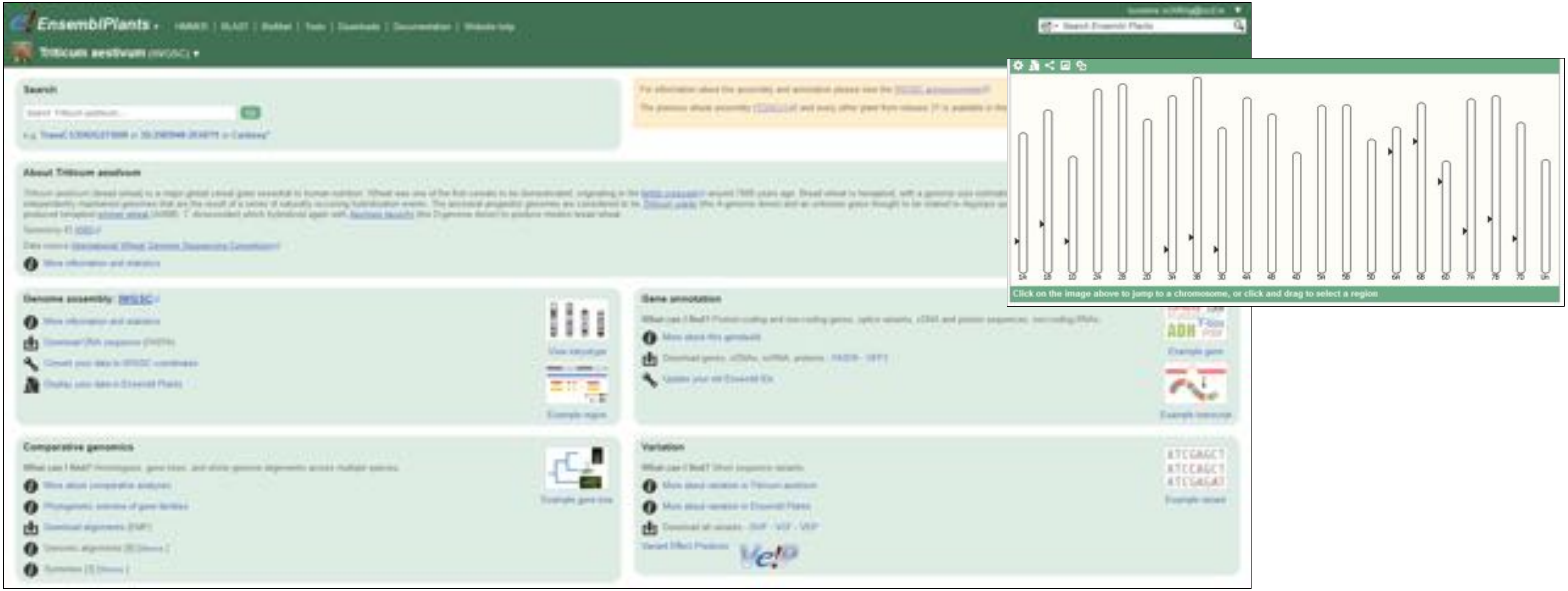
Rows per page:

Gene DB identifier				Gene Length	Gene Chromosome . Primary Identifier	Gene Chromosome Location . Start	Gene Chromosome Location . End	Gene Organism . Short Name	Pfams Identifier
TraesCS1A01G264300				2065	chr1A	459946840	459948904	T. aestivum	PF00319
TraesCS1A01G264300				2065	chr1A	459946840	459948904	T. aestivum	PF01486
TraesCS1B01G275000	NO VALUE	TraesCS1B01G275000	NO VALUE	4393	chr1B	482450864	482455256	T. aestivum	PF00319
TraesCS1B01G275000	NO VALUE	TraesCS1B01G275000	NO VALUE	4393	chr1B	482450864	482455256	T. aestivum	PF01486
TraesCS1D01G264500	NO VALUE	TraesCS1D01G264500	NO VALUE	4705	chr1D	360386617	360391321	T. aestivum	PF00319
TraesCS1D01G264500	NO VALUE	TraesCS1D01G264500	NO VALUE	4705	chr1D	360386617	360391321	T. aestivum	PF01486
TraesCS3A01G406500	NO VALUE	TraesCS3A01G406500	NO VALUE	3229	chr3A	651584391	651587619	T. aestivum	PF00319
TraesCS3A01G406500	NO VALUE	TraesCS3A01G406500	NO VALUE	3229	chr3A	651584391	651587619	T. aestivum	PF01486
TraesCS3B01G440200	NO VALUE	TraesCS3B01G440200	NO	2190	chr3B	679642968	679645157	T. aestivum	PF00319

- Chromosome
- Position
- Length
- PFAM
- GO ID

- FASTA

Ensembl Plants – Data Mining



EnsemblPlants ▾ HMMER | BLAST | BioMart | Tools | Downloads | Documentation | Website help

Search Ensembl Plants

Triticum aestivum (IWGSC) ▾

Search

Search Triticum aestivum... Go

e.g. [illegible]

About Triticum aestivum

[illegible]

[illegible]

Data source [illegible]

More information and statistics

Genome assembly: IWGSC

More information and statistics

Download DNA sequence (FASTA)

Convert your data to IWGSC coordinates

Display your data in Ensembl Plants

View karyotype

Example region

Comparative genomics

What can I find? Homologues, gene trees, and whole genome alignments across multiple species.

More about comparative analyses

Phylogenetic overview of gene families

Download alignments (EMF)

Genomic alignments (8) [Show]

Syntenies (3) [Show]

Example gene tree

Gene annotation

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

More about this genebuild

Download genes, cDNAs, ncRNA, proteins - FASTA - GFF3

Update your old Ensembl IDs

Example gene

Example transcript

Variation

What can I find? Short sequence variants.

More about variation in Triticum aestivum

More about variation in Ensembl Plants

Download all variants - GVF - VCF - VEP

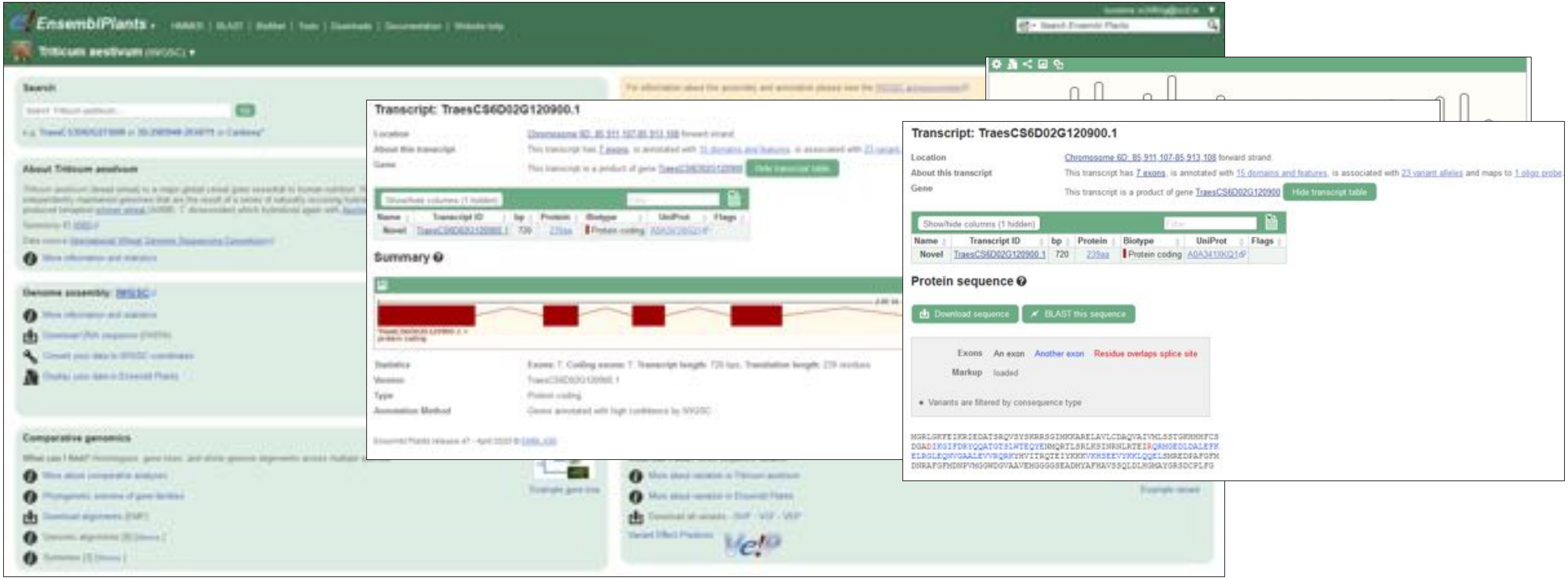
Variant Effect Predictor

Example variant

Click on the image above to jump to a chromosome, or click and drag to select a region

plants.ensembl.org/Triticum_aestivum/

Ensembl Plants – Data Mining



Transcript: TraesCS6D02G120900.1

Summary

Statistics	Exons: 7, Coding exons: 7, Transcript length: 720 bps, Translation length: 239 residues
Version	TraesCS6D02G120900.1
Type	Protein coding
Annotation Method	Genes annotated with high confidence by IWGSC

Transcript: TraesCS6D02G120900.1

Location	Chromosome 6D: 85,911,107-85,913,108 forward strand.
About this transcript	This transcript has 7 exons, is annotated with 15 domains and features, is associated with 23 variant alleles and maps to 1 oligo probe.
Gene	This transcript is a product of gene TraesCS6D02G120900 Hide transcript table

Show/hide columns (1 hidden)

Name	Transcript ID	bp	Protein	Biotype	UniProt	Flags
Novel	TraesCS6D02G120900.1	720	239aa	Protein coding	A0A341XKQ1	

Protein sequence

Download sequence   BLAST this sequence

Exons   An exon   Another exon   Residue overlaps splice site

Markup   loaded

- Variants are filtered by consequence type

```
MGRLGKFEIKRIEDATSRQVSYSKRRSGIMKKARELAVLCDAQVAIVMLSSTGKMMNFCS
DGADIKGIFDRYQQATGTSLWTEQYENMQRTLSRLKSINRNLRTEIRQRMGEDLDALEFK
ELRGLEQNVGAALEVVRQRKYHVITRQTEIYKKKVKHSEEVYKKLQQELSMREDFAFGFM
DNRAFGFHDNPVMGGWDGVAAVENGGGGSEADMYAFHAVSSQLDLHGMAYGRSDCPLFG
```

plants.ensembl.org/Triticum_aestivum/

Rainer Melzer & Susanne Schilling, UCD, Ireland

Ensembl Plants – Data Mining

Transcript: TraesCS6D02G120900.1

Location Chromosome 6D: 85,911,107-85,913,108 forward strand.

About this transcript This transcript has 7 exons, is annotated with 15 domains and features, is associated with 23 variant alleles and maps to 1 oligo probe.

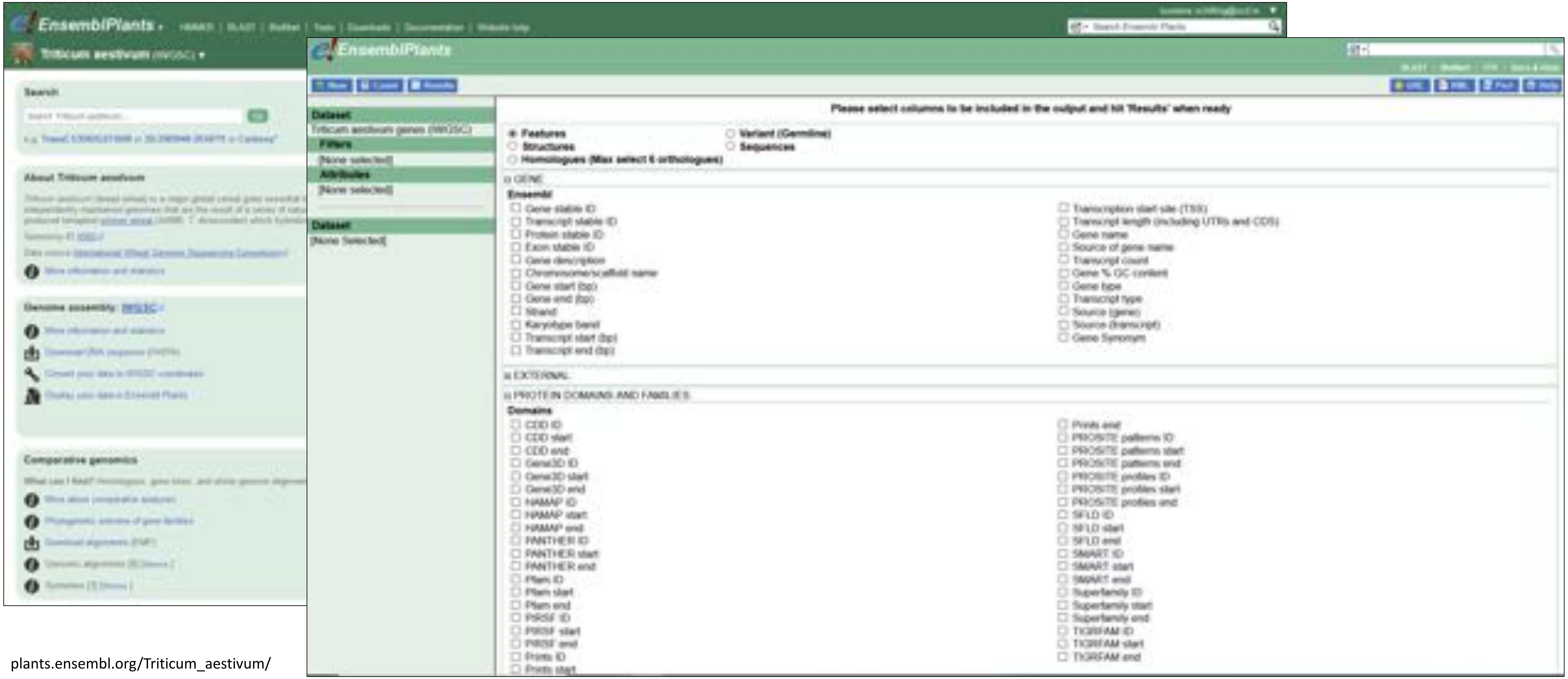
Gene This transcript is a product of gene TraesCS6D02G120900 Hide transcript table

- FASTA

EnsemblPlants
Triticum aestivum
Search
About Triticum aestivum
Genome assembly
Comparative genomics
LEGEND

plants.ensembl.org

Ensembl Plants – Data Mining



Dataset
Triticum aestivum genes (IWGSC)
Filters
[None selected]
Attributes
[None selected]

Dataset
[None Selected]

Please select columns to be included in the output and hit 'Results' when ready

● **Features**
○ **Structures**
○ **Homologues (Max select 6 orthologues)**
○ **Variant (Germline)**
○ **Sequences**

GENE

Ensembl

□ Gene stable ID
□ Transcript stable ID
□ Protein stable ID
□ Exon stable ID
□ Gene description
□ Chromosome/scaffold name
□ Gene start (bp)
□ Gene end (bp)
□ Strand
□ Karyotype band
□ Transcript start (bp)
□ Transcript end (bp)
□ Transcription start site (TSS)
□ Transcript length (including UTRs and CDS)
□ Gene name
□ Source of gene name
□ Transcript count
□ Gene % GC content
□ Gene type
□ Transcript type
□ Source (gene)
□ Source (transcript)
□ Gene Synonym

