



Linking the International Wheat Genome Sequencing Consortium Bread Wheat Reference Genome Sequence to Wheat Genetic and Phenomic Data

Michael Alaux

► To cite this version:

Michael Alaux. Linking the International Wheat Genome Sequencing Consortium Bread Wheat Reference Genome Sequence to Wheat Genetic and Phenomic Data. PAG XXVII - Plant and Animal Genome Conference, Jan 2019, San Diego, United States. hal-04536576

HAL Id: hal-04536576

<https://hal.inrae.fr/hal-04536576>

Submitted on 8 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



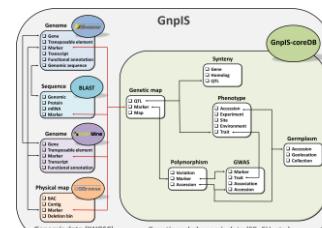
Distributed under a Creative Commons Attribution 4.0 International License

Linking the IWGSC bread wheat reference genome sequence to wheat genetic and phenomic data

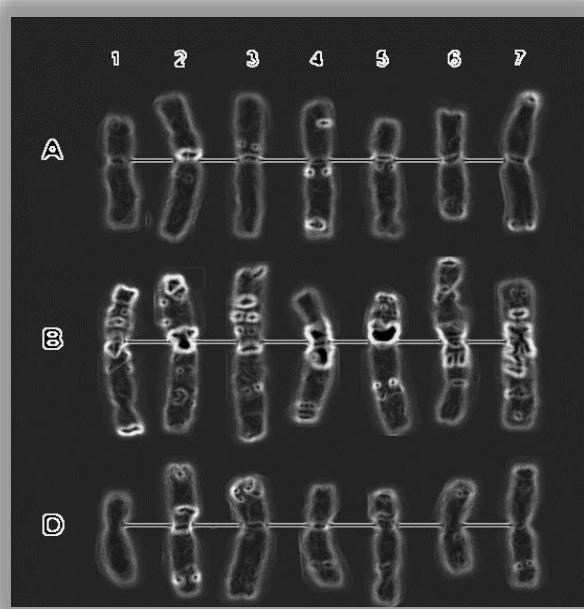
Michael Alaux



Triticeae Genetics and Genomics, Session 1: Progress in structural and functional genomics



Outline



A. The wheat reference sequence

B. A wealth of data available

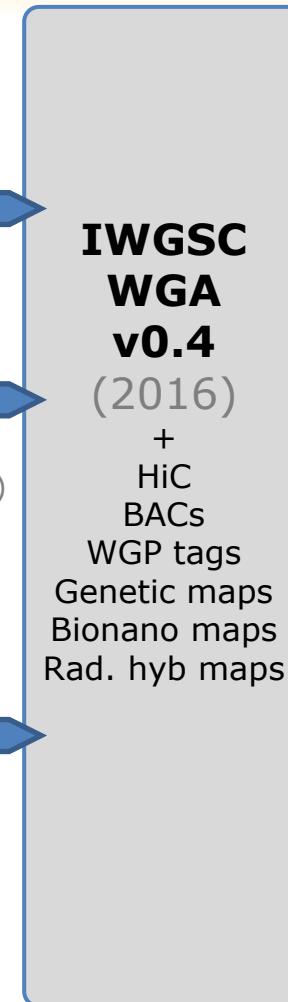
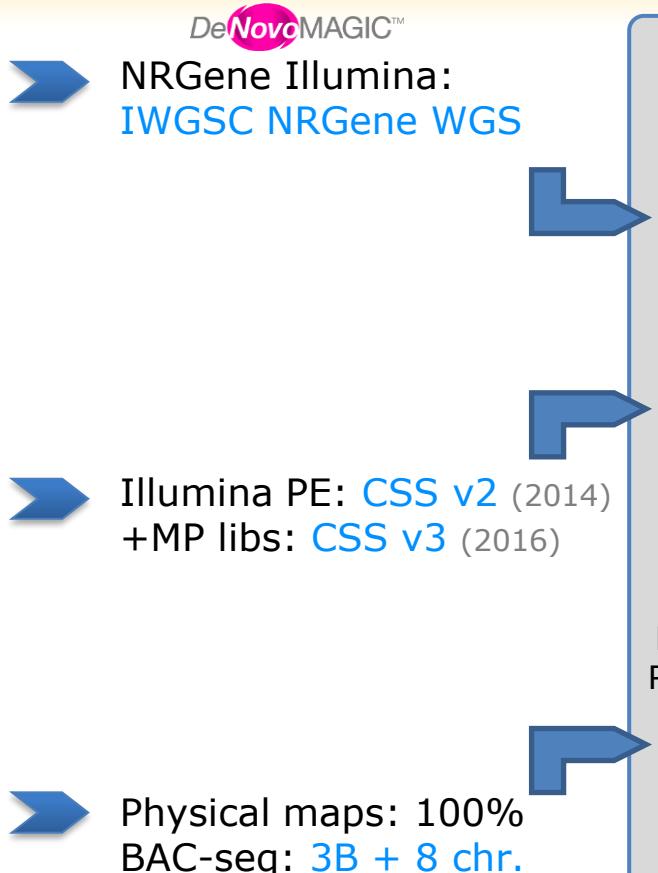
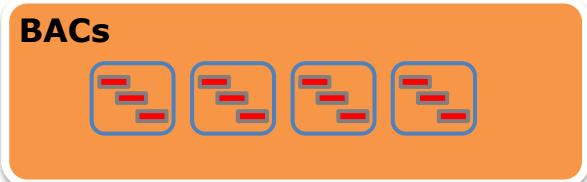
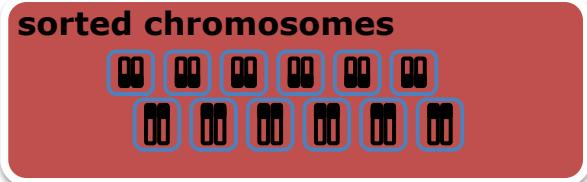
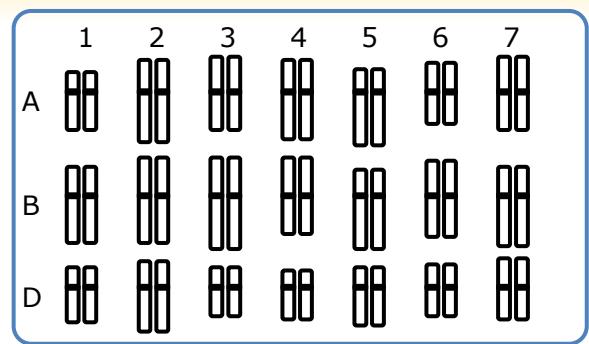
D. Data integration

→ **Wheat@URGI portal**



The wheat reference sequence

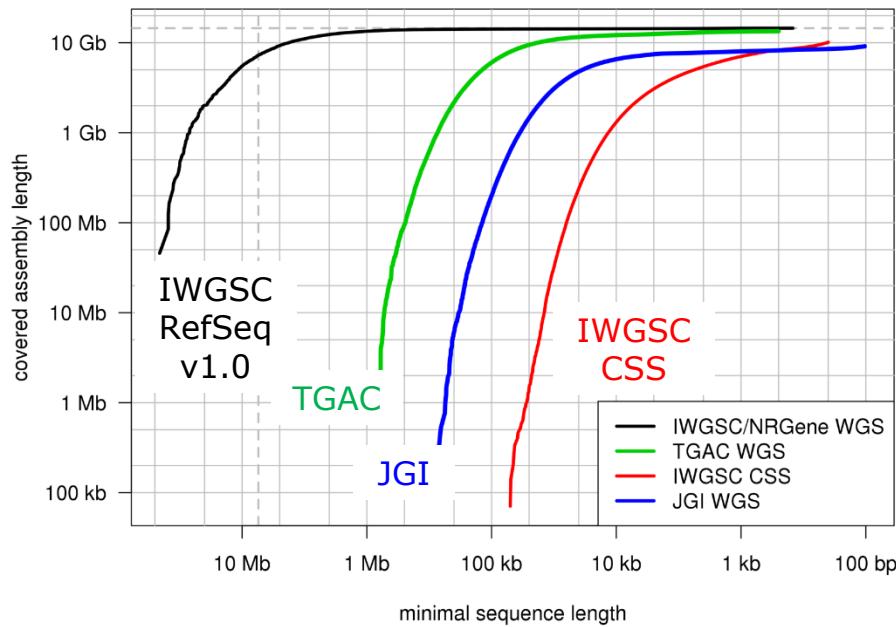
IWGSC sequencing strategy



Courtesy of Fred Choulet

IWGSC RefSeq assembly metrics

total size 14.5 Gb (21 pseudomolecules = 14.1 Gb)
completeness 97-99%
superscaffold N50 23 Mb

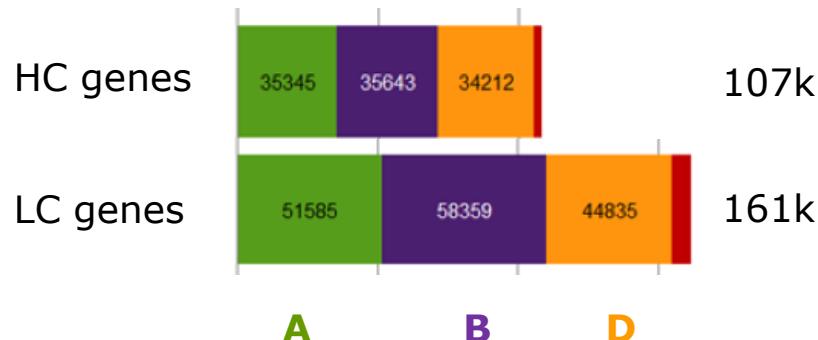


IWGSC RefSeq annotation

Genes

v1.0: automated

v1.1: incorporate 3685 manually annotated genes

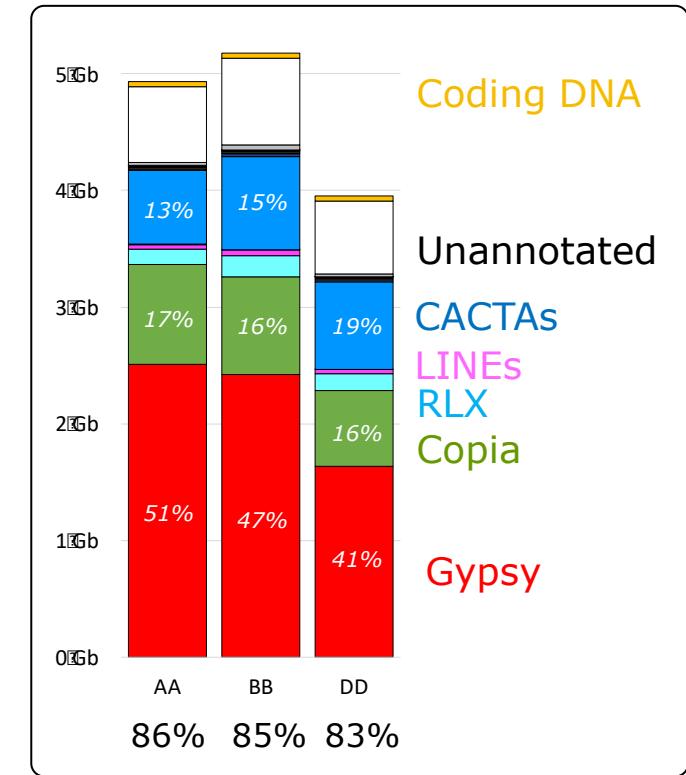
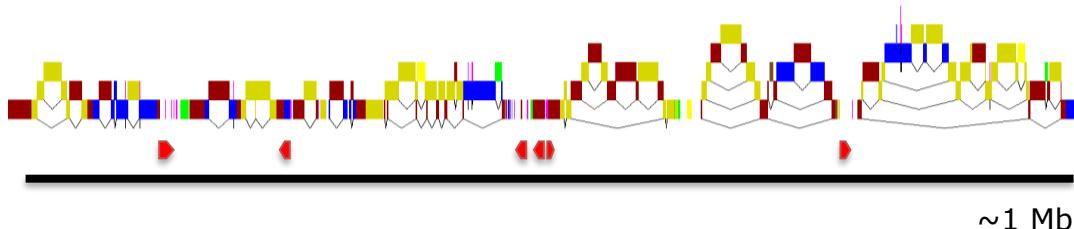


IWGSC, *Science* 2018
<https://doi.org/10.1126/science.aar7191>

IWGSC RefSeq annotation

Transposable Elements

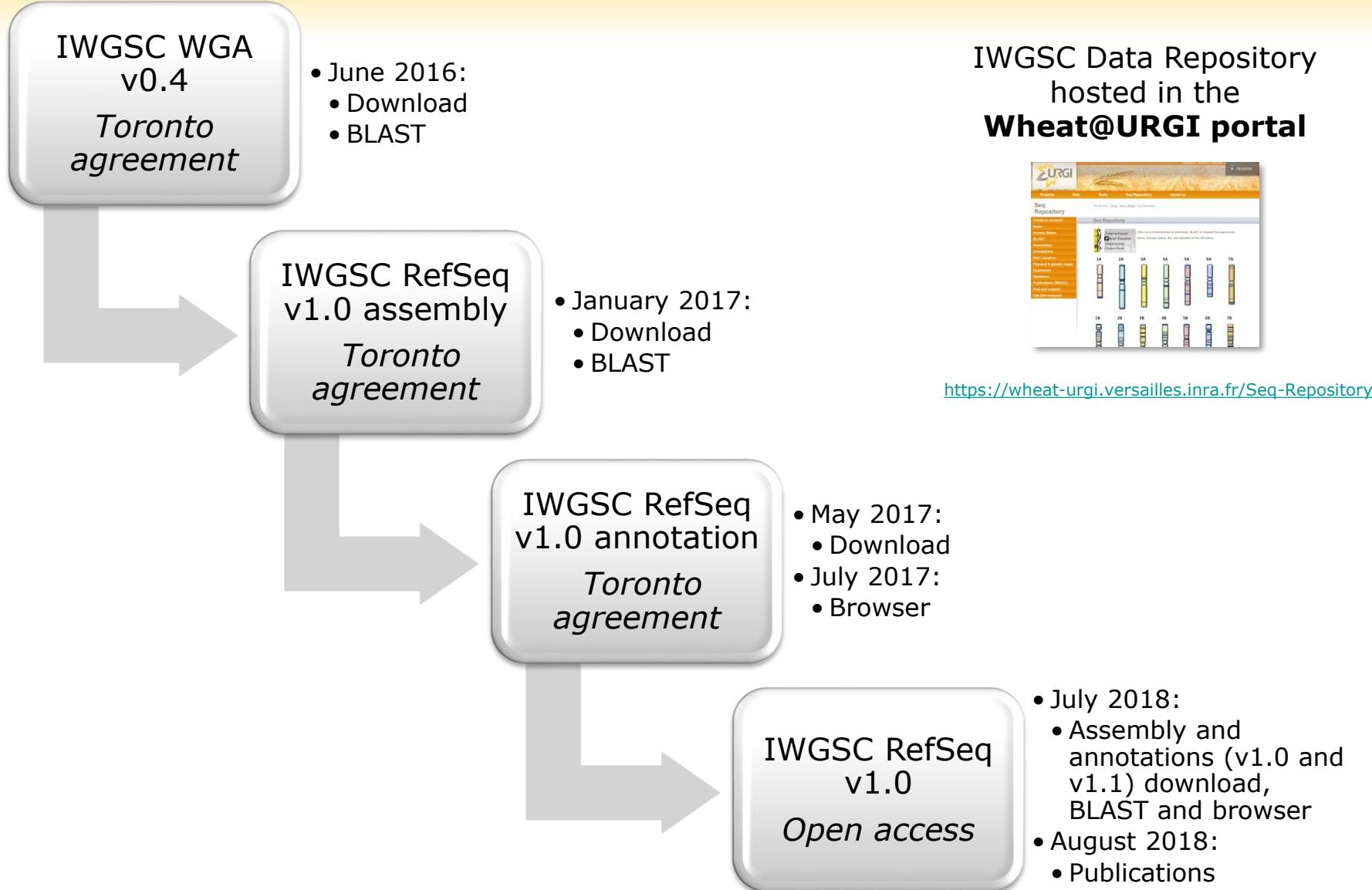
4 million TEs



+ ncRNA, markers, functional annotation, manually curated genes, etc.

Wicker et al., *Genome Biology* 2018
<https://doi.org/10.1186/s13059-018-1479-0>

IWGSC RefSeq data availability



Publications



RESEARCH ARTICLE

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

The International Wheat Genome Sequencing Consortium (IWGSC), IWGSC RefSeq principal investigators:, Rudi Appels^{1,36,*} ...

+ See all authors and affiliations

Science 17 Aug 2018:
Vol. 361, Issue 6403, eaar7191
DOI: 10.1126/science.aar7191

RESEARCH ARTICLE

The transcriptional landscape of polyploid wheat

R. H. Ramírez-González^{1,*}, P. Borrell^{1,*†}, D. Lang², S. A. Harrington¹, J. Brinton¹, L. Venturini³, M. Davey⁴, J. Jacobs⁴, F. van...

+ See all authors and affiliations

Science 17 Aug 2018:
Vol. 361, Issue 6403, eaar6089
DOI: 10.1126/science.aar6089

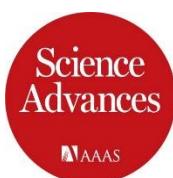
RESEARCH ARTICLE | PLANT SCIENCES

Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat

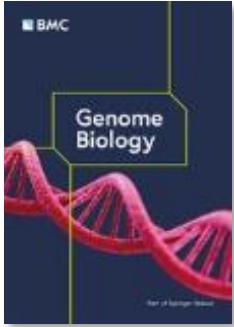
Ángela Juhász^{1,2,*}, Tatiana Belova^{3,*}, Chris G. Florides¹, Csaba Maulis¹, Iris Fischer⁴, Gyöngyvér Gell², Zsófia Birinyi², Jami...

+ See all authors and affiliations

Science Advances 17 Aug 2018:
Vol. 4, no. 8, eaar8602
DOI: 10.1126/sciadv.aar8602



Publications



Research

Chromosome mechanisms

Recent improvement of novo assemblies of

Anupriya Kaur Thi Wulff, Manuel Spe G. Krattinger

Genome Biology 20 Published on: 17 Au

[Full Text](#) [PDF](#)

Research

Impact of transposons on wheat

Transposable elements
The most recent assembly

Thomas Wicker, H González, Romain

Genome Biology 20 Published on: 17 Au

[Full Text](#) [PDF](#)

Research

Optical and resolution of

Numerous scaffold improving the over

Gabriel Keeble-G Frenkel, Abraham Konkin, Pierre Sou

Genome Biology 20 Published on: 17 Au

[Full Text](#) [PDF](#)

Database

Linking the International Wheat Genome

The Wheat@URGI

with access to the bread wheat reference genome sequence produced by the International Wheat Genome S...

Michael Alaux, Jane Rogers, Thomas Letellier, Raphaël Flores, Françoise Alfama, Cyril Pommier, Nacer Mohelbi, Sophie Durand, Erik Kimmel, Célia Michotey, Claire Guerche, Mikaël Loaec, Mathilde Lainé, Delphine Steinbach, Frédéric Choulet, Hélène Rimbert, Philippe Leroy, Nicolas Guilhot, Jérôme Salse, Catherine Feuillet, International Wheat Genome Sequencing Consortium, Etienne Paux, Kellye Eversole, Anne-Françoise Adam-Blondon, and Hadi Quesneville

Alaux et al. *Genome Biology* (2018) 19:111
<https://doi.org/10.1186/s13059-018-1491-4>

Genome Biology

DATABASE

Open Access



Linking the International Wheat Genome Sequencing Consortium bread wheat reference genome sequence to wheat genetic and phenomic data

Michael Alaux¹ , Jane Rogers², Thomas Letellier¹, Raphaël Flores¹, Françoise Alfama¹, Cyril Pommier¹, Nacer Mohelbi¹, Sophie Durand¹, Erik Kimmel¹, Célia Michotey¹, Claire Guerche¹, Mikaël Loaec¹, Mathilde Lainé¹, Delphine Steinbach^{1,4}, Frédéric Choulet³, Hélène Rimbert³, Philippe Leroy³, Nicolas Guilhot³, Jérôme Salse³, Catherine Feuillet^{3,5}, International Wheat Genome Sequencing Consortium⁶, Etienne Paux³, Kellye Eversole⁷, Anne-Françoise Adam-Blondon¹ and Hadi Quesneville¹

Abstract

The Wheat@URGI portal has been developed to provide the international community of researchers and breeders with access to the bread wheat reference genome sequence produced by the International Wheat Genome Sequencing Consortium. Genome browsers, BLAST, and InterMine tools have been established for in-depth exploration of the genome sequence together with additional linked datasets including physical maps, sequence variations, gene expression, and genetic and phenomic data from other international collaborative projects already stored in the GnpS information system. The portal provides enhanced search and browser features that will facilitate the deployment of the latest genomics resources in wheat improvement.