EXTERNAL

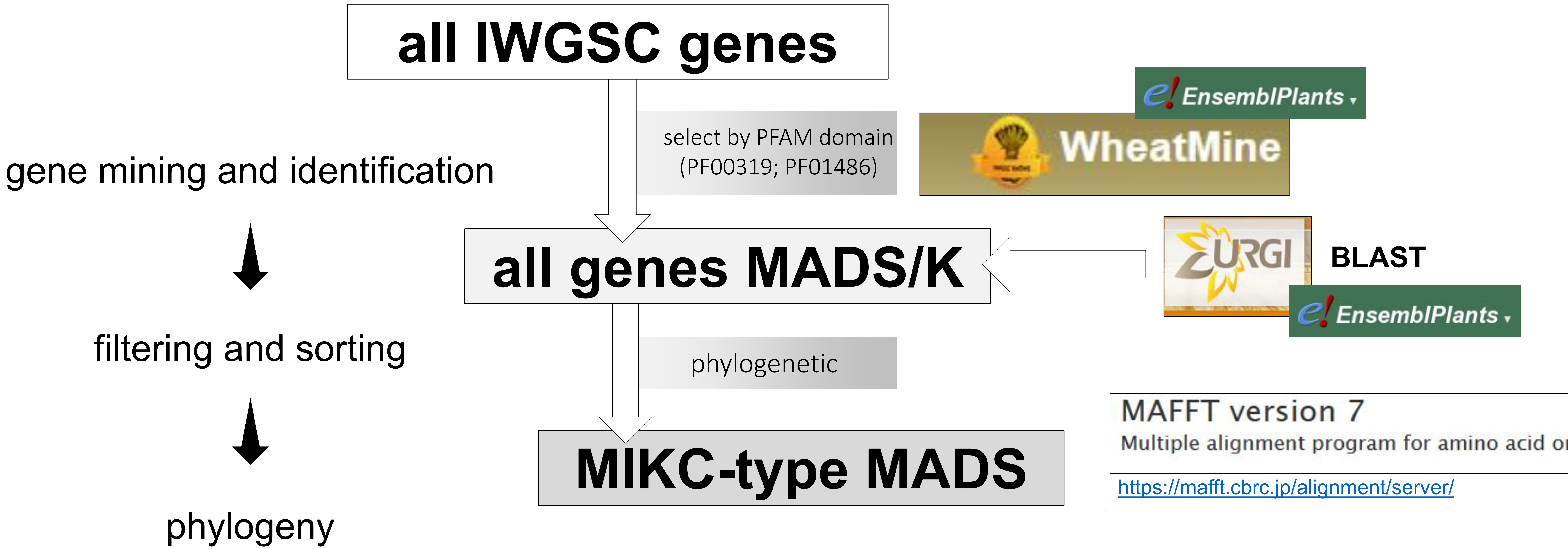
PROTEIN DOMAINS AND FAMILIES

Domains

□ CDD ID
□ CDD start
□ CDD end
□ Gene3D ID
□ Gene3D start
□ Gene3D end
□ HAMAP ID
□ HAMAP start
□ HAMAP end
□ PANTHER ID
□ PANTHER start
□ PANTHER end
□ Pfam ID
□ Pfam start
□ Pfam end
□ PIRSF ID
□ PIRSF start
□ PIRSF end
□ Prints ID
□ Prints start
□ Prints end
□ PROSITE patterns ID
□ PROSITE patterns start
□ PROSITE patterns end
□ PROSITE profiles ID
□ PROSITE profiles start
□ PROSITE profiles end
□ SFLD ID
□ SFLD start
□ SFLD end
□ SMART ID
□ SMART start
□ SMART end
□ Superfamily ID
□ Superfamily start
□ Superfamily end
□ TIGRFAM ID
□ TIGRFAM start
□ TIGRFAM end

plants.ensembl.org/Triticum_aestivum/

From Genes to Phylogenies



gene mining and identification

filtering and sorting

phylogeny

all IWGSC genes

select by PFAM domain (PF00319; PF01486)

all genes MADS/K

BLAST

phylogenetic

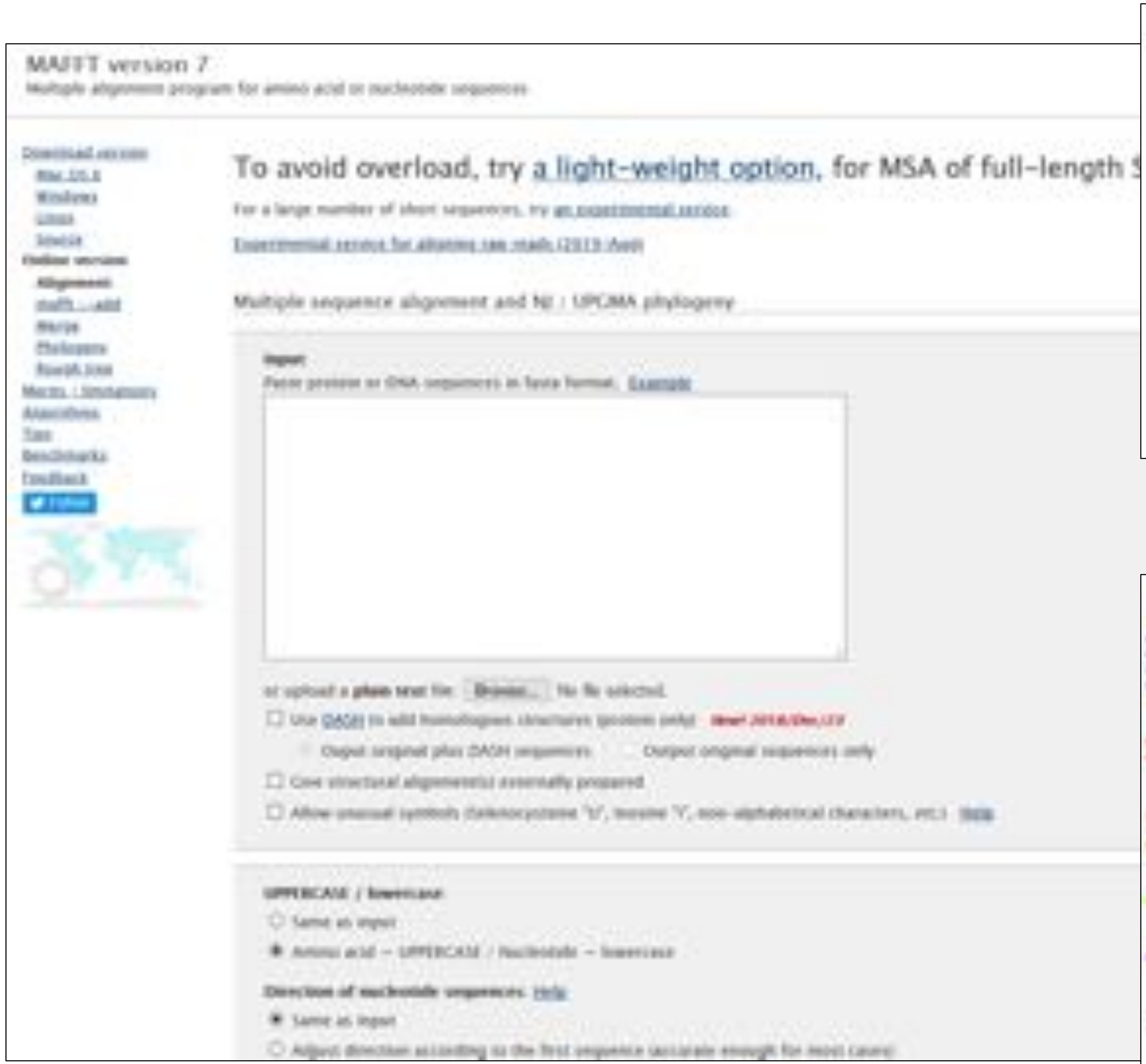
MIKC-type MADS

MAFFT version 7

Multiple alignment program for amino acid or

https://mafft.cbrc.jp/alignment/server/

MAFFT for building alignments



MAFFT version 7

Multiple alignment program for amino acid or nucleotide sequences

To avoid overload, try a light-weight option, for MSA of full-length S

For a large number of short sequences, try an experimental service.

Experimental service for aligning raw reads (2019 Apr)

Multiple sequence alignment and NJ / UPGMA phylogeny

Input:

Paste protein or DNA sequences in fasta format. Example

or upload a plain text file: Browse... No file selected.

□ Use DASH to add homologous structures (protein only) New! 2018/Dec/23

○ Output original plus DASH sequences ○ Output original sequences only

□ Give structural alignment(s) externally prepared

□ Allow unusual symbols (Selenocysteine 'U', Inosine 'i', non-alphabetical characters, etc.) Help

UPPERCASE / lowercase

○ Same as input

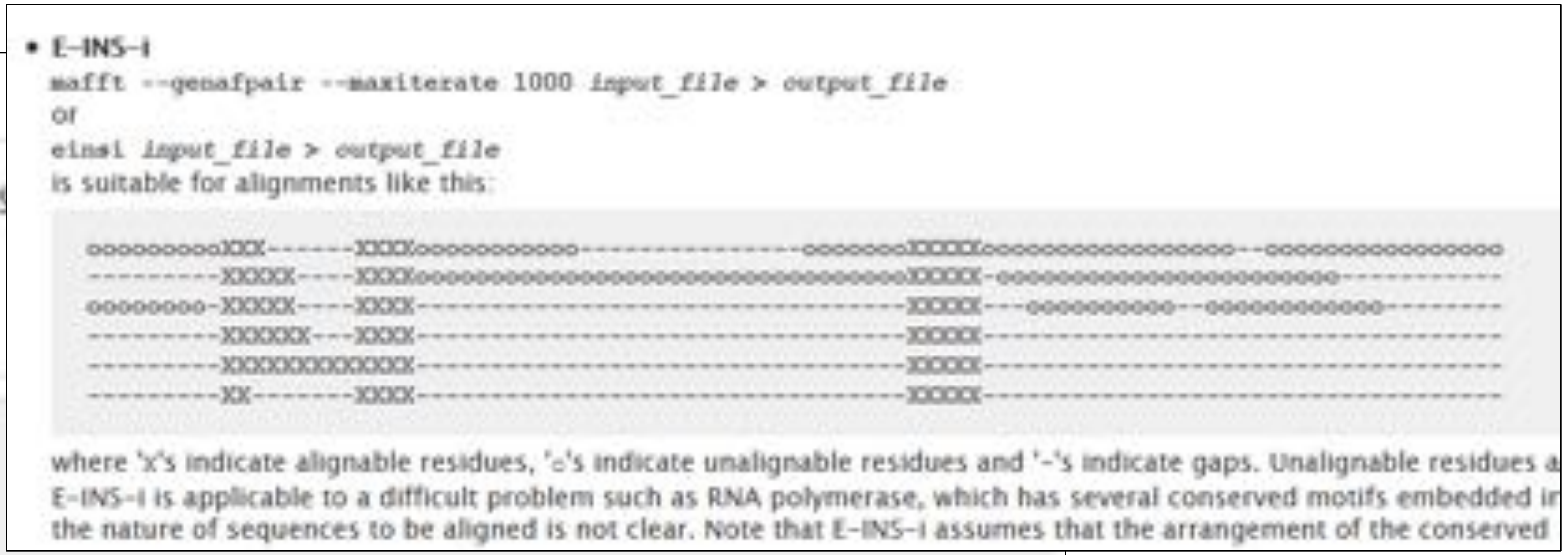
● Amino acid → UPPERCASE / Nucleotide → lowercase

Direction of nucleotide sequences: Help

● Same as input

○ Adjust direction according to the first sequence (accurate enough for most cases)

https://mafft.cbrc.jp/alignment/server/



- E-INS-i

```
mafft --genafpair --maxiterate 1000 input_file > output_file
```

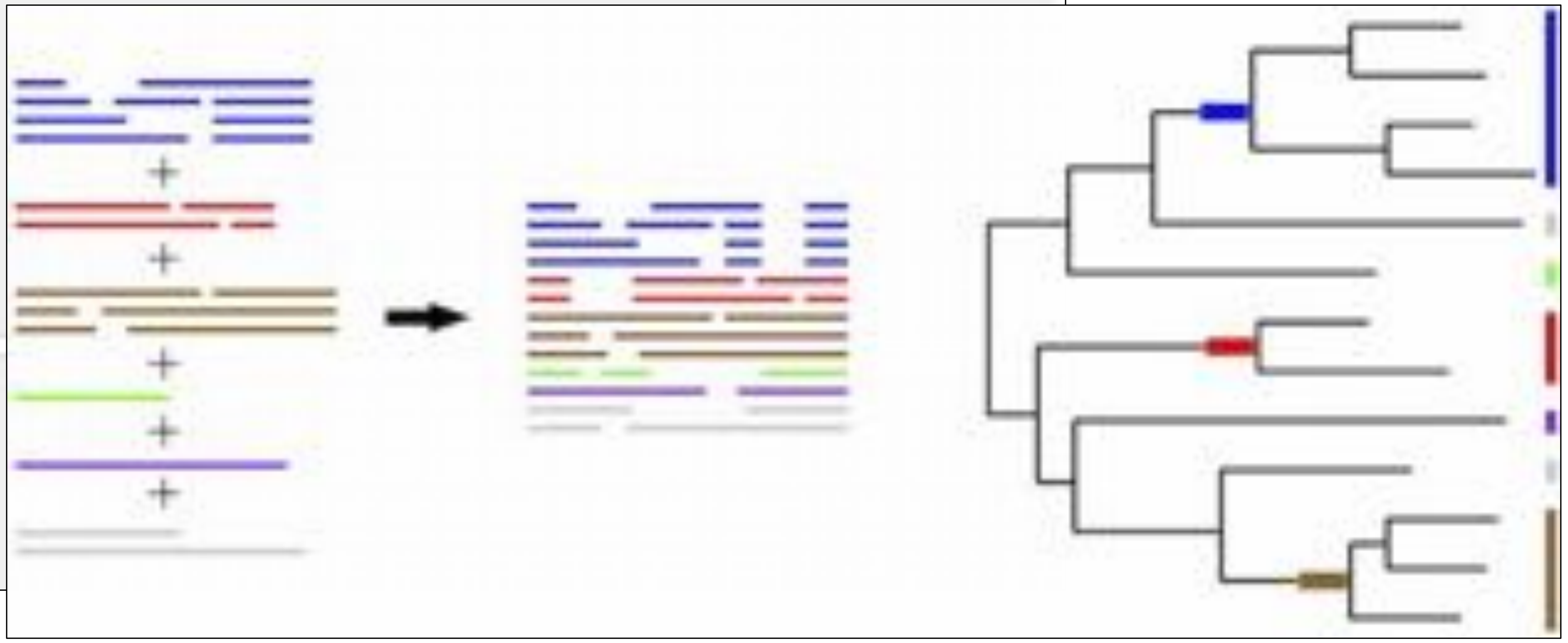
or

```
einsi input_file > output_file
```

is suitable for alignments like this:

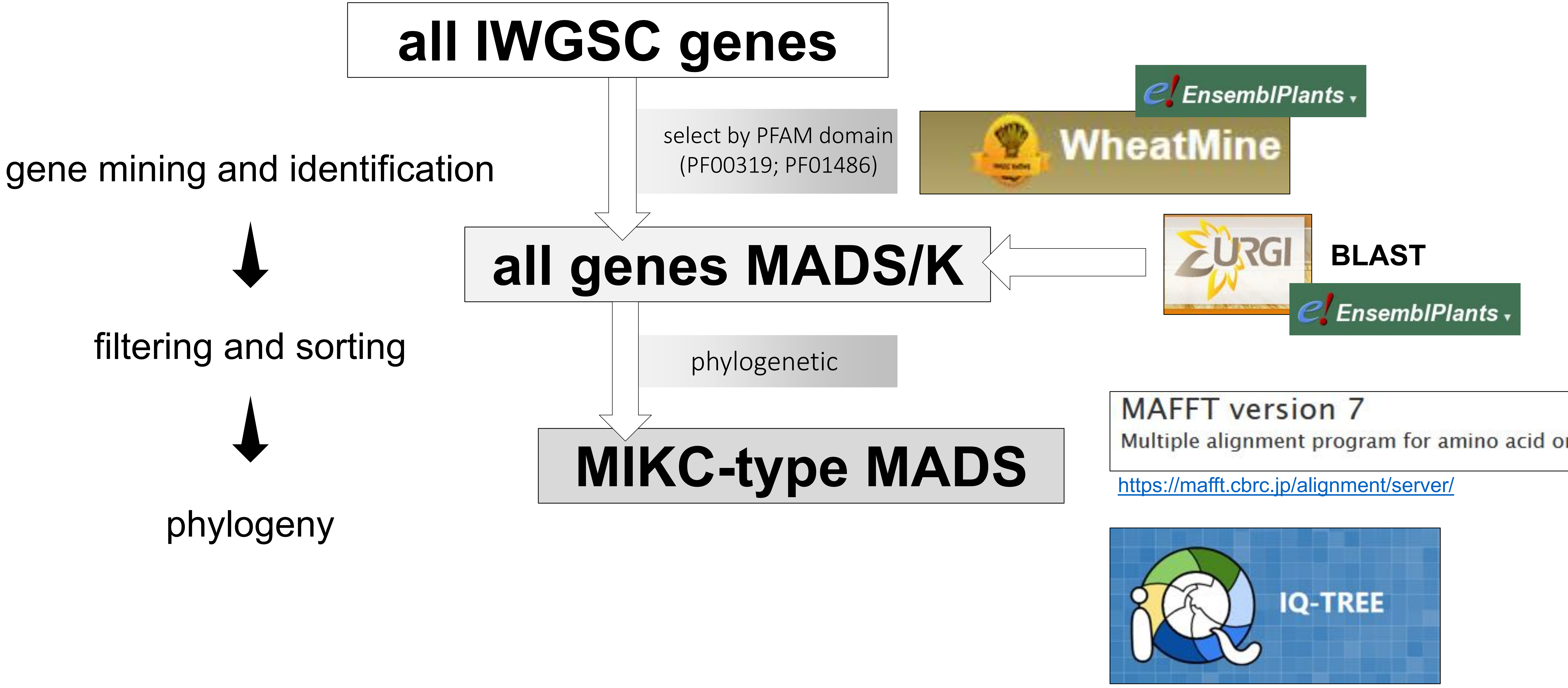
```
oooooooooXXX------XXXXooooooooooo---------------oooooooXXXXXoooooooooooooooo--ooooooooooooooo
---------XXXXX----XXXXoooooooooooooooooooooooooooooooXXXXX-oooooooooooooooooooooooo----------
oooooooo-XXXXX----XXXX------------------------------XXXXX---ooooooooooo--ooooooooooooo--------
---------XXXXXX---XXXX------------------------------XXXXX------------------------------------
---------XXXXXXXXXXXXX------------------------------XXXXX------------------------------------
---------XX-------XXXX------------------------------XXXXX------------------------------------
```

where 'x's indicate alignable residues, 'o's indicate unalignable residues and '-'s indicate gaps. Unalignable residues a
E-INS-i is applicable to a difficult problem such as RNA polymerase, which has several conserved motifs embedded ir
the nature of sequences to be aligned is not clear. Note that E-INS-i assumes that the arrangement of the conserved



Rainer Melzer & Susanne Schilling, UCD, Ireland

From Genes to Phylogenies



gene mining and identification

filtering and sorting

phylogeny

all IWGSC genes

select by PFAM domain (PF00319; PF01486)

all genes MADS/K

phylogenetic

MIKC-type MADS

MAFFT version 7

Multiple alignment program for amino acid o

https://mafft.cbrc.jp/alignment/server/

http://www.iqtree.org/

From Genes to Phylogenies

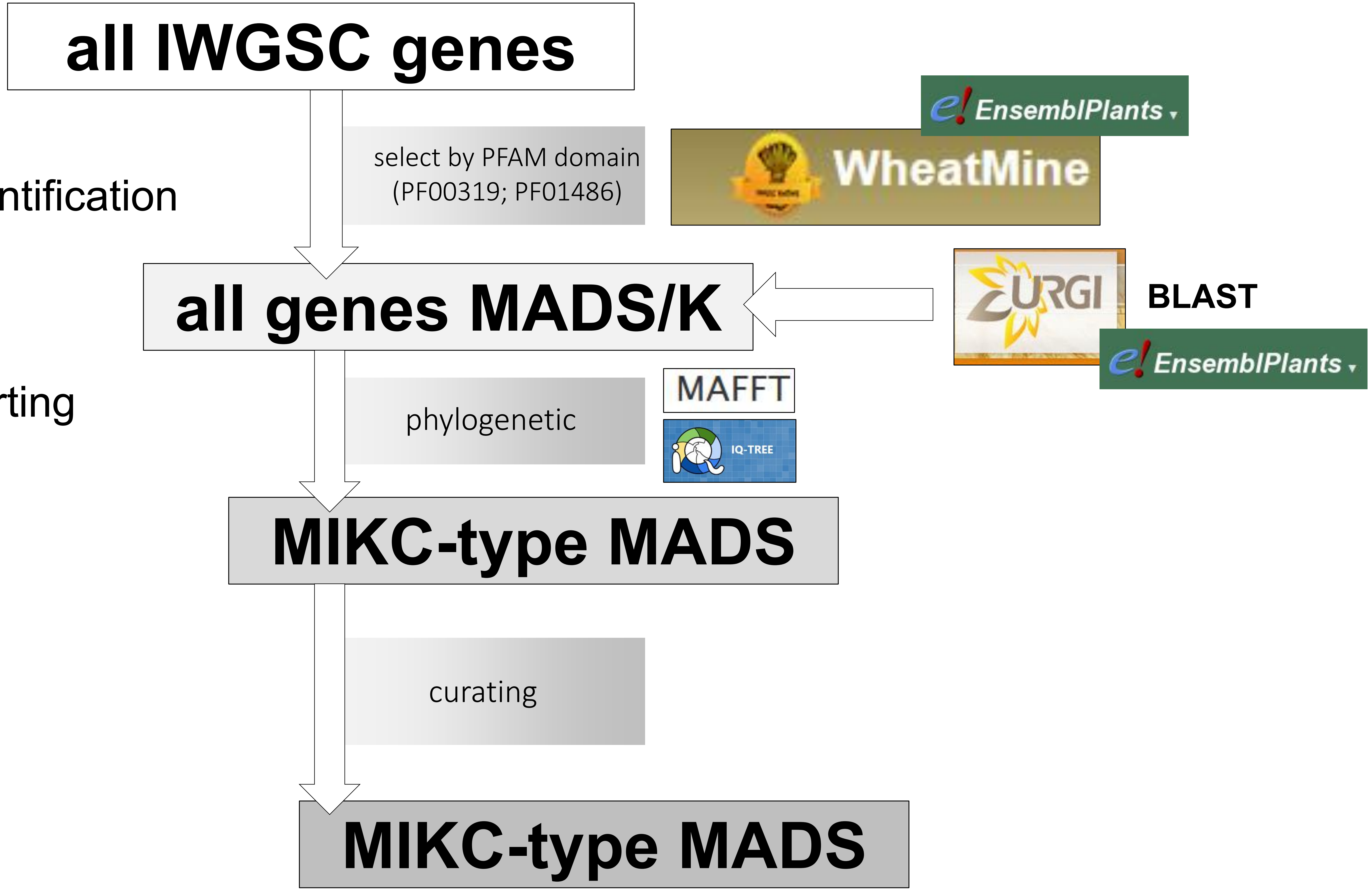
gene mining and identification

↓

filtering and sorting

↓

phylogeny



all IWGSC genes

select by PFAM domain (PF00319; PF01486)

all genes MADS/K

BLAST

phylogenetic

MIKC-type MADS

curating

MIKC-type MADS

Rainer Melzer & Susanne Schilling, UCD, Ireland

From Genes to Phylogenies

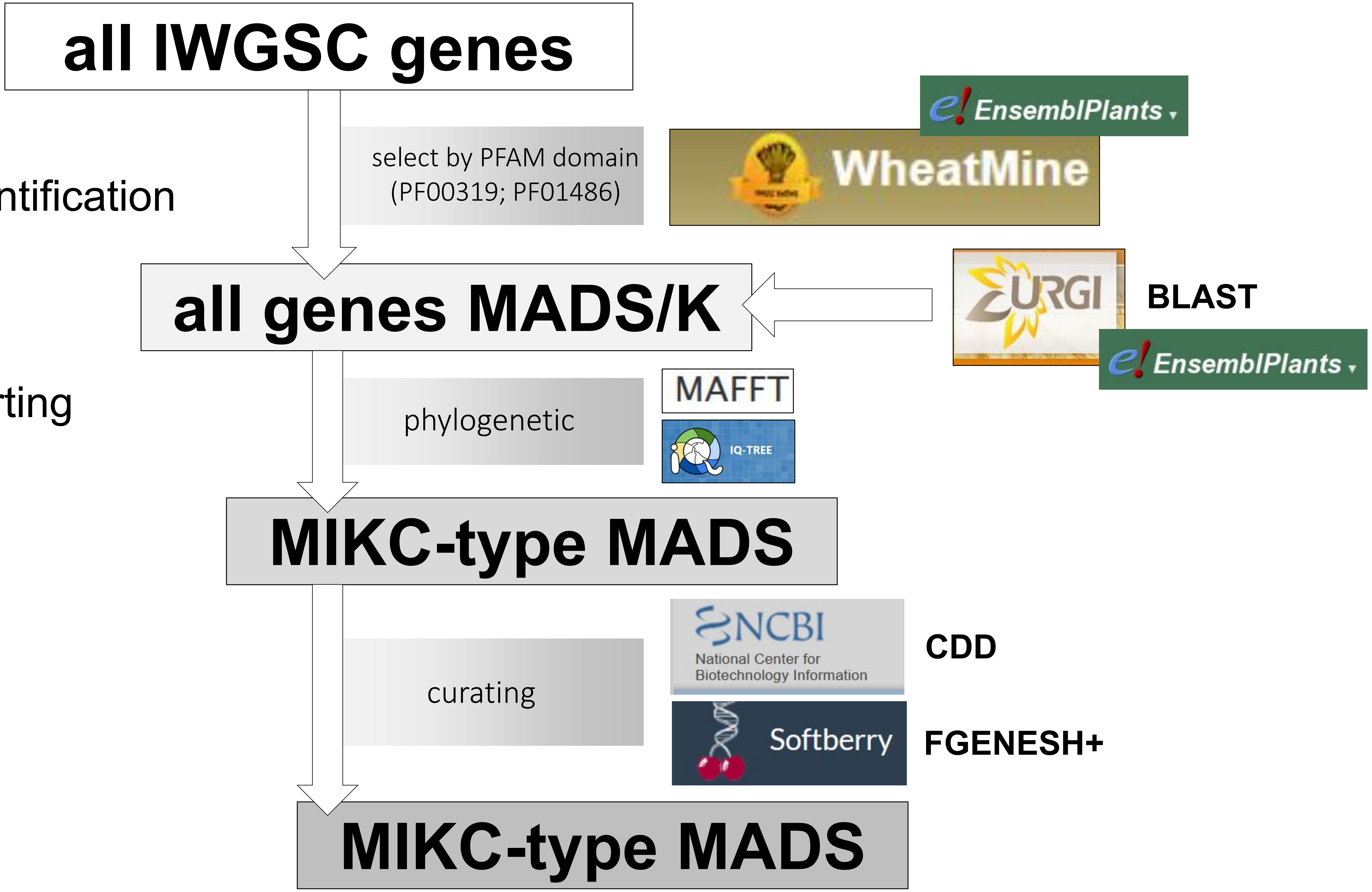
gene mining and identification

↓

filtering and sorting

↓

phylogeny



all IWGSC genes

select by PFAM domain (PF00319; PF01486)

WheatMine

EnsemblPlants

all genes MADS/K

URGI

BLAST

EnsemblPlants

phylogenetic

MAFFT

IQ-TREE

MIKC-type MADS

curating

NCBI National Center for Biotechnology Information

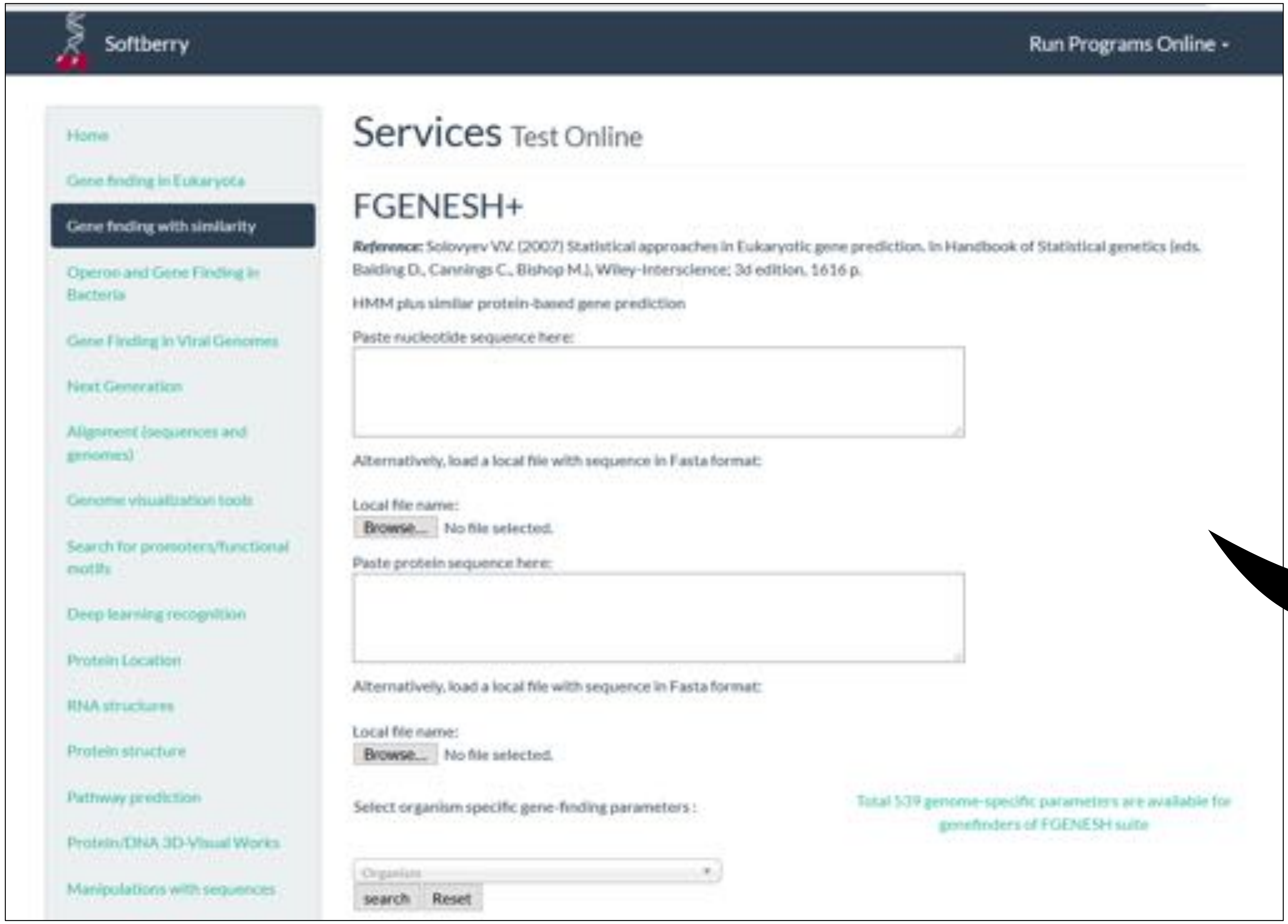
CDD

Softberry

FGENESH+

MIKC-type MADS

From Genes to Phylogenies


Softberry
Run Programs Online ▾
Home
Gene finding in Eukaryota
Gene finding with similarity
Operon and Gene Finding in Bacteria
Gene Finding in Viral Genomes
Next Generation
Alignment (sequences and genomes)
Genome visualization tools
Search for promoters/functional motifs
Deep learning recognition
Protein Location
RNA structures
Protein structure
Pathway prediction
Protein/DNA 3D-Visual Works
Manipulations with sequences

Services Test Online

FGENESH+

Reference: Solovyev V.V. (2007) Statistical approaches in Eukaryotic gene prediction. In Handbook of Statistical genetics (eds. Balding D., Cannings C., Bishop M.), Wiley-Interscience; 3d edition, 1616 p.

HMM plus similar protein-based gene prediction

Paste nucleotide sequence here:

Alternatively, load a local file with sequence in Fasta format:

Local file name:
Browse... No file selected.