Keywords: Data integration, Information system, Big data, Wheat genomics, genetics and phenomics

Background

The International Wheat Genome Sequencing Consortium (IWGSC) [1] is an international collaborative group of growers, academic scientists, and public and private breeders that was established to generate a high-quality reference genome sequence of the hexaploid bread wheat, and to provide breeders with state-of-the-art tools for wheat improvement. The vision of the consortium is that the high-quality, annotated ordered genome sequence integrated with physical maps will serve as a foundation for the accelerated development of improved varieties and will empower all aspects of basic and applied wheat science. A first analysis of the reference sequence produced by the consortium (IWGSC RefSeq v1.0) was recently published [2].

To ensure that wheat breeding and research programs can make the most of this extensive genomic resource, the IWGSC endorsed the establishment of a data repository at URGI (Unité de Recherche Génomique Info/research unit in genomics and bioinformatics) from INRA (Institut National de la Recherche Agronomique/French national institute for agricultural research) to develop databases and browsers with relevant links to public data available worldwide. The IWGSC data repository is thus hosted by URGI to support public and private parties in data management as well as analysis and usage of the sequence data. Wheat functional genomics (expression, methylation, etc.), genetic, and phenomic data have increased concurrently, requiring the development of additional tools and resources to integrate different data for biologists and breeders. To manage this escalation of data, URGI has built this data repository for the wheat community with the following specific aims: (1) to store resources for which no public archive exists (e.g. physical maps,

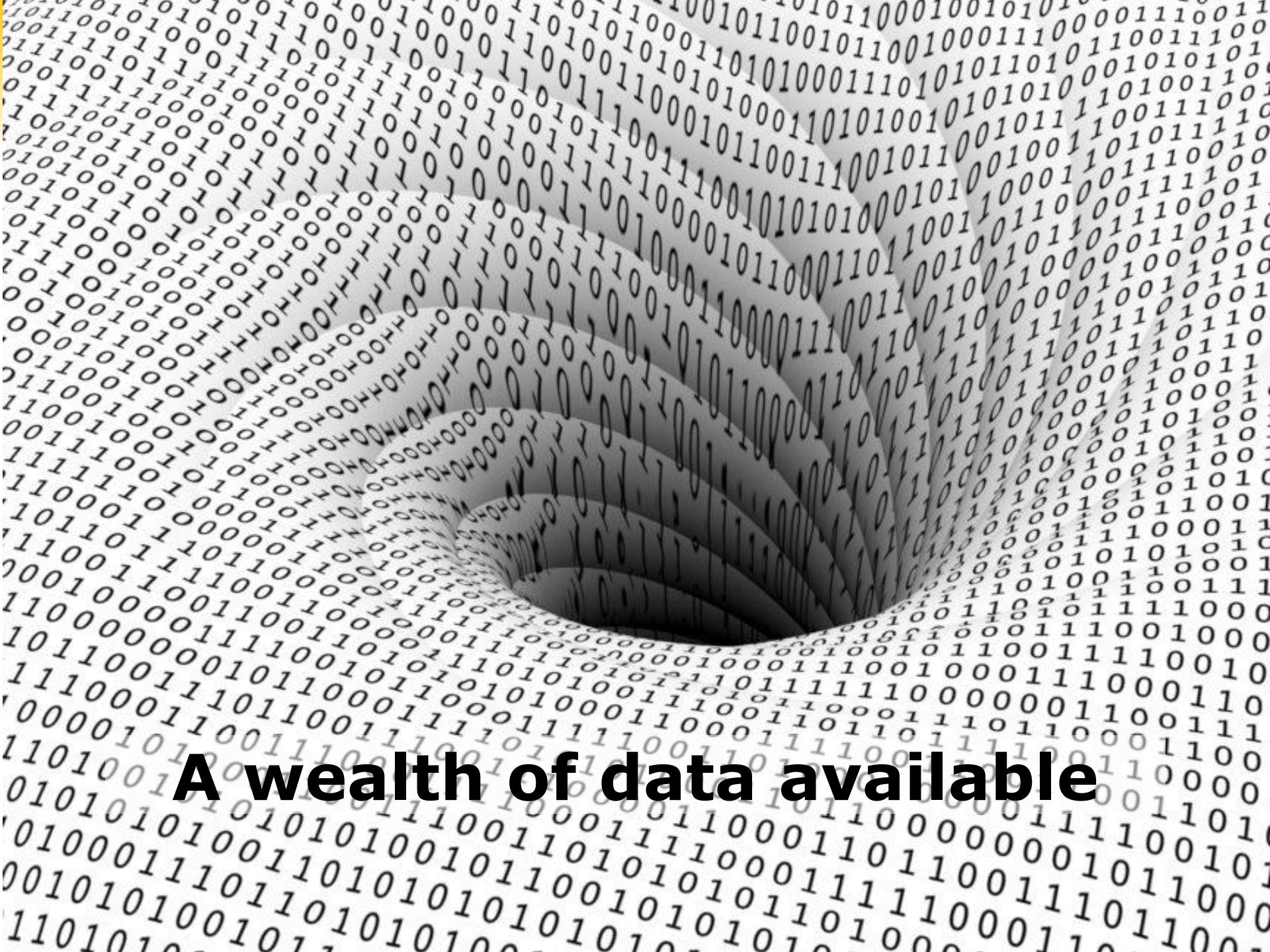
* Correspondence: michael.alaux@inra.fr

¹URGI-INRA, Université Paris-Saclay, 78026 Versailles, France
Full list of author information is available at the end of the article



© The Author(s). 2018 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

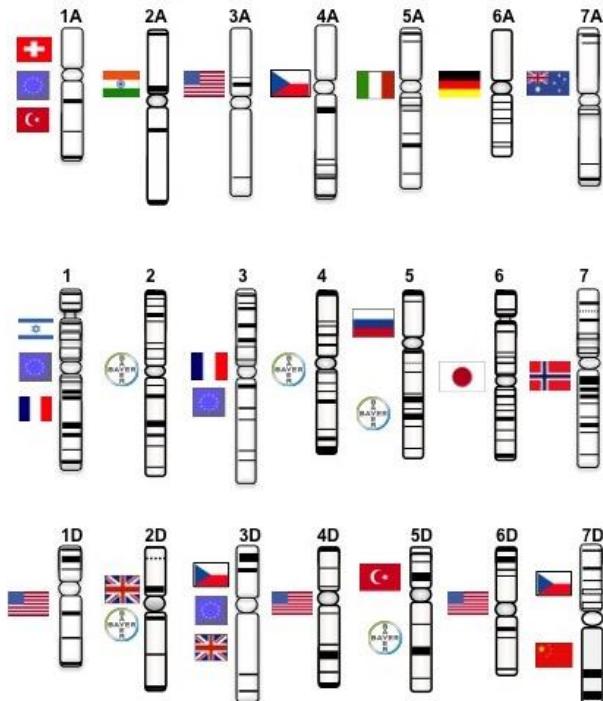
Alaux et al., *Genome Biology* 2018
<https://doi.org/10.1186/s13059-018-1491-4>



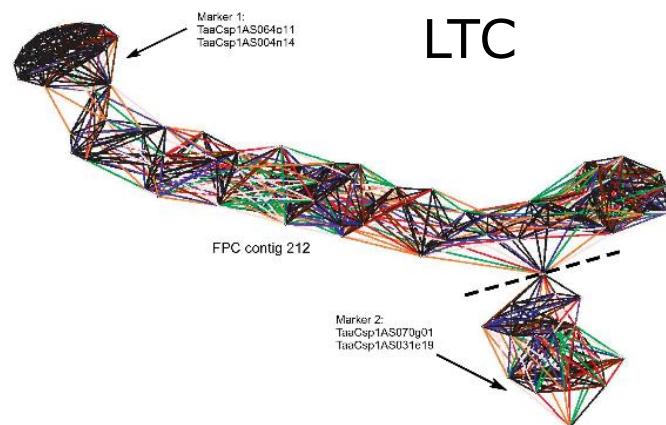
A wealth of data available

Other IWGSC data hosted

Physical maps



T. aestivum
cv Chinese Spring
I W G S C



Frenkel et al., *BMC Bioinformatics* 2010
<https://doi.org/10.1186/1471-2105-11-584>

Other IWGSC data hosted

Expression data

Transcriptome atlas

RNA-seq mapping of 850 samples – 32 tissues

- 85% HC genes expressed
- 49% LC genes expressed

Ramírez-González et al., *Science* 2018
<https://doi.org/10.1126/science.aar6089>

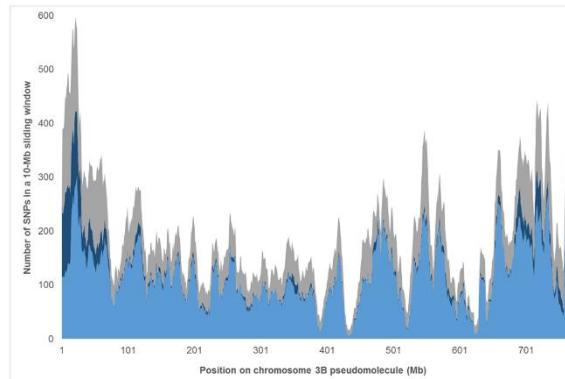
Variation data

SNPs

GBS

Exome capture

Varietal SNPs



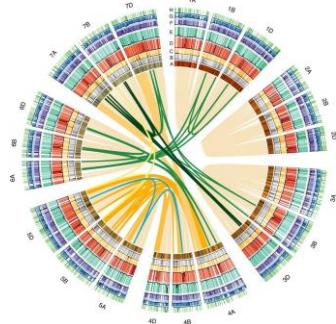
Rimbert et al., *PLOS ONE* 2018
<https://doi.org/10.1371/journal.pone.0186329>

Dubcovsky, Akhunov

Wheat gene pool

Other CS assembly

TGACv1



Clavijo et al., *Genome Research* 2017
<https://doi.org/10.1101/gr.217117.116>

Other wheat species

- Diploid progenitor of *Aegilops tauschii*
- *Triticum durum*
- etc.

Wheat genetic and phenomic data

from French and European projects since 2000

Data type	Object	Total number	No. open access	No. restricted access to projects
Genetic resources	Taxon	56	56	0
	Accession	12,839	10,016	2823
Genetic maps	Map	30	29	1
	Marker	704,822	34,164	670,658
	QTL	749	465	284
SNP discovery	Sequence Variation	4,189,312,581	90	4,189,312,491
	SNP, indel	724,132	95	724,037
Genotyping (high throughput)	Experiment	23	2	21
	Sample	8885	47	8872
	Marker	668,540	0	668,540
Phenotyping	Trial	850	821	29
	Plot	3660	2985	901
	Variable	282	89	195
	Observation	1,171,172	527,981	643,191
GWAS	Analysis	1555	43	1512
	Sample	2365	1839	526
	Variable	359	37	322
	Marker	123,866	4109	119,757
	Association	824,217	48,596	775,621





Data integration

Make your data great



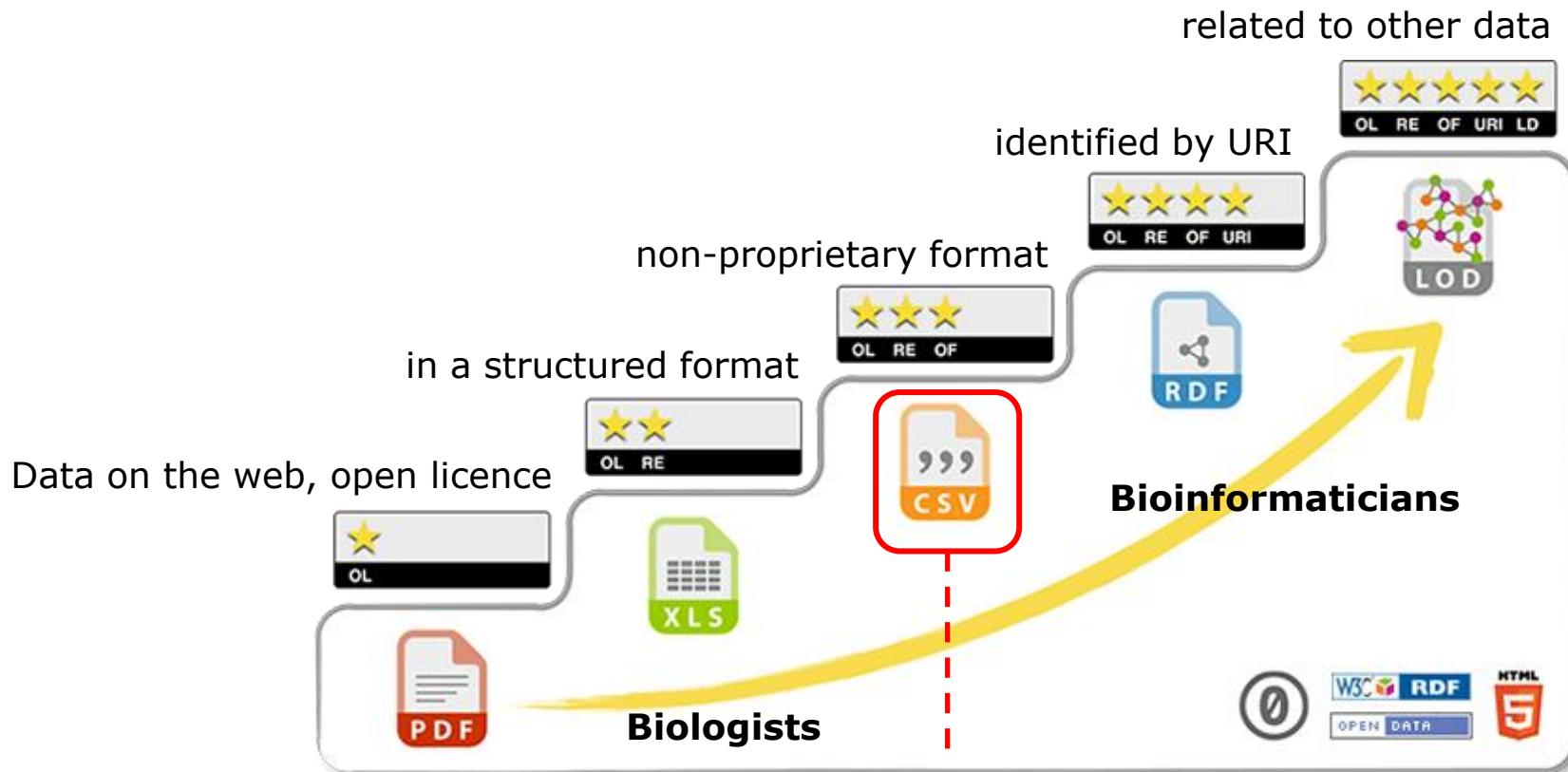
Persistent identifier
(e.g. URI, DOI)
Metadata
Indexed

Data portal
Open data

Web API
Metadata
DMP
Ontologies

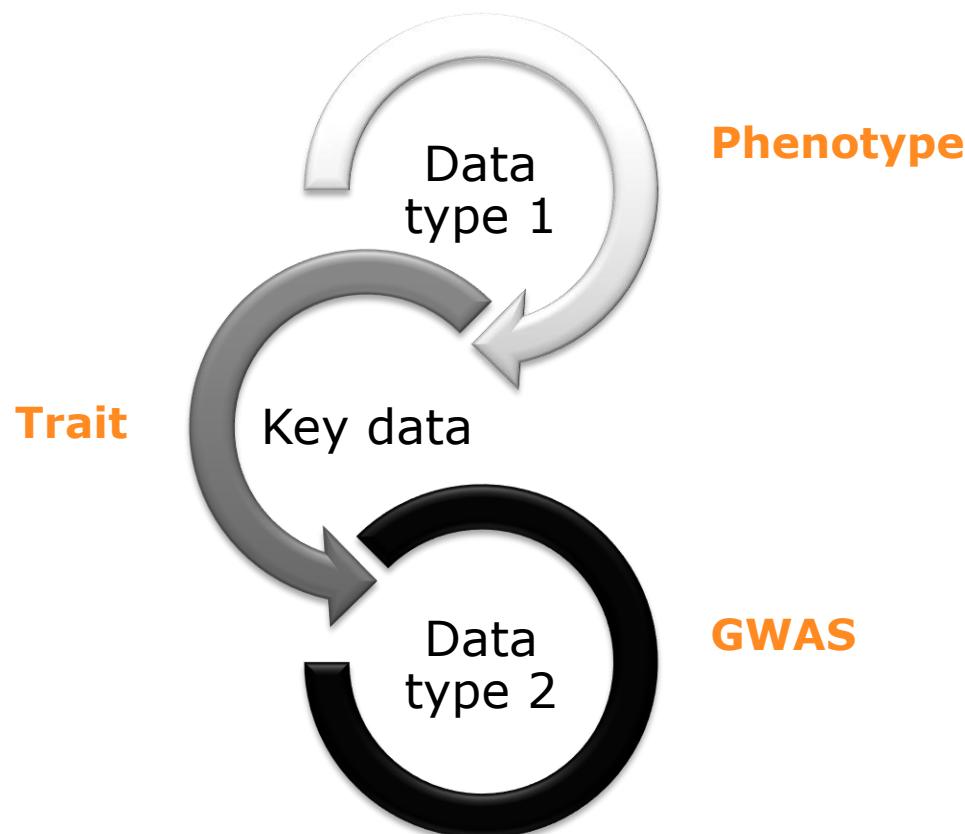
Standards
Licence
Documentation
Publication