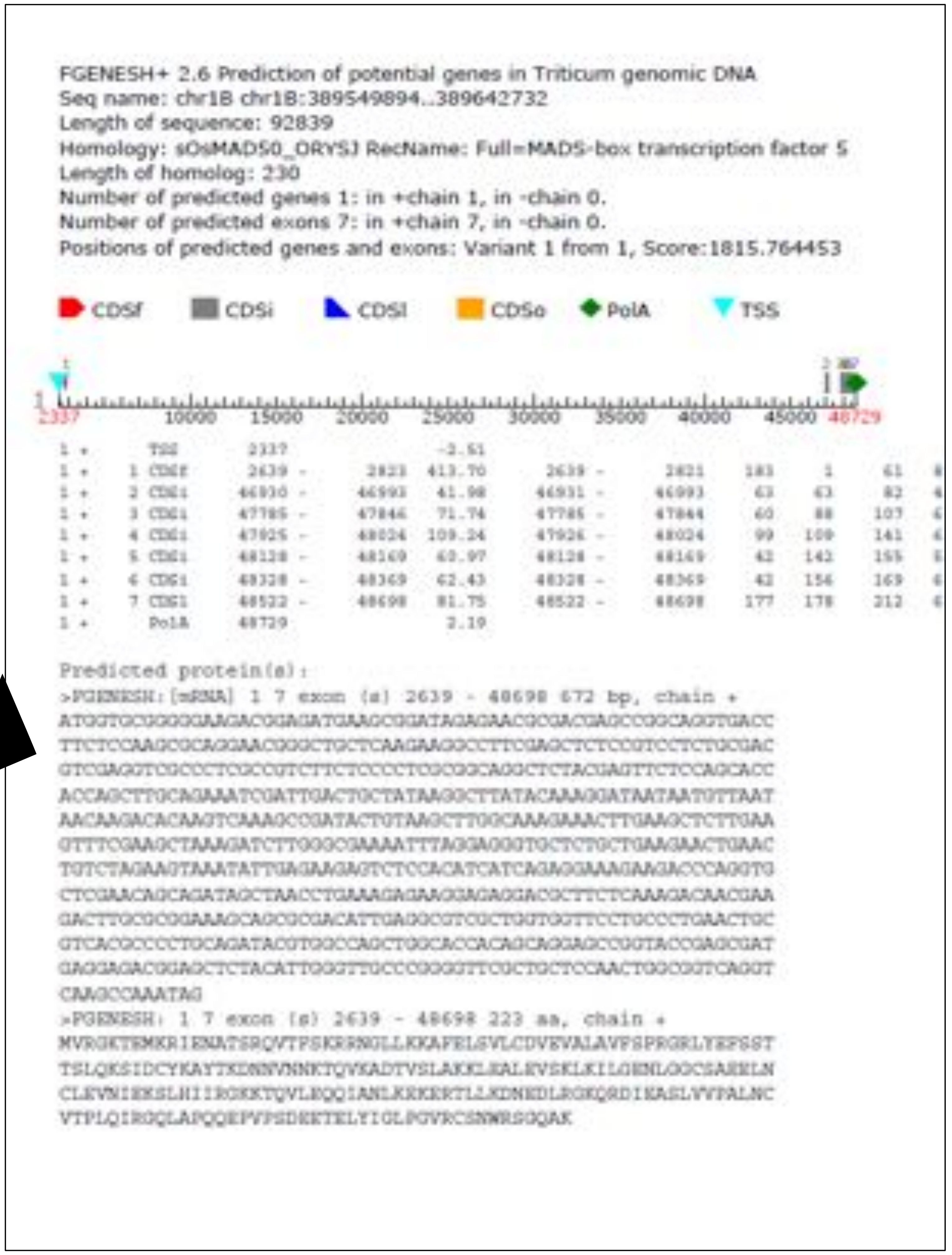
Paste protein sequence here:

Alternatively, load a local file with sequence in Fasta format:

Local file name:
Browse... No file selected.

Select organism specific gene-finding parameters :

Total 539 genome-specific parameters are available for genefinders of FGENESH suite



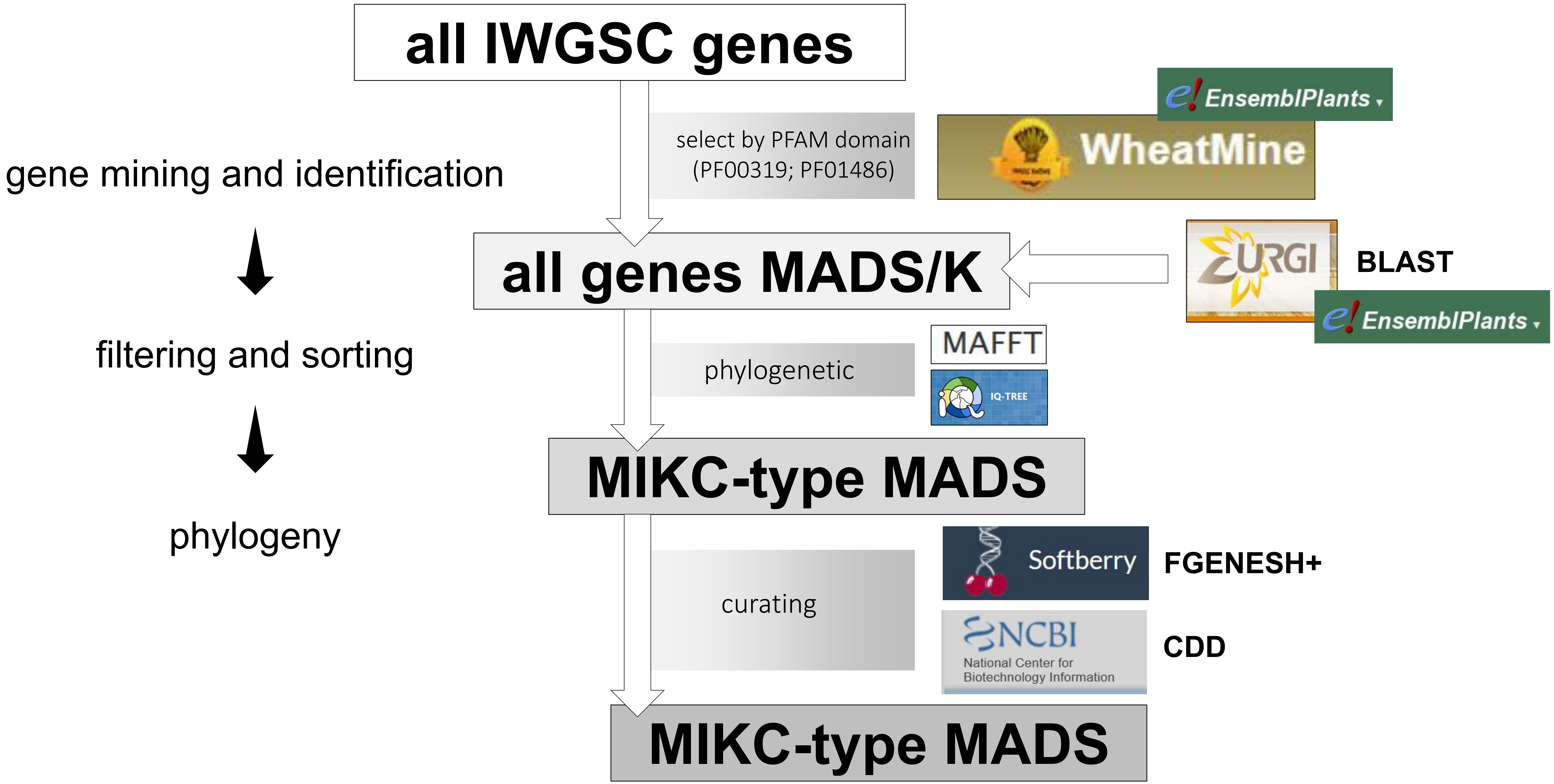
FGENESH+ 2.6 Prediction of potential genes in Triticum genomic DNA
Seq name: chr1B chr1B:389549894..389642732
Length of sequence: 92839
Homology: sOsMADS0_ORYSJ RecName: Full=MADS-box transcription factor 5
Length of homolog: 230
Number of predicted genes 1: in +chain 1, in -chain 0.
Number of predicted exons 7: in +chain 7, in -chain 0.
Positions of predicted genes and exons: Variant 1 from 1, Score:1815.764453

```
1 +      TSS     2337              -2.51
1 +    1 CDSf    2639 -     2823  413.70     2639 -     2821   183     1    61   6
1 +    2 CDSi   46930 -    46993   41.98    46931 -    46993    63    63    82   4
1 +    3 CDSi   47785 -    47846   71.74    47785 -    47844    60    88   107   6
1 +    4 CDSi   47925 -    48024  109.24    47926 -    48024    99   109   141   6
1 +    5 CDSi   48128 -    48169   60.97    48128 -    48169    42   142   155   5
1 +    6 CDSi   48328 -    48369   62.43    48328 -    48369    42   156   169   6
1 +    7 CDSl   48522 -    48698   81.75    48522 -    48698   177   178   212   6
1 +      PolA   48729               2.19

Predicted protein(s):
>FGENESH:[mRNA] 1 7 exon (s) 2639 - 48698 672 bp, chain +
ATGGTGCGGGGGAAGACGGAGATGAAGCGGATAGAGAACGCGACGAGCCGGCAGGTGACC
TTCTCCAAGCGCAGGAACGGGCTGCTCAAGAAGGCCTTCGAGCTCTCCGTCCTCTGCGAC
GTCGAGGTCGCCCTCGCCGTCTTCTCCCCTCGCGGCAGGCTCTACGAGTTCTCCAGCACC
ACCAGCTTGCAGAAATCGATTGACTGCTATAAGGCTTATACAAAGGATAATAATGTTAAT
AACAAGACACAAGTCAAAGCCGATACTGTAAGCTTGGCAAAGAAACTTGAAGCTCTTGAA
GTTTCGAAGCTAAAGATCTTGGGCGAAAATTTAGGAGGGTGCTCTGCTGAAGAACTGAAC
TGTCTAGAAGTAAATATTGAGAAGAGTCTCCACATCATCAGAGGAAAGAAGACCCAGGTG
CTCGAACAGCAGATAGCTAACCTGAAAGAGAAGGAGAGGACGCTTCTCAAAGACAACGAA
GACTTGCGCGGAAAGCAGCGCGACATTGAGGCGTCGCTGGTGGTTCCTGCCCTGAACTGC
GTCACGCCCCTGCAGATACGTGGCCAGCTGGCACCACAGCAGGAGCCGGTACCGAGCGAT
GAGGAGACGGAGCTCTACATTGGGTTGCCCGGGGTTCGCTGCTCCAACTGGCGGTCAGGT
CAAGCCAAATAG
>FGENESH: 1 7 exon (s) 2639 - 48698 223 aa, chain +
MVRGKTEMKRIENATSRQVTFSKRRNGLLKKAFELSVLCDVEVALAVFSPRGRLYEFSST
TSLQKSIDCYKAYTKDNNVNNKTQVKADTVSLAKKLEALEVSKLKILGENLGGCSAEELN
CLEVNIEKSLHIIRGKKTQVLEQQIANLKEKERTLLKDNEDLRGKQRDIKASLVVPALNC
VTPLQIRGQLAPQQEPVPSDEETELYIGLPGVRCSNWRSGQAK
```

Rainer Melzer & Susanne Schilling, UCD, Ireland

From Genes to Phylogenies



gene mining and identification

filtering and sorting

phylogeny

all IWGSC genes

select by PFAM domain (PF00319; PF01486)