5 ★ Open Data

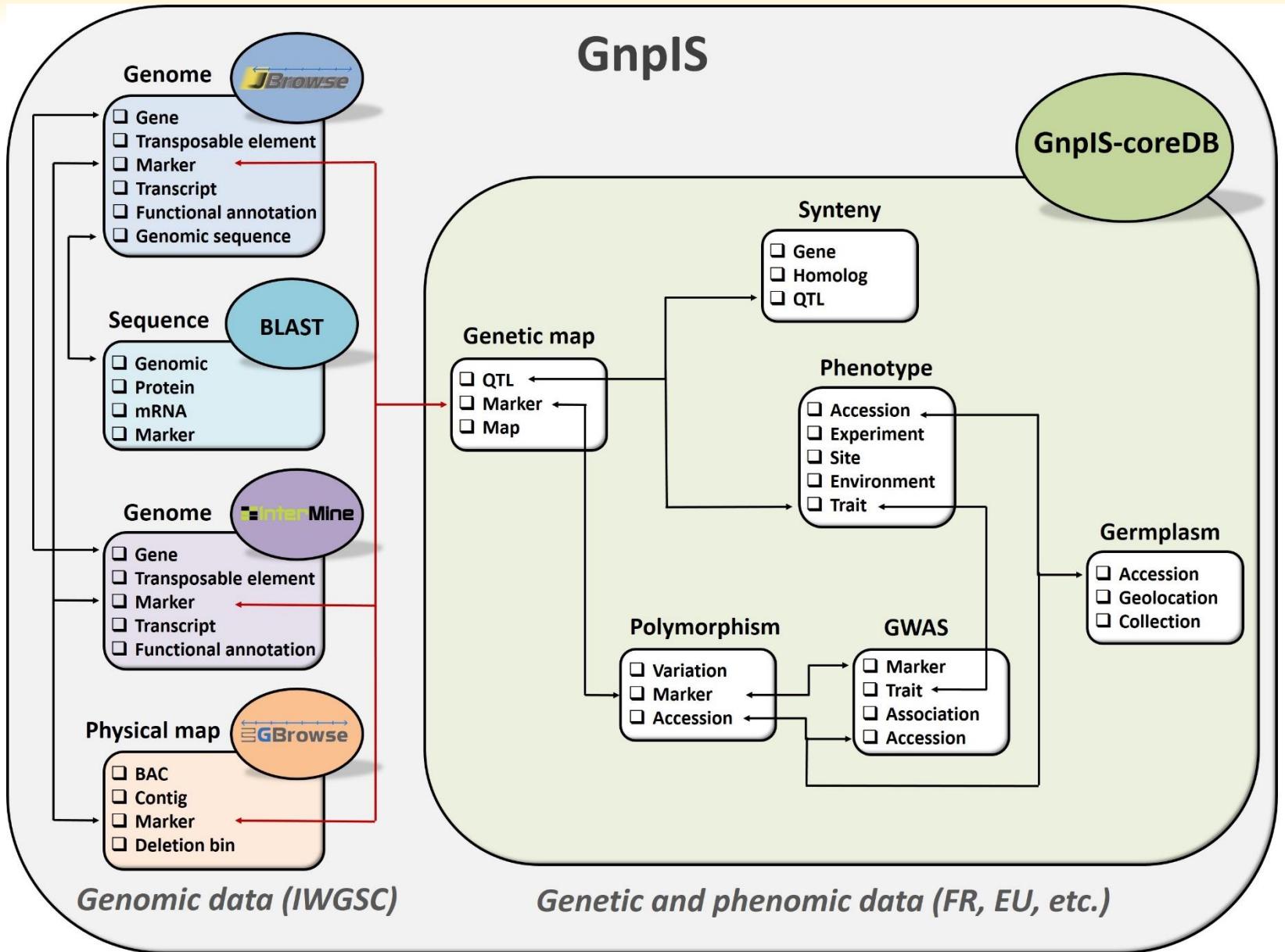


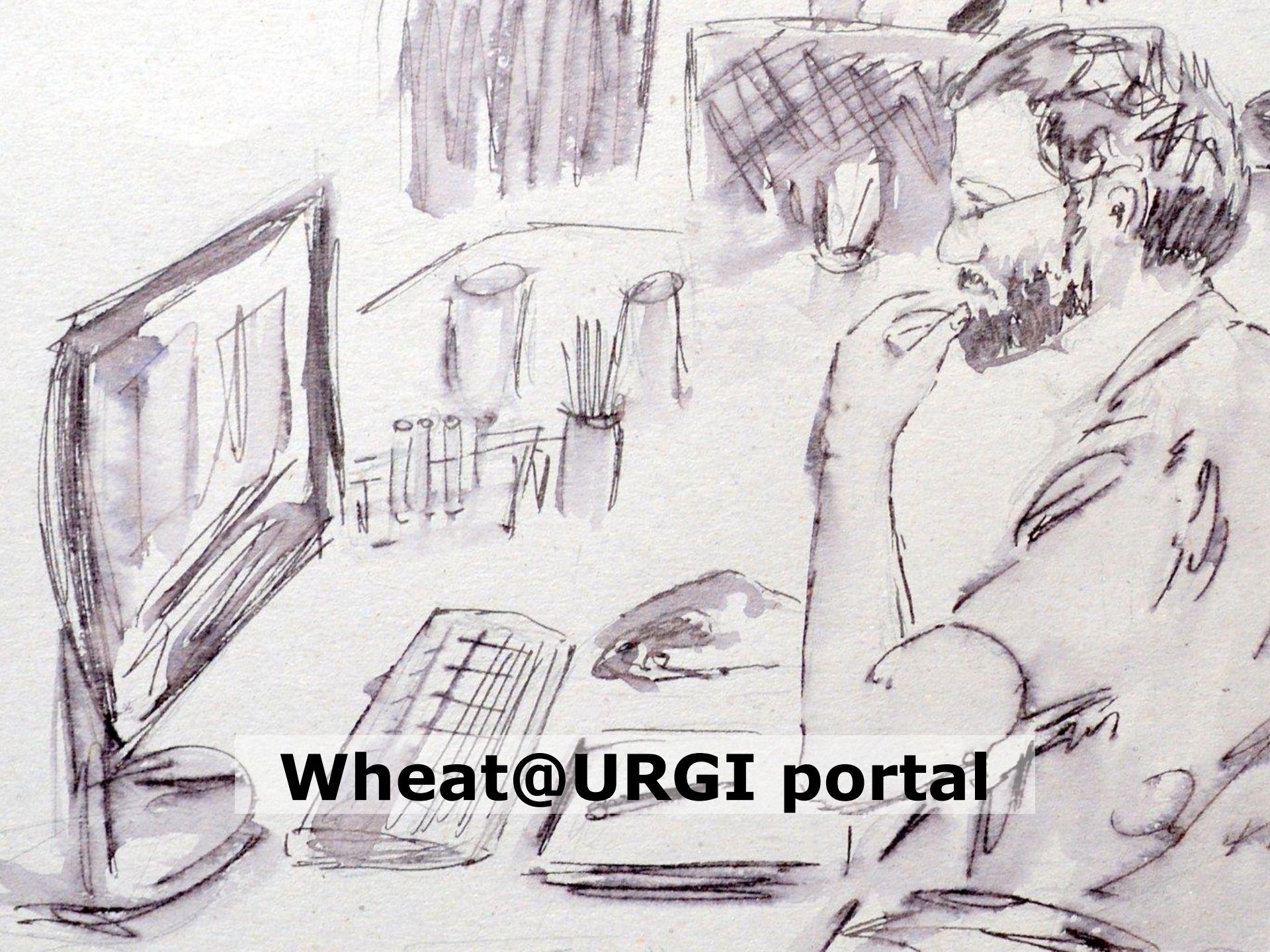
Integration using key data

Key data are interoperability « pivots »



Data integration in GnpIS





Wheat@URGI portal

Website

The image shows the homepage of the URGI website. At the top left is the URGI logo. The top right features links for FEEDBACK, CONTACT, SITE MAP, and REGISTER. A large, stylized yellow wheat ear graphic is positioned in the center-left. On the left side, there's a 'QUICK SEARCH' bar with the text 'Xwmc430' and a 'SUBMIT' button, followed by examples: Xwmc430, QTL, TaeCsp3B. Below it is an 'ADVANCED TOOLS' section with a 'WHEATMINE' button. The main menu at the top includes Projects, Data, Tools, Seq Repository, About us, and a 'WHAT'S NEW?' section with an RSS feed icon. The 'WHAT'S NEW?' section lists several news items with dates and titles, such as 'URGI talks at PAG 2019' (07 Jan 2019), 'GnpIS version 18.2 is available' (09 Nov 2018), and 'The Wheat Code is Finally Cracked' (17 Aug 2018). The bottom left shows an 'EVENTS & PUBLICATIONS' section with news items like 'Linking the International Wheat Genome Sequencing Consortium bread wheat reference ...' (23 Aug 2018) and a 'MORE... >' button. The bottom right corner contains the name Michael Alaux.

URGI

FEEDBACK | CONTACT | SITE MAP

REGISTER

Projects Data Tools Seq Repository About us

WHAT'S NEW ? RSS

QUICK SEARCH

Xwmc430 SUBMIT

Examples: [Xwmc430](#), [QTL](#), [TaeCsp3B](#)

ADVANCED TOOLS

[WHEATMINE](#)

Sequences

Physical maps

Genetic maps

Markers

QTLs , MetaQTLs

Germplasms

Phenotypes

SNPs

Synteny

07 Jan 2019
URGI talks at PAG 2019
• Linking the International Wheat Genome Sequencing Consortium Bread Wheat Reference ...

09 Nov 2018
GnpIS version 18.2 is available
GnpIS version 18.2 is availableMajor news are:
o All public genetic resources ...

03 Sep 2018
Data manager permanent position (for INRA staff only)
Dans le cadre de la bourse annuelle de mobilité, l'URGI recrute un(e) Ingénieur-e d'études ...

17 Aug 2018
The Wheat Code is Finally Cracked
IWGSC published the 17th of August in the international journal Science a detailed description ...

01 Aug 2018
GnpIS version 18.1 is available
GnpIS version 18.1 is availableMajor news are:
▪ Wheat IWGSC RefSeq v1.0 data are ...

05 Jul 2018
All IWGSC RefSeq resources publicly available
All IWGSC reference sequence resources are now publicly available at URGI. Look at the

Events & Publications RSS

07 Jan 2019
URGI talks at PAG 2019
• Linking the International Wheat Genome Sequencing Consortium Bread Wheat Reference ...

23 Aug 2018
Linking the International Wheat Genome Sequencing Consortium bread wheat reference ...
ACL (papers with reading committee)
Genome Biology , 2018, 19 (1) : 1-10.<https://doi.org/10.1186/s13002-018-0750-2>

[MORE... >](#)

<https://wheat-urgi.versailles.inra.fr/>

Michael Alaux

Tools - developed by URGI

IWGSC Data Repository

Webpages + File Download
including README, MD5



27k downloads in 2018

Index of /download/iwgsc/IWGSC_RefSeq_Annotations/v1.0

[ICO]	Name	Last modified	Size	Description
[DIR]	Parent Directory		-	
[]	iwgsc_refseqv1.0_FunctionalAnnotation_v1.zip	06-Dec-2017 14:49	19M	
[TXT]	iwgsc_refseqv1.0_FunctionalAnnotation_v1.zip.md5.txt	25-Jan-2018 16:36	81	
[]	iwgsc_refseqv1.0_HighConf_2017Mar13.gff3.zip	24-Mar-2017 17:26	11M	
[TXT]	iwgsc_refseqv1.0_HighConf_2017Mar13.gff3.zip.md5.txt	05-Dec-2017 15:01	79	
[]	iwgsc_refseqv1.0_HighConf_CDS_2017Mar13.fa.zip	24-Mar-2017 17:27	43M	
[TXT]	iwgsc_refseqv1.0_HighConf_CDS_2017Mar13.fa.zip.md5.txt	05-Dec-2017 15:01	81	
[]	iwgsc_refseqv1.0_HighConf_PROTEIN_2017Mar13.fa.zip	24-Mar-2017 17:27	26M	
[TXT]	iwgsc_refseqv1.0_HighConf_PROTEIN_2017Mar13.fa.zip.md5.txt	05-Dec-2017 15:01	85	
[]	iwgsc_refseqv1.0_HighConf_REPR_CDS_2017Apr03.fa.zip	11-Apr-2017 15:45	40M	
[TXT]	iwgsc_refseqv1.0_HighConf_REPR_CDS_2017Apr03.fa.zip.md5.txt	05-Dec-2017 15:01	86	
[]	iwgsc_refseqv1.0_HighConf_REPR_PROTEIN_2017Apr03.fa.zip	11-Apr-2017 15:45	25M	
[TXT]	iwgsc_refseqv1.0_HighConf_REPR_PROTEIN_2017Apr03.fa.zip.md5.txt	05-Dec-2017 15:01	90	
[]	iwgsc_refseqv1.0_HighConf_UTR_2017May05.gff3.zip	09-May-2017 09:53	18M	
[TXT]	iwgsc_refseqv1.0_HighConf_UTR_2017May05.gff3.zip.md5.txt	05-Dec-2017 15:01	83	
[]	iwgsc_refseqv1.0_ISBP.fa.zip	03-Jan-2018 18:06	209M	
[TXT]	iwgsc_refseqv1.0_ISBP.fa.zip.md5.txt	05-Dec-2017 15:01	64	
[]	iwgsc_refseqv1.0_LowConf_2017Mar13.gff3.zip	24-Mar-2017 17:27	7.3M	
[TXT]	iwgsc_refseqv1.0_LowConf_2017Mar13.gff3.zip.md5.txt	05-Dec-2017 15:01	78	
[]	iwgsc_refseqv1.0_LowConf_CDS_2017Mar13.fa.zip	24-Mar-2017 17:27	30M	
[TXT]	iwgsc_refseqv1.0_LowConf_CDS_2017Mar13.fa.zip.md5.txt	05-Dec-2017 15:01	80	
[]	iwgsc_refseqv1.0_LowConf_PROTEIN_2017Mar13.fa.zip	24-Mar-2017 17:27	18M	
[TXT]	iwgsc_refseqv1.0_LowConf_PROTEIN_2017Mar13.fa.zip.md5.txt	05-Dec-2017 15:01	84	
[]	iwgsc_refseqv1.0_LowConf_REPR_CDS_2017Apr03.fa.zip	11-Apr-2017 15:45	30M	
[TXT]	iwgsc_refseqv1.0_LowConf_REPR_CDS_2017Apr03.fa.zip.md5.txt	05-Dec-2017 15:01	85	
[]	iwgsc_refseqv1.0_LowConf_REPR_PROTEIN_2017Apr03.fa.zip	11-Apr-2017 15:45	18M	
[TXT]	iwgsc_refseqv1.0_LowConf_REPR_PROTEIN_2017Apr03.fa.zip.md5.txt	05-Dec-2017 15:01	89	
[]	iwgsc_refseqv1.0_LowConf_UTR_2017May05.gff3.zip	09-May-2017 09:53	10M	
[TXT]	iwgsc_refseqv1.0_LowConf_UTR_2017May05.gff3.zip.md5.txt	05-Dec-2017 15:01	82	
[]	iwgsc_refseqv1.0_Marker_mapping_summary_2017Mar13.zip	24-Mar-2017 17:27	39M	
[TXT]	iwgsc_refseqv1.0_Marker_mapping_summary_2017Mar13.zip.md5.txt	05-Dec-2017 15:01	88	

Tools - developed by URGI

GnpIS-coreDB

Information System to manage genetic and phenomic data



+ Synteny data



18k visits in 2018



Phenotypes

- GxE
- QTLs
- GWAS



Genetics

- Genetic maps
- Markers, Polymorphisms, Genotyping
- Genetic resources



Tools – instantiated by URGI

BLAST

IWGSC BLAST

URGI International Wheat Genome Sequencing Consortium

Home Databases About Help Contact us

BLAST parameter settings

Enter query sequences here in Fasta format

Or upload sequence fasta file (max 10MB)

Program: blastn Group: Wheat

currently selected database(s): IWGSC RefSeq v1.0 all chromosomes

Basic Search - using default BLAST parameters

Basic search Reset

Query	Databanks	Subject		Score	Identities (Query length)	Percentage	Expect	Start	End
Synth12	IWGSC RefSeq v1.0 chromosome 3B only	<input type="checkbox"/> chr3B		2805	1559/1560 (1560)	99	0.0	10949433	10950991
Synth12	IWGSC RefSeq v1.0 chromosome 4A only	<input type="checkbox"/> chr4A		848	630/731 (1560)	86	0.0	589960558	589961274
Synth12	IWGSC RefSeq v1.0 chromosome 7D only	<input type="checkbox"/> chr7D		845	633/732 (1560)	86	0.0	73268459	73269168
Synth12	IWGSC RefSeq v1.0 chromosome 5A only	<input type="checkbox"/> chr5A		836	625/731 (1560)	85	0.0	651285927	651286633
Synth12	IWGSC RefSeq v1.0 chromosome 3D only	<input type="checkbox"/> chr3D		816	631/752 (1560)	84	0.0	600648937	600649643
Synth12	IWGSC RefSeq v1.0 chromosome 7A only	<input type="checkbox"/> chr7A		814	634/744 (1560)	85	0.0	637182879	637182163
Synth12	IWGSC RefSeq v1.0 chromosome 5B only	<input type="checkbox"/> chr5B		809	623/735 (1560)	85	0.0	28361821	28361109
Synth12	IWGSC RefSeq v1.0 chromosome 4D only	<input type="checkbox"/> chr4D		809	629/749 (1560)	84	0.0	87367538	87366825
Synth12	IWGSC RefSeq v1.0 chromosome 2A only	<input type="checkbox"/> chr2A		809	622/730 (1560)	85	0.0	458531602	458530889
Synth12	IWGSC RefSeq v1.0 chromosome 1D only	<input type="checkbox"/> chr1D		798	631/750 (1560)	84	0.0	416979484	416978759
Synth12	IWGSC RefSeq v1.0 chromosome 5D only	<input type="checkbox"/> chr5D		796	603/706 (1560)	85	0.0	343115315	343115996
Synth12	IWGSC RefSeq v1.0 chromosome 3A only	<input type="checkbox"/> chr3A		796	605/710 (1560)	85	0.0	555399983	555399288



280k searches in 2018

Tools – instantiated by URGI

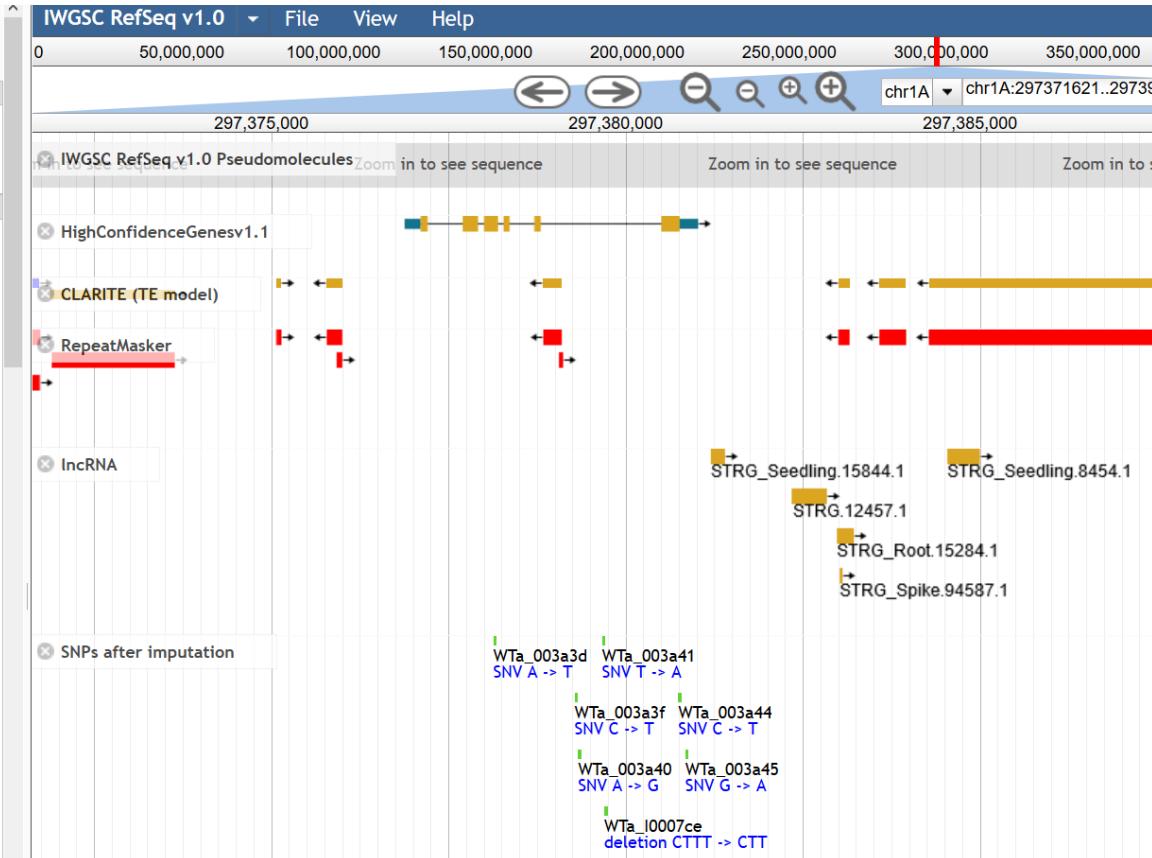
JBrowse



40k visits in 2018

Available Tracks

- filter by text
- IWGSC Reference Sequence**
 - IWGSC RefSeq v1.0 Pseudomolecules
 - IWGSC RefSeq v1.0 Superscaffolds
 - IWGSC RefSeq v1.0 Scaffolds
- IWGSC Official Annotation**
 - Gene Model**
 - Annotation v1.0**
 - HighConfidenceGenesv1.0
 - LowConfidenceGenesv1.0
 - Annotation v1.1**
 - HighConfidenceGenesv1.1
 - LowConfidenceGenesv1.1
 - Manually curated genes**
 - CRT-Binding Factors (CBF) gene family
 - Nucleotide-binding and Leucine-rich Repeat (NLR) gene family
 - Pentatricopeptide repeats (PPR) gene family
 - Prolamin and Globulin genes
 - Amino acid transporters
 - Cell Wall-Associated Kinase (WAK)
 - Transposable Elements**
 - CLARITE (TE model)
 - RepeatMasker
 - RNA**
 - lncRNA
 - miRNA



Tools – instantiated by URGI

InterMine

WheatMine Wheat IWGSC RefSeq v1.0

Home Templates Lists QueryBuilder

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

WheatMine contains data from the IWGSC F

Query for wheatmine content:

Gene : **TraesCS1A01G006900** *T. aestivum*

DB identifier TraesCS1A01G006900 Is Repr 1
Score 49.0 Score Type IWGSCv1.0_UTR

SHARE

Quick Links: Summary Genomics Other

Genome feature
Region: gene Length: 1954 Fasta...
Location: chr1A:3669097-3671050 reverse strand

Genome Browser
Click and drag the browser to move the view. Drag and drop tracks from left menu into the main panel to see the data. Clicking on individual features to open a report page for that feature.
* denotes SNPs that are mapped to multiple genome position. Centre on TraesCS1A01G006900

Available Tracks

- IWGSC Reference Sequence
- IWGSC RefSeq v1.0 Pseudomolecules
- IWGSC RefSeq v1.0 Superscaffolds
- IWGSC RefSeq v1.0 Scaffolds

- IWGSC Official Annotation
- Gene Model
- HighConfidenceGenesv1.0
- LowConfidenceGenesv1.0

- Transposable Elements
- CLARITE (TE model)
- RepeatMasker

IWGSC RefSeq v1.0 File View Help

50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 300,000,000 350,000,000 400,000,000 450,000,000 500,000,000 550,000,000

chr1A chr1A:3668902_3671245 (2.35 Kb) Go

3,669,000 3,669,500 3,670,000 3,670,500 3,671,000

miRNA

IWGSC RefSeq v1.0 Scaffolds

HighConfidenceGenesv1.0

LowConfidenceGenesv1.0

3 Data Sets

Name

High confidence genes

Functional annotation

Genes and markers sequences

15 Overlapping Features

Class	DB identifier	Length	Chromosome Location
NcRNA	STRG_Root.62029.2.path1	57855	chr1A: 3639864-3697718
NcRNA	STRG_Root.62029.2.mrna1	57855	chr1A: 3639864-3697718
NcRNA	STRG_Seedling.59.1.path1	2320	chr1A: 3668200-3670519
NcRNA	STRG_Seedling.59.1.mrna1	2320	chr1A: 3668200-3670519
NcRNA	STRG_Leaf.63.1.mrna1	2348	chr1A: 3668230-3670577
NcRNA	STRG_Leaf.63.1.path1	2348	chr1A: 3668230-3670577
NcRNA	STRG_Leaf.63.1.mrna1.exon2	76	chr1A: 3670502-3670577
NcRNA	STRG_Seedling.59.1.mrna1.exon2	18	chr1A: 3670502-3670519
Scaffold	chr1A_super3	3727638	chr1A: 1633103-5360740
Scaffold	scaffold121111	1501578	chr1A: 2943949-4445526