all genes MADS/K

BLAST

phylogenetic

MIKC-type MADS

curating

FGENESH+

CDD

MIKC-type MADS

Phylogeny tools

MAFFT version 7

Multiple alignment program for amino acid or nucleotide sequences

https://mafft.cbrc.jp/alignment/server/

IQ-TREE

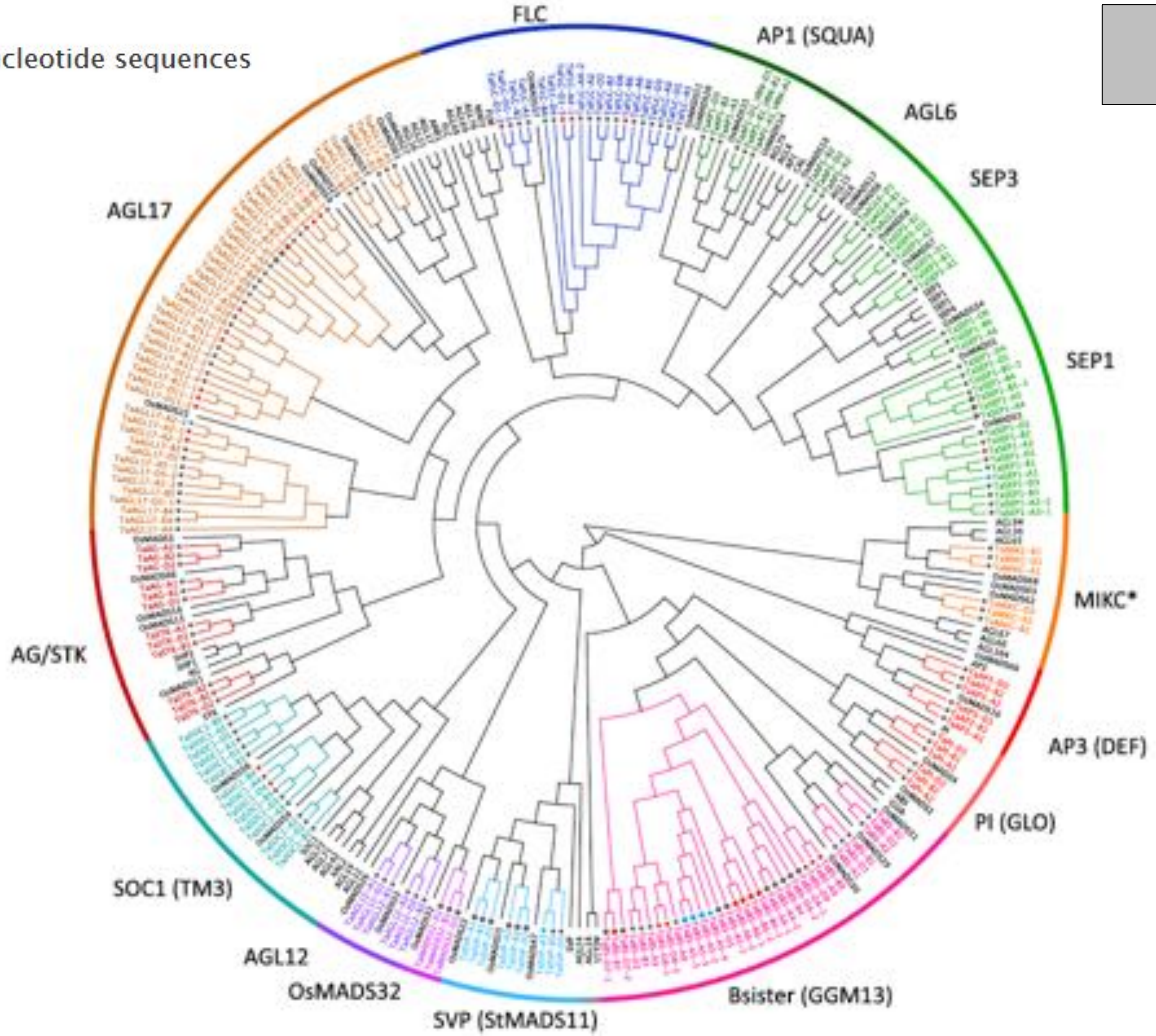
http://www.iqtree.org/

geneious

https://www.geneious.com/

FigTree v1.4.4

https://github.com/rambaut/figtree/releases

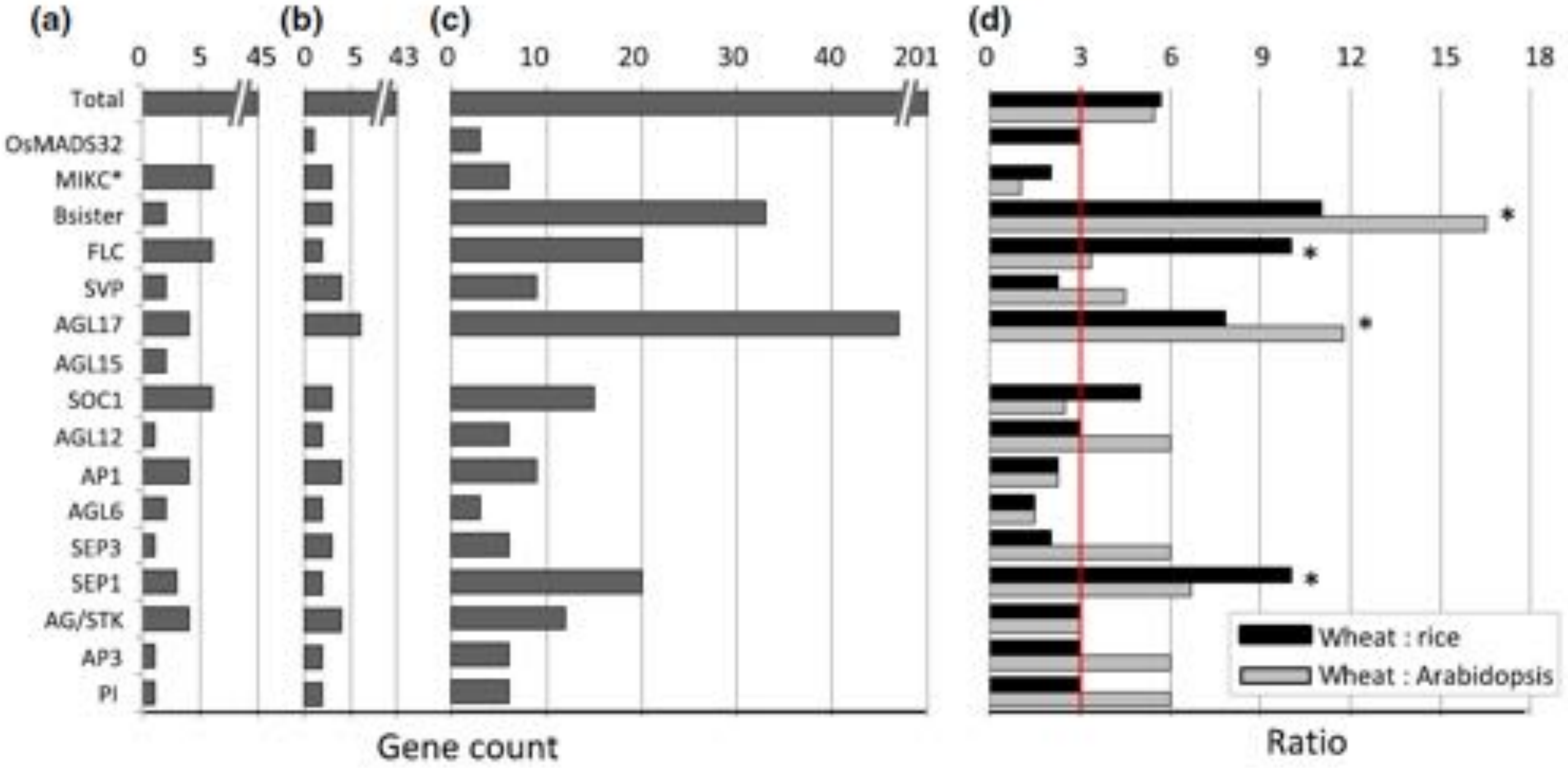


MIKC-type MADS

wheat	201
rice	43
Arabidopsis	45

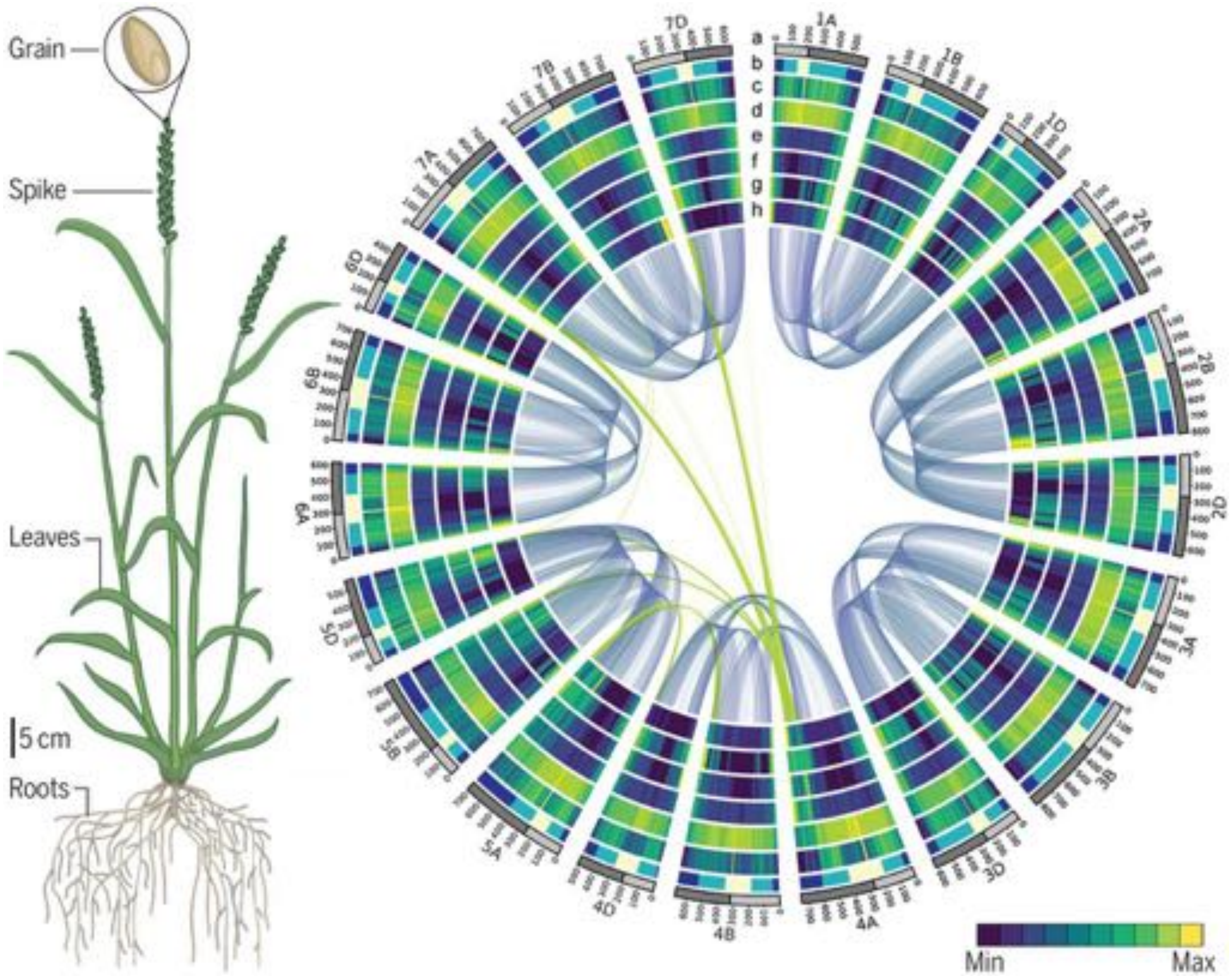
Schilling et al. 2020

Counting, comparing and stats



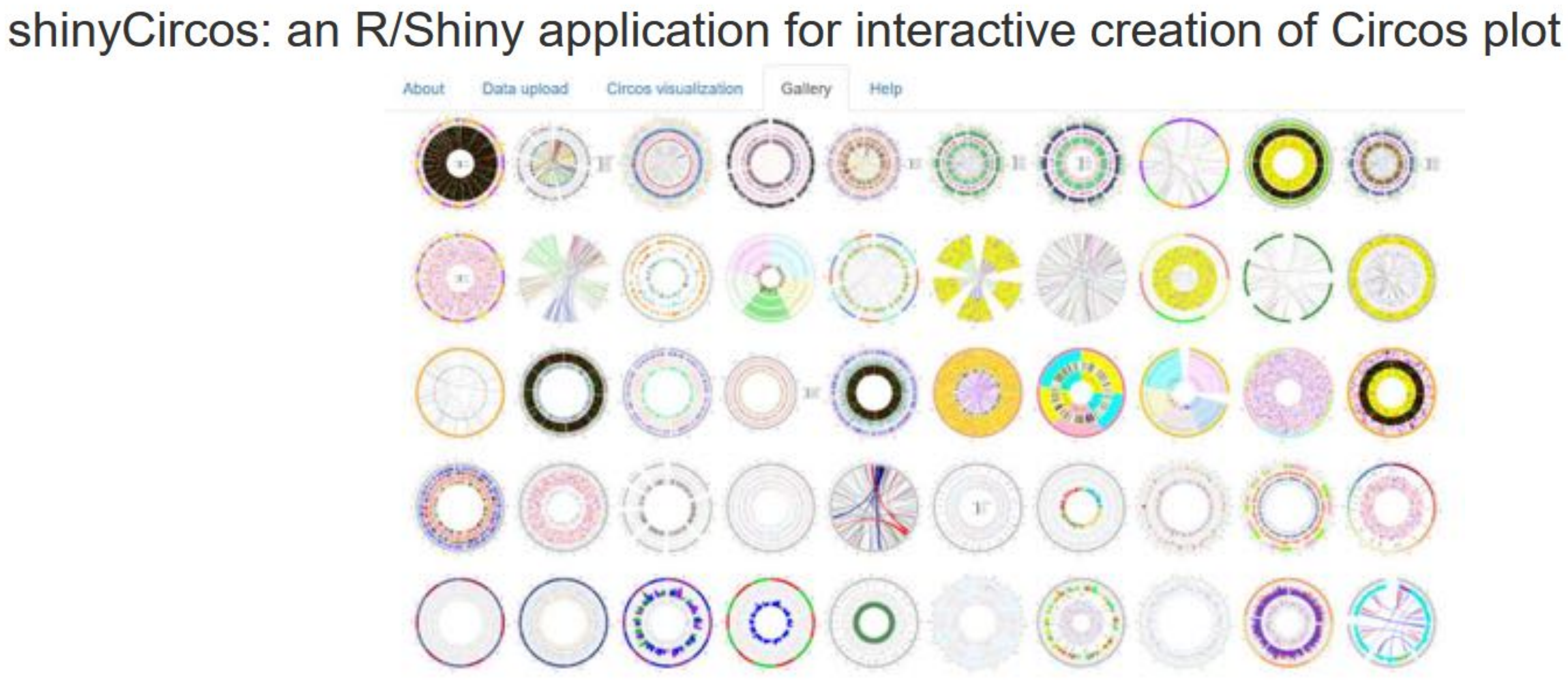
Schilling et al. 2020

Circos plots to visualize whole genomes



IWGSC 2018, Science

Circos plot with Shiny Circos

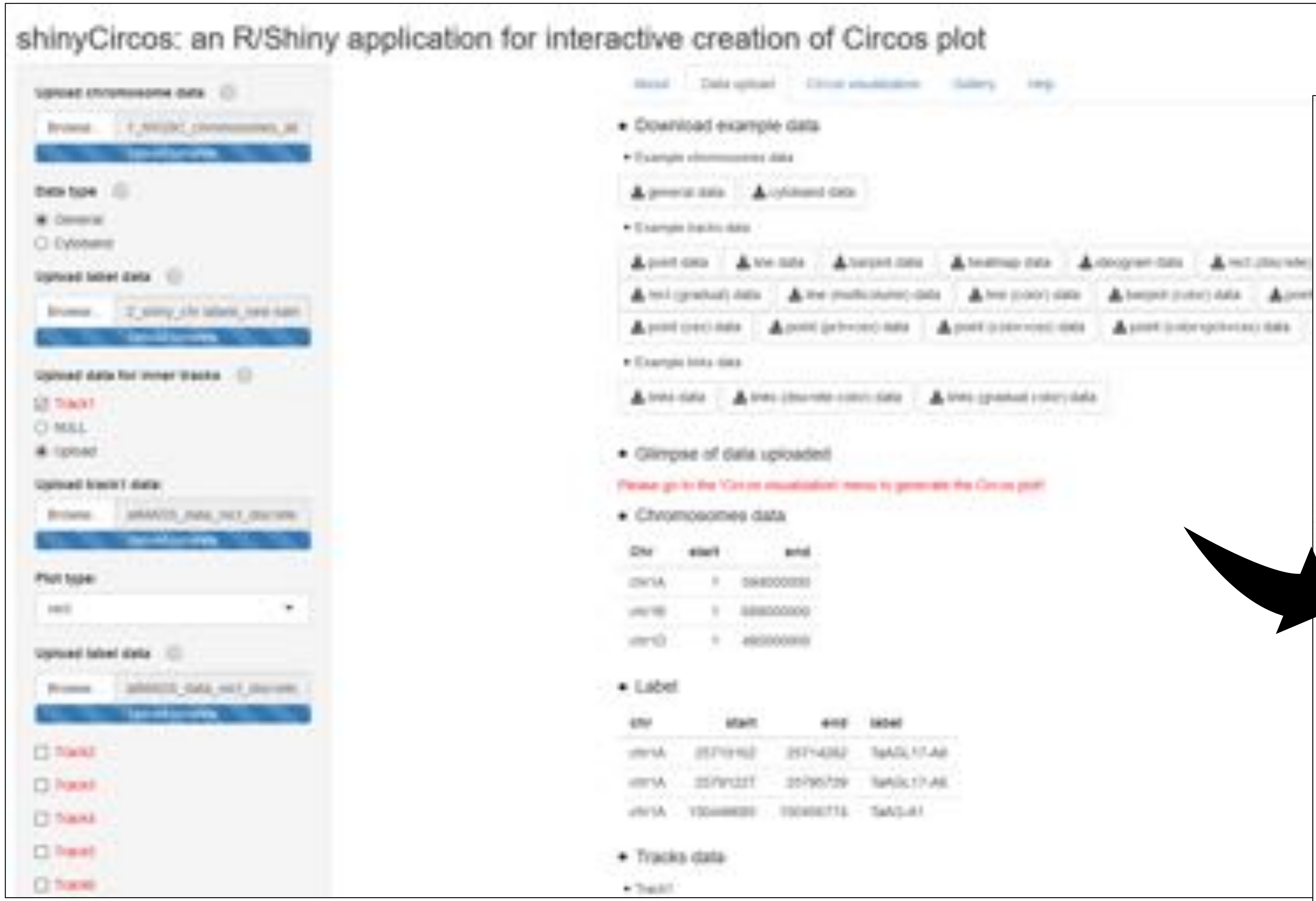


shinyCircos: an R/Shiny application for interactive creation of Circos plot

https://github.com/venyao/shinyCircos

http://150.109.59.144:3838/shinyCircos/

Circos plot with Shiny Circos



shinyCircos: an R/Shiny application for interactive creation of Circos plot

Upload chromosome data

Browse... 1_wheat_chromosomes_all

Data type

● General

○ Cytoband

Upload label data

Browse... 2_shiny_chr labels_rare.txt

Upload data for inner tracks

☑ Track1

○ NULL

● Upload

Upload track1 data

Browse... wheat_data_rect_discrete

Plot type:

rect

Upload label data

Browse... wheat_data_rect_discrete

□ Track2

□ Track3

□ Track4

□ Track5

□ Track6

- Download example data
  - Example chromosomes data

  - Example tracks data

  - Example links data

- Glimpse of data uploaded

Please go to the 'Circos visualization' menu to generate the Circos plot!