4 GOS

Identifier	Description
GO:0005515	MF: protein binding
GO:0005524	MF: ATP binding
GO:0004672	MF: protein kinase activity
GO:0006468	BP: protein phosphorylation

7 Interpros

Identifier	Description
IPR000719	Protein kinase domain
IPR011009	Protein kinase-like domain
IPR017441	Protein kinase, ATP binding site
IPR032675	Leucine-rich repeat domain, L domain-like
IPR016111	Leucine-rich repeat
IPR003591	Leucine-rich repeat, typical subtype
IPR008266	Tyrosine-protein kinase, active site

2 Pfams

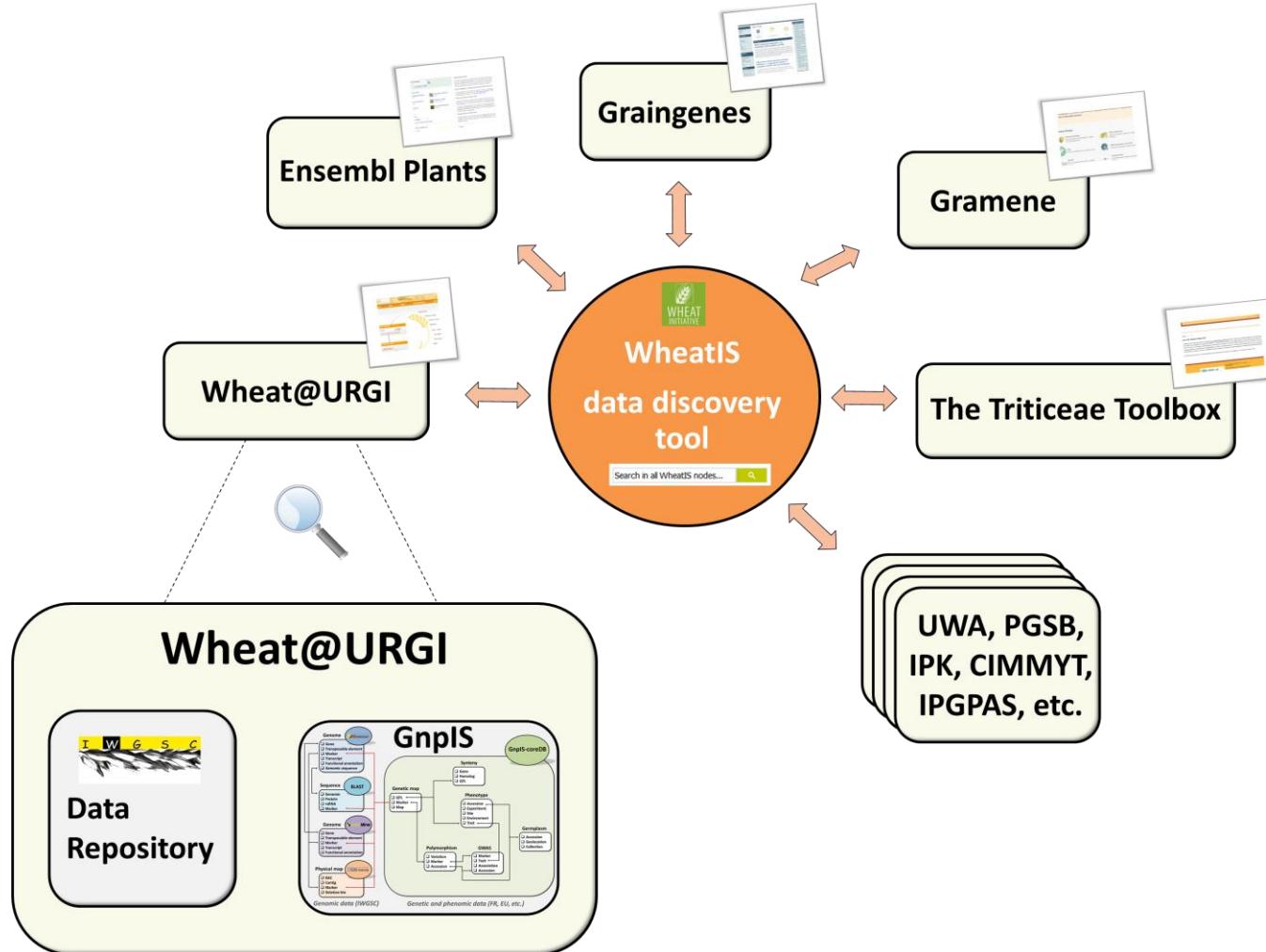
Identifier	Description
PF00069	Protein kinase domain
PF13855	Leucine rich repeat

4k visits in 2018



Tools – collaborative development

WheatIS data discovery



Tools – collaborative development

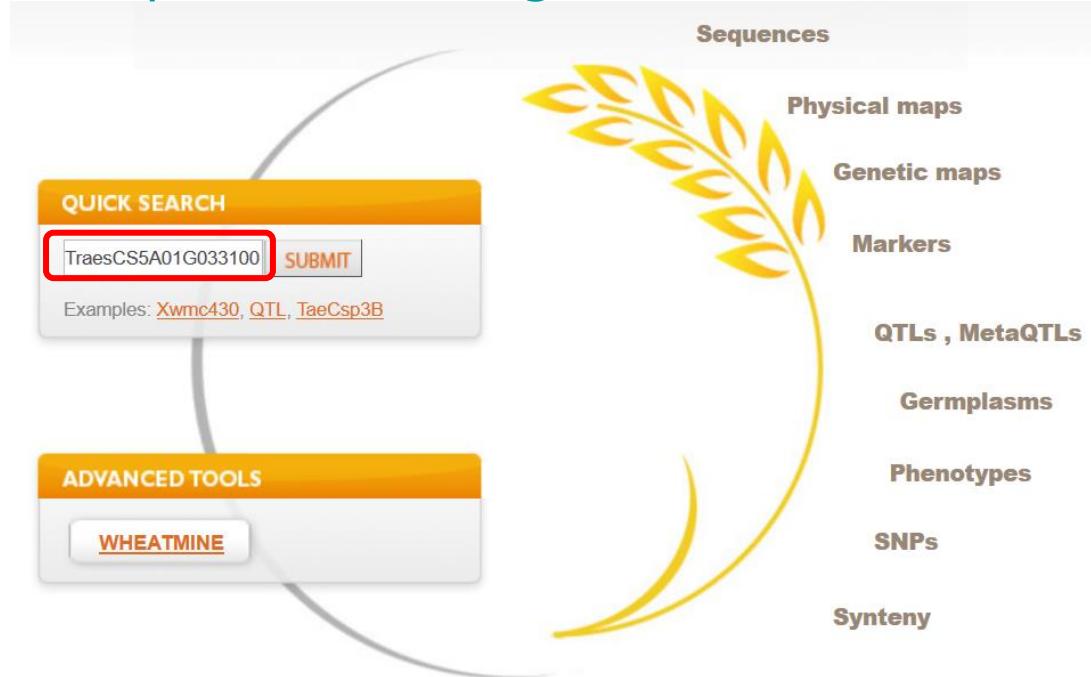
The screenshot shows the WheatIS search interface. The search term 'fhb' has been entered into the search bar, resulting in 1-10 of 583 entries. The results are displayed in a table with columns: Link to source, Source, Type, Taxon, and Description. The descriptions highlight various resistance genes and their associations with Fusarium head blight (FHB) in wheat. The interface includes a sidebar for 'How To Join' and sections for 'Database', 'Type', and 'Species' filters.

Link to source	Source	Type	Taxon	Description
10.1007/s10681-006-9153-0	OpenMinTeD	Bibliography	Triticum	Resistance gene analogs associated with Fusarium head blight resistance in wheat . Fusarium head blight (FHB) is one of the most destructive [...]
10.1111/j.1364-3703.2006.00349.x	OpenMinTeD	Bibliography	Triticum	Characterization of <i>Arabidopsis thaliana</i> - <i>Fusarium graminearum</i> interactions and identification of variation in resistance among ecotypes. <i>Fusarium graminearum</i> causes fusarium head blight (FHB) of wheat lines for resistance to Fusarium head blight . The major quantitative trait locus (QTL) on 3[...]
10.1139/G06-010	OpenMinTeD	Bibliography	Triticum	A resistance-like gene identified by EST mapping and its association with a QTL controlling Fusarium head blight infection on wheat chromosome 3BS. Fusarium head blight (FHB) is a ma[...]
10.1007/s00122-006-0249-7	OpenMinTeD	Bibliography	Triticum	Fine mapping Fhb1, a major gene controlling fusarium head blight resistance in bread wheat (<i>Triticum</i> [...])
10.1007/s10681-005-9047-6	OpenMinTeD	Bibliography	Triticum	Genetic diversity of resistance genes controlling fusarium head blight with simple sequence repeat markers in thirty-six wheat accessions from east asian origin. Fusarium head bligh[...]
10.1007/s00299-005-0059-4	OpenMinTeD	Bibliography	Triticum	A model wheat cultivar for transformation to improve resistance to Fusarium head blight . Fusarium head blight (FHB), caused primarily by Fu[...]
10.1270/jsbbs.56.25	OpenMinTeD	Bibliography	Triticum	QTL analysis of fusarium head blight severity in recombinant inbred population derived from a cross between two-rowed barley varieties. Fusarium head blight (FHB) severity was evalu[...]
10.1111/j.1439-0523.2006.01182.x	OpenMinTeD	Bibliography	Triticum	Validation of two major quantitative trait loci for fusarium head blight resistance in Chinese wheat line W1[...]
10.1007/s00122-005-0156-3	OpenMinTeD	Bibliography	Triticum	Mapping QTL associated with resistance to Fusarium head blight in the Nanda2419 x Wangshuibai population. II: Type I resistance Fusari[...]