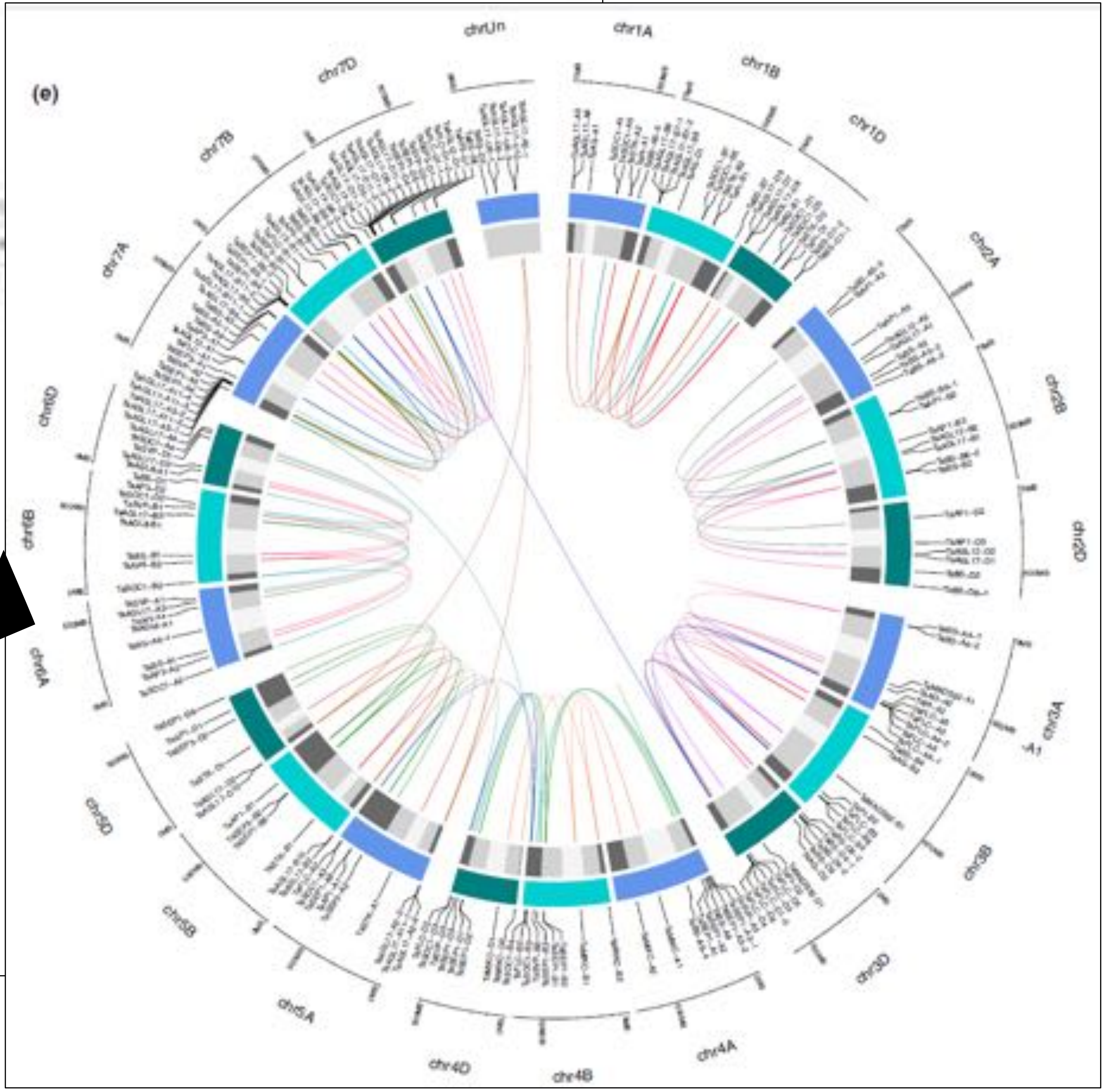
- Chromosomes data

Chr	start	end
chr1A	1	594000000
chr1B	1	689000000
chr1D	1	495000000

- Label

chr	start	end	label
chr1A	25719162	25714282	TaAGL17-A6
chr1A	25791227	25795729	TaAGL17-A6
chr1A	150448830	150456774	TaVIL-A1

- Tracks data
  - Track1



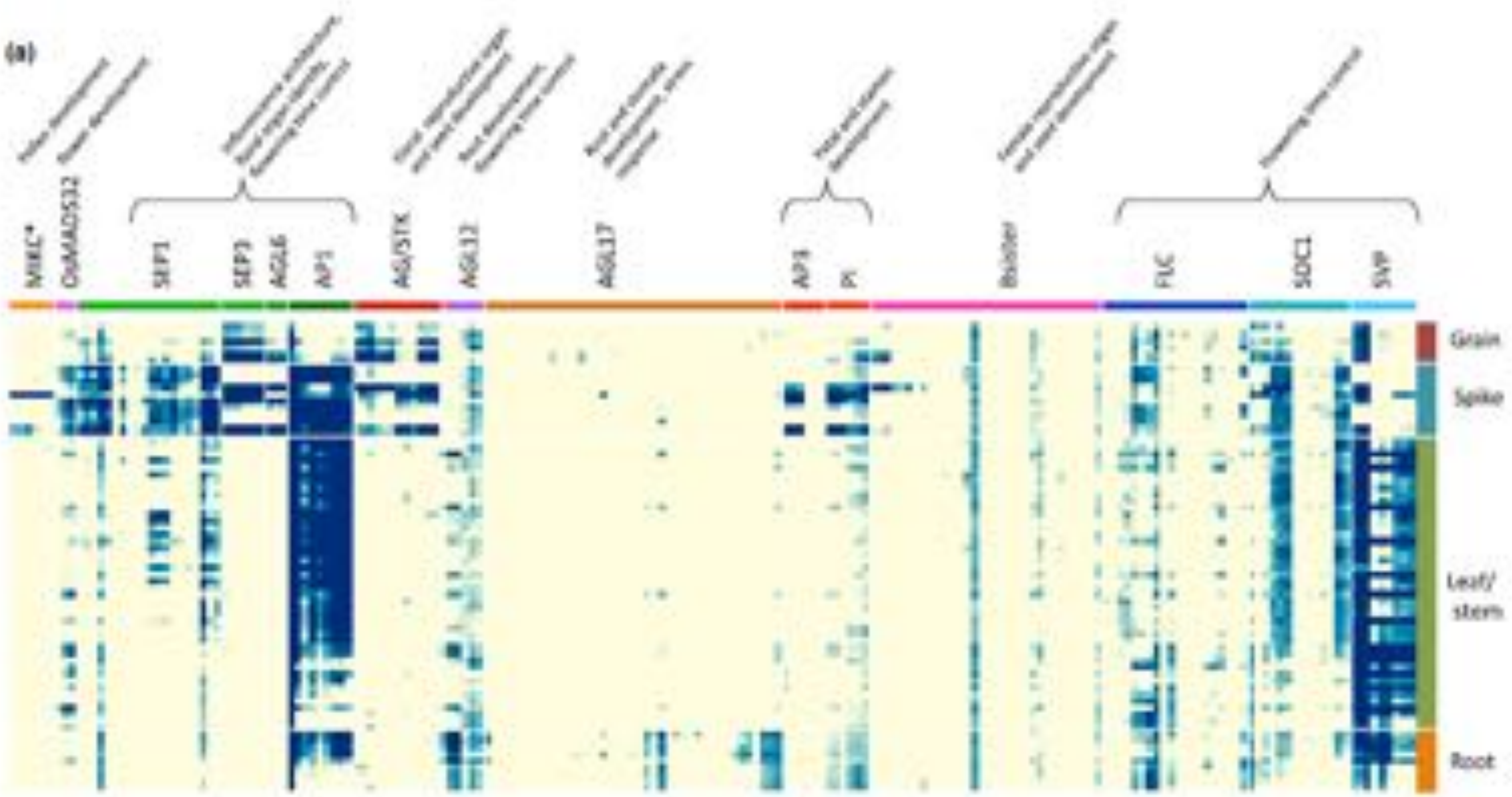
RStudio

https://github.com/venyao/shinyCircos

http://150.109.59.144:3838/shinyCircos/

Rainer Melzer & Susanne Schilling, UCD, Ireland

Expression analysis

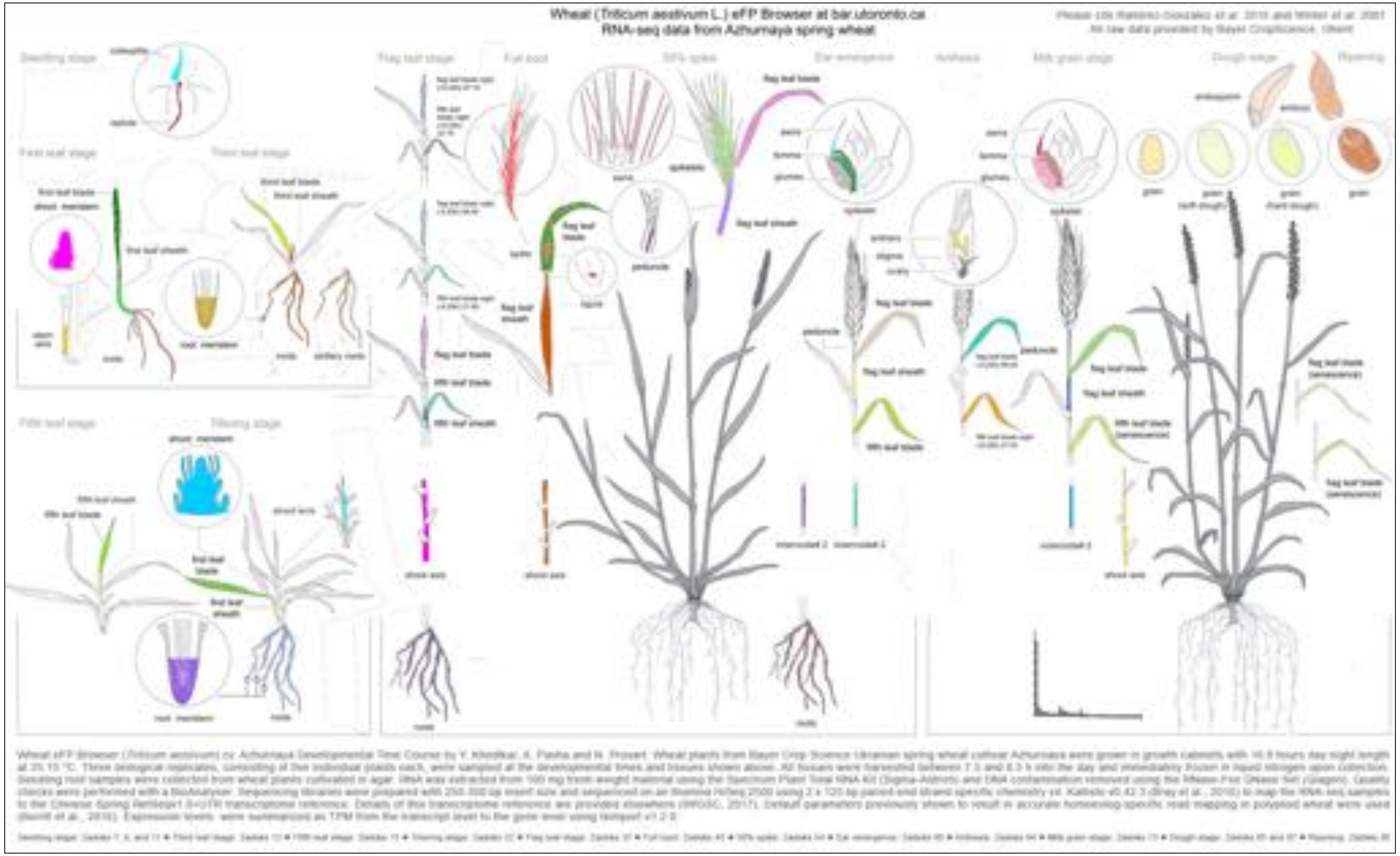


Data mining

Visualisation

Schilling et al. 2020

Expression data



Wheat (Triticum aestivum L.) eFP Browser at bar.utoronto.ca
RNA-seq data from Azhurnaya spring wheat

Please cite Ramírez-González et al. 2018 and Winter et al. 2007
All raw data provided by Bayer CropScience, Ghent

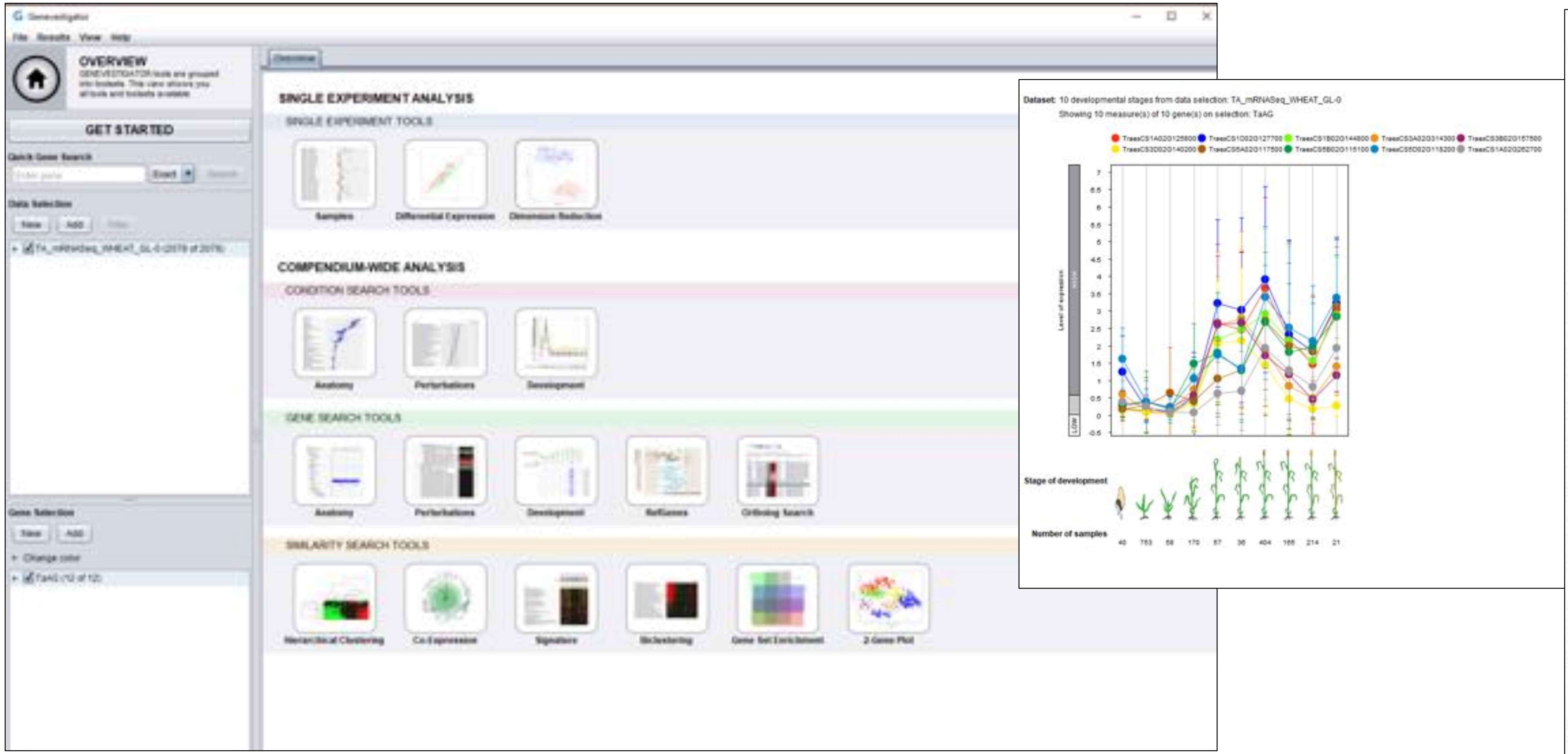
Wheat eFP Browser (Triticum aestivum) cv. Azhurnaya Developmental Time Course by V. Khedikar, A. Pasha and N. Provart. Wheat plants from Bayer Crop Science Ukrainian spring wheat cultivar Azhurnaya were grown in growth cabinets with 16:8 hours day:night length at 20:15 °C. Three biological replicates, consisting of five individual plants each, were sampled at the developmental times and tissues shown above. All tissues were harvested between 7.5 and 8.5 h into the day and immediately frozen in liquid nitrogen upon collection. Seedling root samples were collected from wheat plants cultivated in agar. RNA was extracted from 100 mg fresh weight material using the Spectrum Plant Total RNA Kit (Sigma-Aldrich) and DNA contamination removed using the RNase-Free DNase Set (Qiagen). Quality checks were performed with a BioAnalyser. Sequencing libraries were prepared with 250-350 bp insert size and sequenced on an Illumina HiSeq 2500 using 2 x 125 bp paired-end and strand specific chemistry. Kallisto v0.42.3 (Bray et al., 2016) to map the RNA-seq samples to the Chinese Spring RefSeqv1.0+UTR transcriptome reference. Details of this transcriptome reference are provided elsewhere (IWGSC, 2017). Default parameters previously shown to result in accurate homoeolog-specific read mapping in polyploid wheat were used (Borrill et al., 2016). Expression levels were summarized as TPM from the transcript level to the gene level using tximport v1.2.0.

Seedling stage: Zadoks 7, 9, and 11 ◆ Third leaf stage: Zadoks 13 ◆ Fifth leaf stage: Zadoks 15 ◆ Tillering stage: Zadoks 22 ◆ Flag leaf stage: Zadoks 37 ◆ Full boot: Zadoks 45 ◆ 50% spike: Zadoks 54 ◆ Ear emergence: Zadoks 60 ◆ Anthesis: Zadoks 64 ◆ Milk grain stage: Zadoks 71 ◆ Dough stage: Zadoks 85 and 87 ◆ Ripening: Zadoks 90

http://bar.utoronto.ca/efp_wheat/cgi-bin/efpWeb.cgi

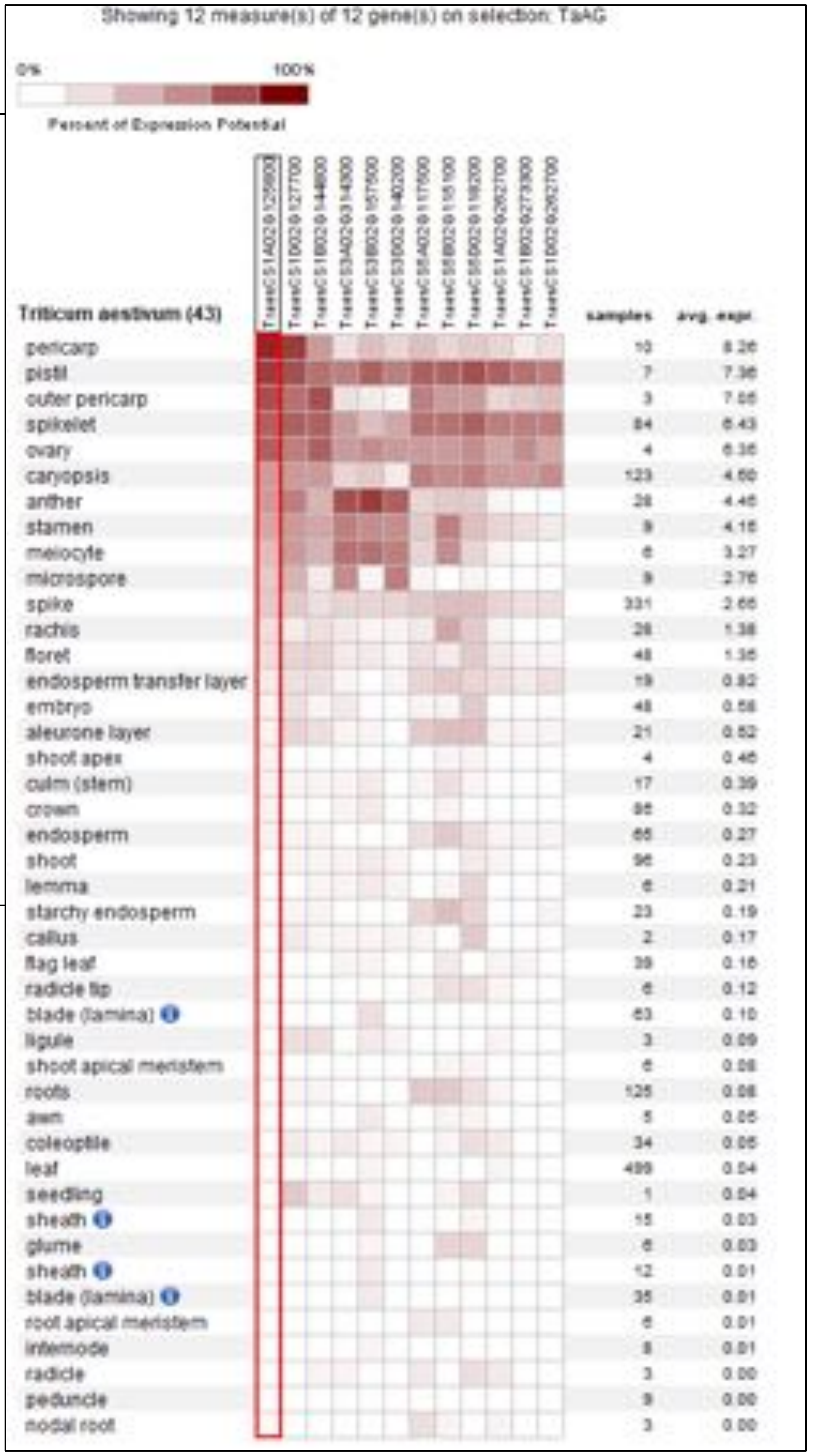
BAR Wheat eFP Browser

Genevestigator



Dataset: 10 developmental stages from data selection: TA_mRNASeq_WHEAT_GL-0

Showing 10 measure(s) of 10 gene(s) on selection: TaAG



Showing 12 measure(s) of 12 gene(s) on selection: TaAG

Triticum aestivum (43)	samples	avg. expr.
pericarp	10	8.26
pistil	7	7.36
outer pericarp	3	7.06
spikelet	84	6.43
ovary	4	6.36
caryopsis	123	4.60
anther	28	4.46
stamen	9	4.16
meiocyte	6	3.27
microspore	9	2.76
spike	331	2.66
rachis	28	1.38
floret	48	1.36
endosperm transfer layer	19	0.82
embryo	48	0.56
aleurone layer	21	0.52
shoot apex	4	0.46
culm (stem)	17	0.39
crown	86	0.32
endosperm	66	0.27
shoot	96	0.23
lemma	6	0.21
starchy endosperm	23	0.19
callus	2	0.17
flag leaf	39	0.16
radicle tip	6	0.12
blade (lamina)	63	0.10
ligule	3	0.09
shoot apical meristem	6	0.08
roots	125	0.08
awn	5	0.06
coleoptile	34	0.06
leaf	499	0.04
seedling	1	0.04
sheath	15	0.03
glume	6	0.03
sheath	12	0.01
blade (lamina)	35	0.01
root apical meristem	6	0.01
internode	8	0.01
radicle	3	0.00
peduncle	9	0.00
nodal root	3	0.00

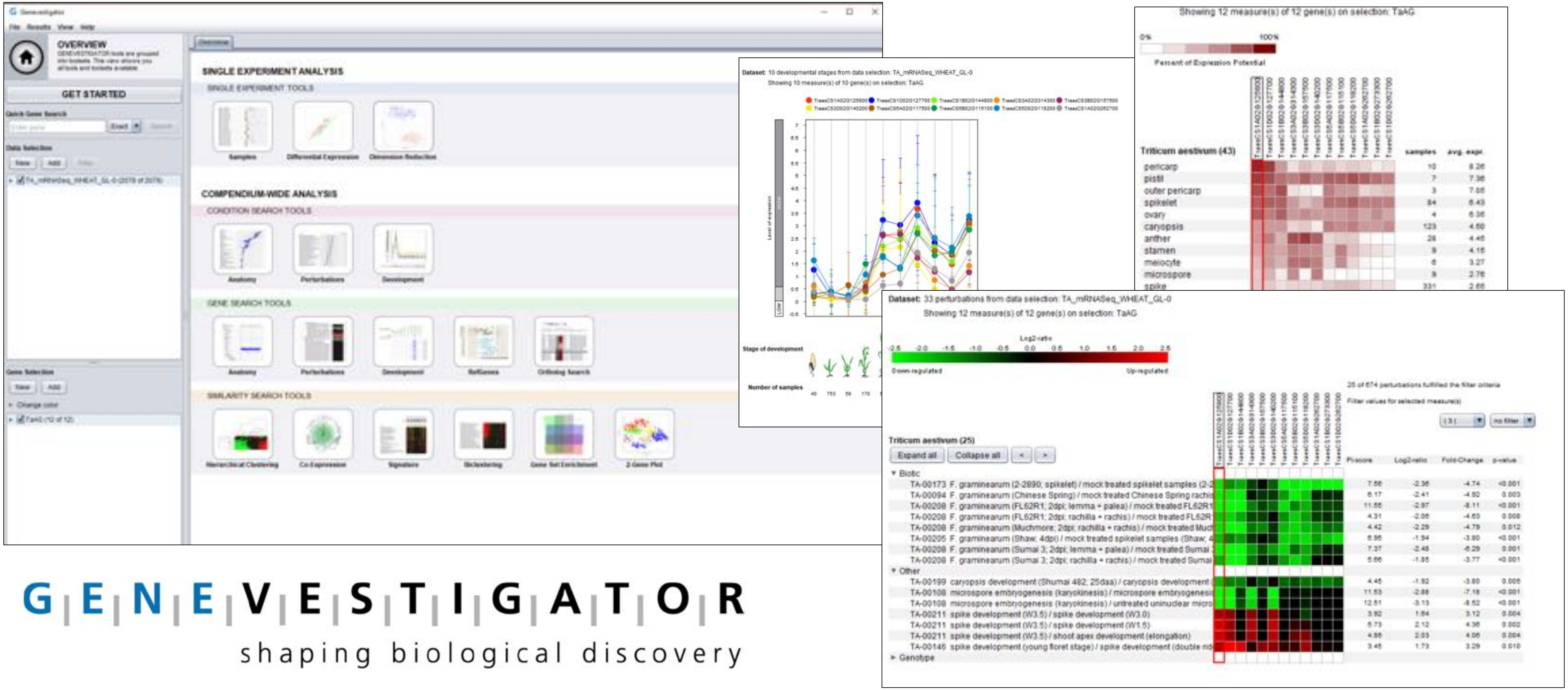
GENEVESTIGATOR

shaping biological discovery

https://genevestigator.com/

Rainer Melzer & Susanne Schilling, UCD, Ireland

Genevestigator



Genevestigator

File Results View Help

OVERVIEW
GENEVESTIGATOR tools are grouped into toolsets. This view shows you all tools and toolsets available.

GET STARTED

Quick Gene Search

Data Selection

☑ TA_mRNASeq_WHEAT_GL-0 (2078 of 2078)

Gene Selection

Change color

☑ TaAG (12 of 12)

Overview

SINGLE EXPERIMENT ANALYSIS

SINGLE EXPERIMENT TOOLS

Samples · Differential Expression · Dimension Reduction

COMPENDIUM-WIDE ANALYSIS

CONDITION SEARCH TOOLS

Anatomy · Perturbations · Development

GENE SEARCH TOOLS

Anatomy · Perturbations · Development · RefGenes · Ortholog Search

SIMILARITY SEARCH TOOLS

Hierarchical Clustering · Co-Expression · Signature · Biclustering · Gene Set Enrichment · 2-Gene Plot

Dataset: 10 developmental stages from data selection: TA_mRNASeq_WHEAT_GL-0

Showing 10 measure(s) of 10 gene(s) on selection: TaAG

Showing 12 measure(s) of 12 gene(s) on selection: TaAG

Triticum aestivum (43)	samples	avg. expr.
pericarp	10	8.26
pistil	7	7.36
outer pericarp	3	7.05
spikelet	84	6.43
ovary	4	6.36
caryopsis	123	4.60
anther	28	4.46
stamen	9	4.16
meiocyte	6	3.27
microspore	9	2.76
spike	331	2.66

Dataset: 33 perturbations from data selection: TA_mRNASeq_WHEAT_GL-0

Showing 12 measure(s) of 12 gene(s) on selection: TaAG

25 of 674 perturbations fulfilled the filter criteria

Filter values for selected measure(s)

Triticum aestivum (25)

Expand all · Collapse all · < · >

Triticum aestivum (25)	Pi-score	Log2-ratio	Fold-Change	p-value
▾ Biotic				
TA-00173 F. graminearum (2-2890; spikelet) / mock treated spikelet samples (2-2	7.56	-2.36	-4.74	<0.001
TA-00094 F. graminearum (Chinese Spring) / mock treated Chinese Spring rachis	6.17	-2.41	-4.82	0.003
TA-00208 F. graminearum (FL62R1; 2dpi; lemma + palea) / mock treated FL62R1	11.56	-2.97	-8.11	<0.001
TA-00208 F. graminearum (FL62R1; 2dpi; rachilla + rachis) / mock treated FL62R1	4.31	-2.06	-4.63	0.008
TA-00208 F. graminearum (Muchmore; 2dpi; rachilla + rachis) / mock treated Much	4.42	-2.29	-4.79	0.012
TA-00205 F. graminearum (Shaw; 4dpi) / mock treated spikelet samples (Shaw; 4	6.96	-1.94	-3.80	<0.001
TA-00208 F. graminearum (Sumai 3; 2dpi; lemma + palea) / mock treated Sumai 3	7.37	-2.48	-6.29	0.001
TA-00208 F. graminearum (Sumai 3; 2dpi; rachilla + rachis) / mock treated Sumai	5.66	-1.85	-3.77	<0.001
▾ Other				
TA-00199 caryopsis development (Shumai 482; 25daa) / caryopsis development (	4.45	-1.92	-3.80	0.005
TA-00108 microspore embryogenesis (karyokinesis) / microspore embryogenesis	11.53	-2.88	-7.18	<0.001
TA-00108 microspore embryogenesis (karyokinesis) / untreated uninuclear micro	12.51	-3.13	-8.62	<0.001
TA-00211 spike development (W3.5) / spike development (W3.0)	3.92	1.64	3.12	0.004
TA-00211 spike development (W3.5) / spike development (W1.5)	5.73	2.12	4.36	0.002
TA-00211 spike development (W3.5) / shoot apex development (elongation)	4.86	2.03	4.06	0.004
TA-00146 spike development (young floret stage) / spike development (double rid	3.45	1.73	3.29	0.010
▸ Genotype				

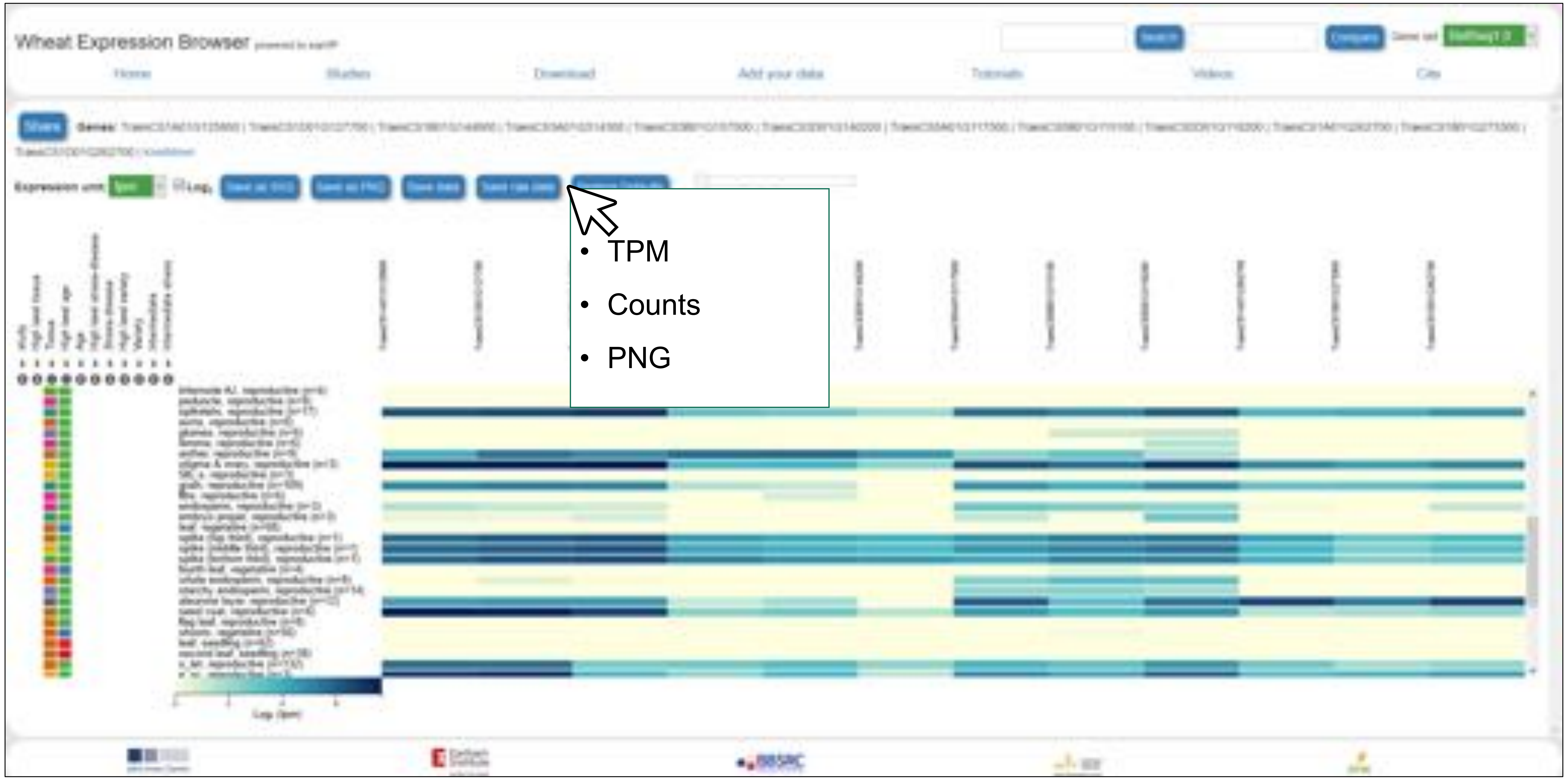
G|E|N|E|V|E|S|T|I|G|A|T|O|R

shaping biological discovery

https://genevestigator.com/

Rainer Melzer & Susanne Schilling, UCD, Ireland

Wheat-expression.com



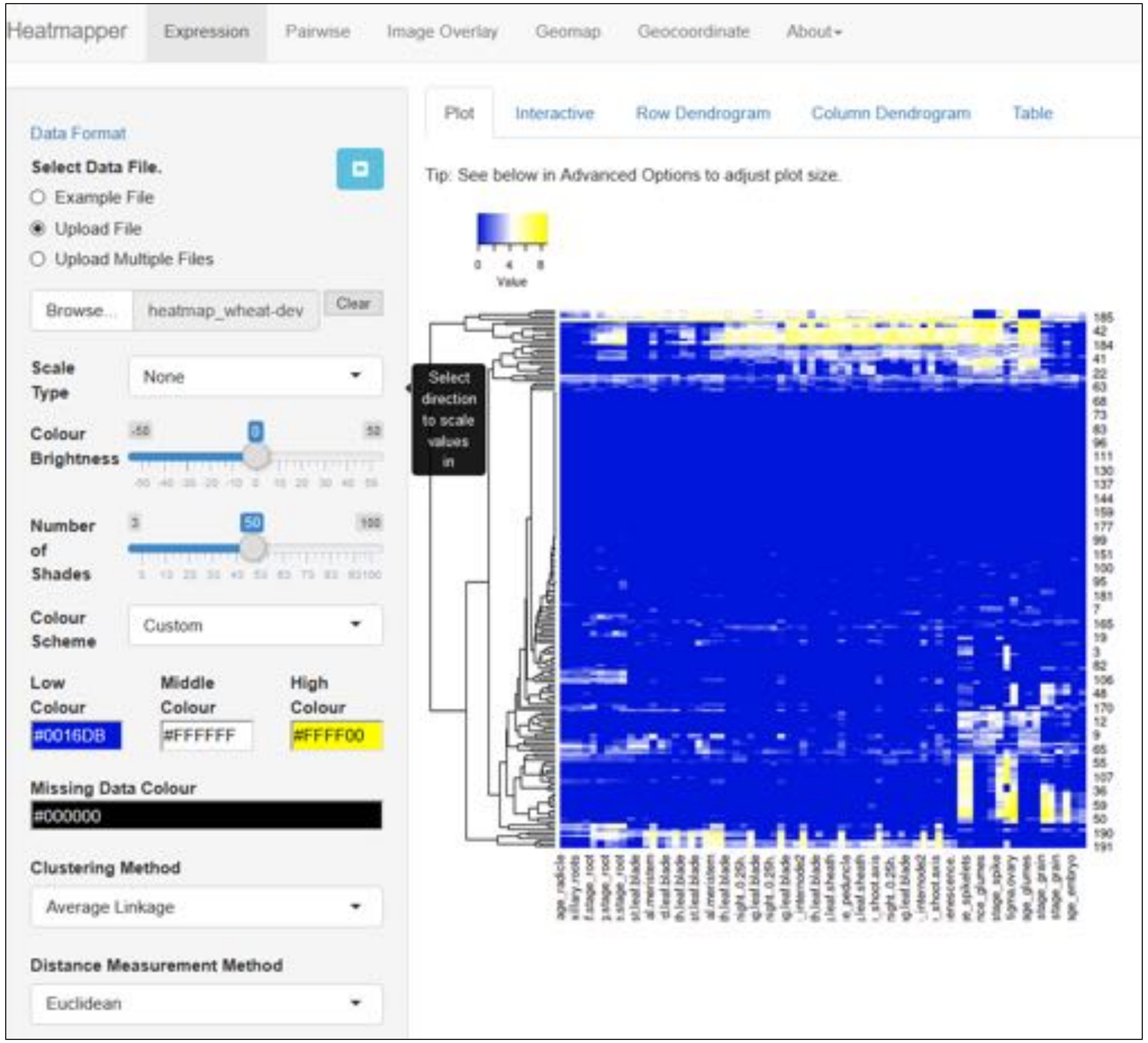
Wheat Expression Browser

- TPM
- Counts
- PNG

http://www.wheat-expression.com/

Rainer Melzer & Susanne Schilling, UCD, Ireland

Heatmap tools



Heatmapper | Expression | Pairwise | Image Overlay | Geomap | Geocoordinate | About▾

Data Format

Select Data File.

○ Example File

● Upload File

○ Upload Multiple Files

Browse... heatmap_wheat-dev Clear

Scale Type: None

Colour Brightness: 0 (-50 to 50)

Number of Shades: 50 (3 to 100)

Colour Scheme: Custom

Low Colour: #0016DB

Middle Colour: #FFFFFF

High Colour: #FFFF00

Missing Data Colour: #000000

Clustering Method: Average Linkage

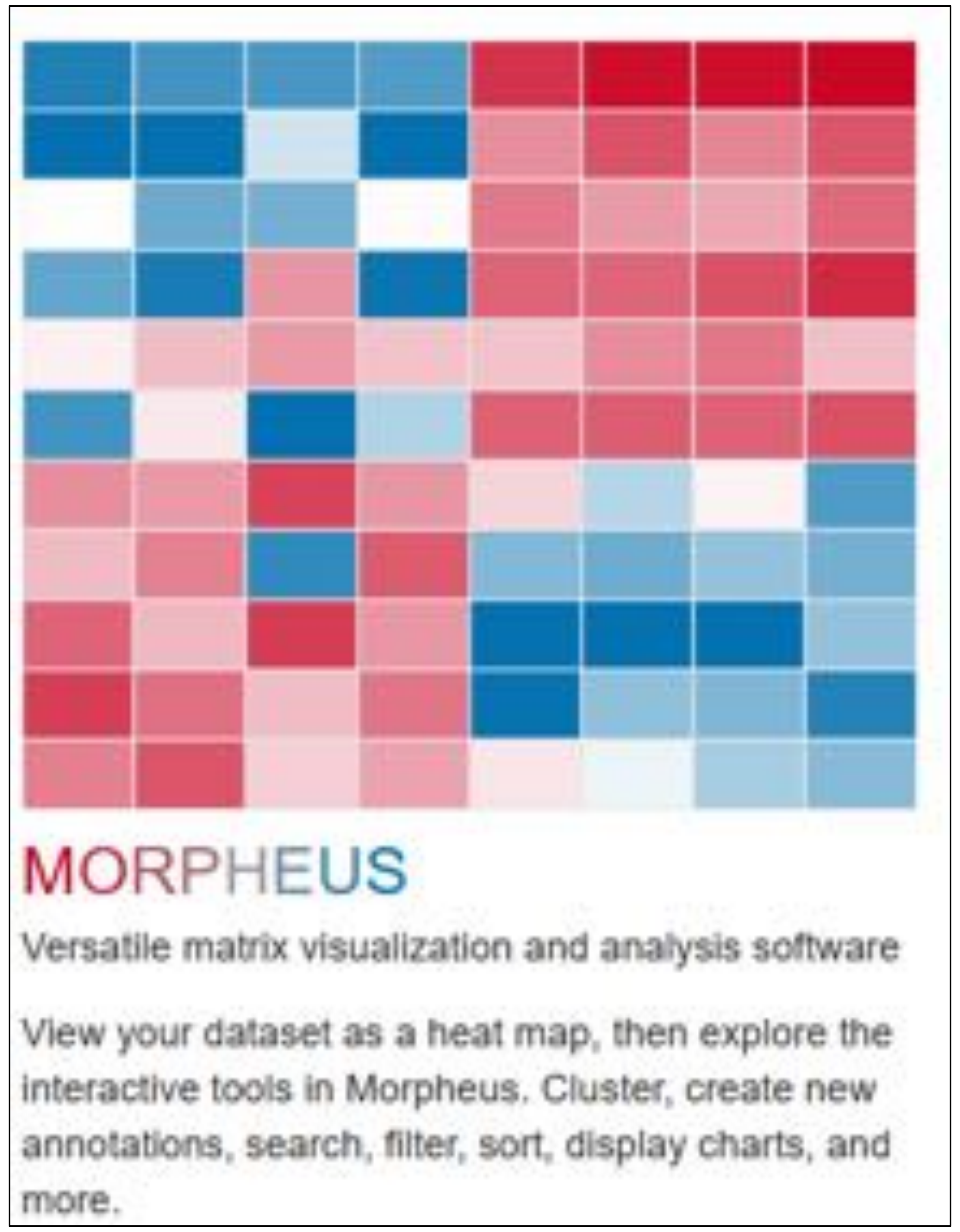
Distance Measurement Method: Euclidean

Plot | Interactive | Row Dendrogram | Column Dendrogram | Table

Tip: See below in Advanced Options to adjust plot size.

Select direction to scale values in

http://www.heatmapper.ca/expression/



MORPHEUS

Versatile matrix visualization and analysis software

View your dataset as a heat map, then explore the interactive tools in Morpheus. Cluster, create new annotations, search, filter, sort, display charts, and more.

https://software.broadinstitute.org/morpheus/

More wheat resources…

JBrowse

https://urgi.versailles.inra.fr/jbrowseiwgsc/gmod_jbrowse/

URGI GnpIS GENETIC AND GENOMIC INFORMATION SYSTEM

https://urgi.versailles.inra.fr/synteny/synteny/

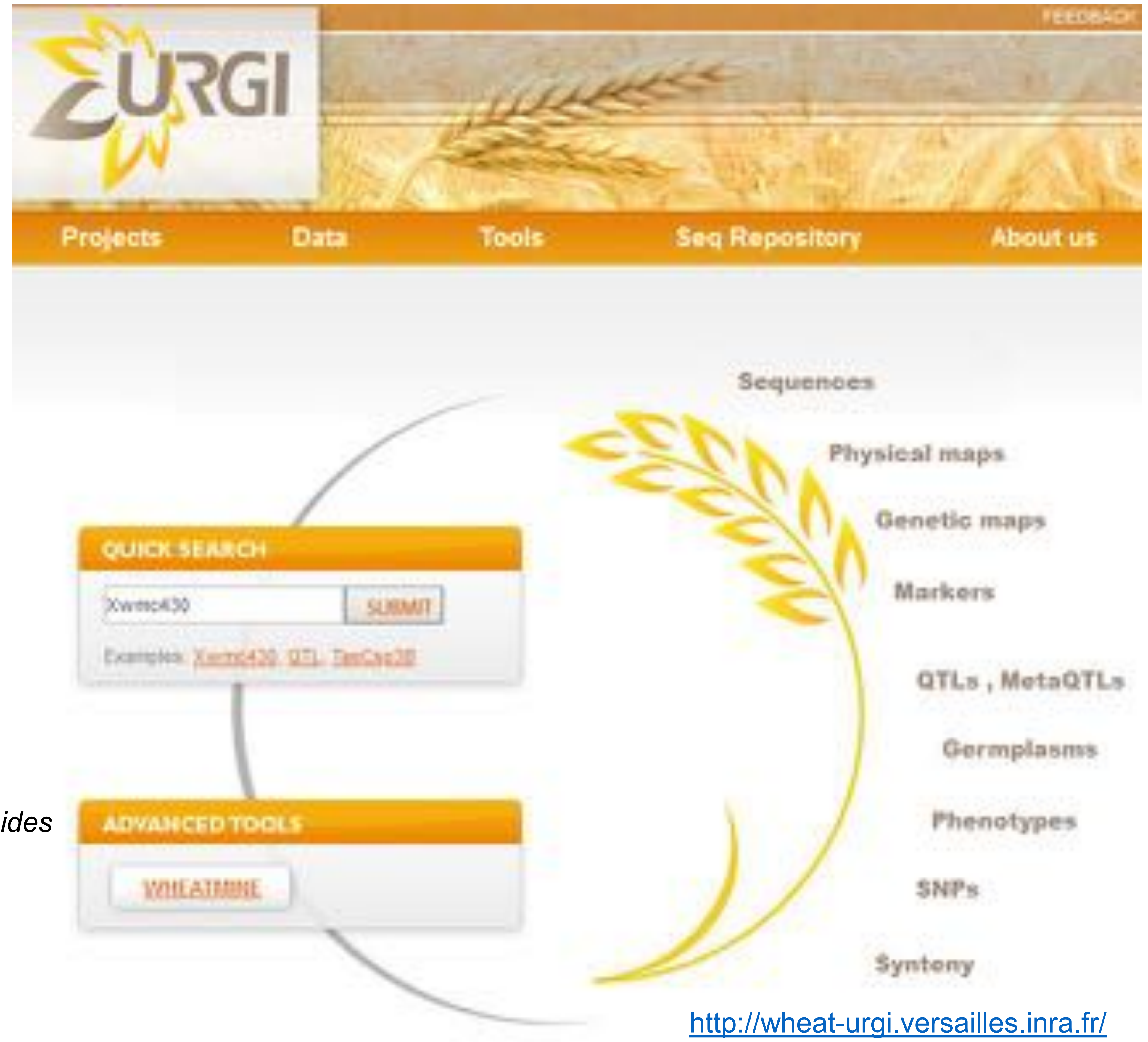
EnsemblPlants

plants.ensembl.org/

Aegilops tauschii
Tritcum urartu

Synteny
EMS mutants

Aegilops speltoides



http://wheat-urgi.versailles.inra.fr/

Other resources…

NCBI
National Center for Biotechnology Information
https://www.ncbi.nlm.nih.gov/

RStudio
https://rstudio.com/

MAFFT version 7
Multiple alignment program for amino acid or nucleotide sequences
https://mafft.cbrc.jp/alignment/server/

IQ-TREE
http://www.iqtree.org/

http://hmmer.org/

Galaxy
https://usegalaxy.org/

START: Shiny Transcriptome Analysis Resource Tool
https://kcvi.shinyapps.io/START/

FigTree v1.4.4
https://github.com/rambaut/figtree/releases

https://www.youtube.com/

Biostars
BIOINFORMATICS EXPLAINED
https://www.biostars.org/

edX
https://www.edx.org/

EMBL-EBI Hinxton
https://www.ebi.ac.uk/training/online/

Thank you for listening

Questions?

SiRui Pan

Lars Jermiin

Alice Kennedy

susanne.schilling@ucd.ie

rainer.melzer@ucd.ie

Susanne Schilling & Rainer Melzer
School of Biology and Environmental Science
University College Dublin