Use case: From a gene to genetic studies

<https://wheat-urgi.versailles.inra.fr/>



WheatIS

Filters

Clear

Database

IWGSC@GNPIS (2)

Type

GENOME ANNOTATION (2)

Species

TRITICUM AESTIVUM (2)

Search

Link to source	Source	Type	Taxon	Description
chr5A:30214480..30215784	IWGSC@GnPlS	Genome annotation	Triticum aestivum	chr5A:30214480..30215784, Start = 30214480 , End = 30215784 , Strand = -1 , Source = IWGSCv1.0_UTR , Seq_id = chr5A , Name = TraesCS5A01G033100 , Score = 40 , Primconf = HC , Type = gene , Id = TraesCS5A01G033100 , Sub[...]
chr5A:19177589..19177898	IWGSC@GnPlS	Genome annotation	Triticum aestivum	chr5A:19177589..19177898, Start = 19177589 , End = 19177898 , Strand = -1 , Source = IWGSCv1.0_UTR , Seq_id = chr5A , Name = TraesCS5A01G033100LC , Score = 11 , Primconf = LC , Type = gene , Id = TraesCS5A01G033100LC , [...]

Use case: From a gene to genetic studies

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

IWGSC BLAST

International Wheat Genome Sequencing Consortium

Home Databanks About Help Contact us

BLAST parameter settings

Enter query sequences here in Fasta format:

```
>gene1
CAGGAGCGGTAGATCCATCGGTGGCCATCTCGTCGATATGGAGAAATTACCATGTAACGTATGCA
GACATTTAACATGCCACCTTTTGACTATGCGCTCATGTTTGAACCTTTGATACAGTATAAAG
AAAGAACTGATAATATCAGAAGATGATTCAAGAGATTGATGGACATATGCATAAAGTGAATGCC
AGTAGATTCAACTAGACTAGGTAGTATTTGGTGTAGTATGTAGTACCTCCGTCTCAAATGTAAGA
TAAAACCTCACACTAGACTAGGTAGTATTTGGTGTAGTATGTAGTACCTCCGTCTCAAATGTAAGA
CGTTTTGCGAGTTCAAAACCAACCGAGTAATATCAGGTCAGGGCAACCTGTGTCAATTAC
```

Or upload sequence fasta file (max 2M): [Parcourir...](#) Aucun fichier sélectionné.

Program: blastn Group: Wheat_IWGSC_RefSeq_v1_chromosomes Database(s):

- IWGSC RefSeq v1.0 all chromosomes
- IWGSC RefSeq v1.0 chromosome 1A only
- IWGSC RefSeq v1.0 chromosome 1B only
- IWGSC RefSeq v1.0 chromosome 1D only
- IWGSC RefSeq v1.0 chromosome 2A only
- IWGSC RefSeq v1.0 chromosome 2B only
- IWGSC RefSeq v1.0 chromosome 2D only
- IWGSC RefSeq v1.0 chromosome 3A only
- IWGSC RefSeq v1.0 chromosome 3B only
- IWGSC RefSeq v1.0 chromosome 3D only

currently selected database(s)

IWGSC RefSeq v1.0 all chromosomes

Basic Search - using default BLAST parameter settings

Query	Databanks	Subject	HSP	JBrowse	Score	Identities (Query length)	Percentage	Expect	Start	End
gene1	IWGSC RefSeq v1.0 chromosome 5A only	<input type="checkbox"/> chr5A 			2352	1304/1304 (1304)	100	0.0	30215784	30214481
gene1	IWGSC RefSeq v1.0 chromosome 4B only	<input type="checkbox"/> chr4B 			931	785/966 (1304)	81	0.0	659003290	659002331
gene1	IWGSC RefSeq v1.0 chromosome 4D only	<input type="checkbox"/> chr4D 			560	515/658 (1304)	78	4e-156	508643283	508642663
gene1	IWGSC RefSeq v1.0 chromosome 7B only	<input type="checkbox"/> chr7B 			241	254/329 (1304)	77	5e-60	59742901	59743224
gene1	IWGSC RefSeq v1.0 chromosome 7D only	<input type="checkbox"/> chr7D 			237	253/329 (1304)	77	6e-59	102538104	102538427
gene1	IWGSC RefSeq v1.0 chromosome 1B only	<input type="checkbox"/> chr1B 			232	309/422 (1304)	73	2e-57	594492563	594492150
gene1	IWGSC RefSeq v1.0 chromosome 7A only	<input type="checkbox"/> chr7A 			228	237/307 (1304)	77	3e-56	106679123	106679424
gene1	IWGSC RefSeq v1.0 chromosome 1A only	<input type="checkbox"/> chr1A 			223	312/430 (1304)	73	1e-54	535078923	535078502

Use case: From a gene to genetic studies

WheatIS

Filters

- Clear**
- Database**
 - IWGSC@GNPIS (2)
- Type**
 - GENOME ANNOTATION (2)
- Species**
 - TRITICUM AESTIVUM (2)

Search

TraesCS5A01G033100

1-2 of 2

Link to source	Source	Type	Taxon	Description
chr5A:30214480..30215784	IWGSC@GnpIS	Genome annotation	Triticum aestivum	chr5A:30214480..30215784 IWGSCv1.0_UTR.: HC , Type = gene , I
chr5A:19177589..19177898	IWGSC@GnpIS	Genome annotation	Triticum aestivum	chr5A:19177589..19177898 IWGSCv1.0_UTR.: = LC , Type = gene

Databanks

Query	Databanks	Subject	Score
gene1	IWGSC RefSeq v1.0 chromosome 5A only	<input type="checkbox"/> chr5A   	2352
gene1	IWGSC RefSeq v1.0 chromosome 4B only	<input type="checkbox"/> chr4B   	931
gene1	IWGSC RefSeq v1.0 chromosome 4D only	<input type="checkbox"/> chr4D   	560
gene1	IWGSC RefSeq v1.0 chromosome 7B only	<input type="checkbox"/> chr7B   	241
gene1	IWGSC RefSeq v1.0 chromosome 7D only	<input type="checkbox"/> chr7D   	237
gene1	IWGSC RefSeq v1.0 chromosome 1B only	<input type="checkbox"/> chr1B   	232
gene1	IWGSC RefSeq v1.0 chromosome 7A only	<input type="checkbox"/> chr7A   	228
gene1	IWGSC RefSeq v1.0 chromosome 1A only	<input type="checkbox"/> chr1A   	223

IWGSC RefSeq v1.0 Pseudomolecules

Zoom in to see sequence

Zoom

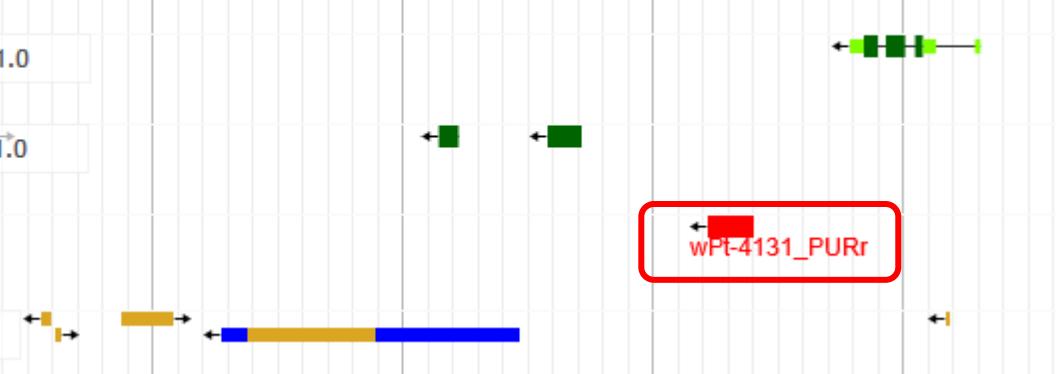
HighConfidenceGenesv1.0

LowConfidenceGenesv1.0

DArT public

CLARITE (TE model)

wPt-4131_PURr



Use case: From a gene to genetic studies

Marker: wPt4131

Association results

DETAILS

Marker	wPt4131
Taxon :	Wheat
Marker	Marker
Marker	Marker
Target :	Grain yield
Associated	Associated

Filters:

- Traits**: Dough quality: M-Consistency, Dough quality: M-Elasticity, Dough quality: M-Peak Time, Dough quality: M-Soft Break, Grain Hardness
- Treatments**: A
- Locations**: Chr 4B
- Years**: 2000-2008

ASSOCIATION RESULTS

Boxplot showing association results for traits A across chromosomes 0-4B. The plot shows significant associations (red dots) for traits like Dough quality: M-Consistency and Dough quality: M-Elasticity.

ASSOCIATION RESULTS

Table of associated measures:

Measure	Description	Experimentation
GY_RER-F7_MO_99	grain yield	RER-F7_MO_99
GY_cf9_Np	grain yield	cf9_Np
GY_ms9_Np	grain yield	ms9_Np
GY_ms9_NI	grain yield	ms9_NI
GY_sb9_NI	grain yield	sb9_NI
GY_nw9_Np	grain yield	nw9_Np
GY_nw9_NI	grain yield	nw9_NI
GY_ml8_Np	grain yield	ml8_Np
GY_cf8_Np	grain yield	cf8_Np
GY_RER-F7_RN_99	grain yield	RER-F7_RN_99
GY_RER-F7_LM_99	grain yield	RER-F7_LM_99
GY_RER-F7_CVIL_99	grain yield	RER-F7_CVIL_99
GY_RER-F7_CHAL_99	grain yield	RER-F7_CHAL_99
GY_RER-F7_CF_99	grain yield	RER-F7_CF_99

ASSOCIATED PHENOTYPES

See phenotypes for trait WIPO:0000217

CC_Qualité Wholemeal Flour color: ColFa-a*

Clermont-F

Phenotypes

Back to Form

DETAILS ON ITEM

Marker	cfn0655053
--------	------------

Trait card

TRAIT DETAILS

Trait name	grain_yield
Trait code	WIPO:0000217
Trait description	grain yield
Unit	q/t
Short remark	-
Theme	Quality
Genoplante classification	Agronomic
MetaQTLs	-

ASSOCIATED MEASURES

Number of measures: 14

Measure	Description	Experimentation
GY_RER-F7_MO_99	grain yield	RER-F7_MO_99
GY_cf9_Np	grain yield	cf9_Np
GY_ms9_Np	grain yield	ms9_Np
GY_ms9_NI	grain yield	ms9_NI
GY_sb9_NI	grain yield	sb9_NI
GY_nw9_Np	grain yield	nw9_Np
GY_nw9_NI	grain yield	nw9_NI
GY_ml8_Np	grain yield	ml8_Np
GY_cf8_Np	grain yield	cf8_Np
GY_RER-F7_RN_99	grain yield	RER-F7_RN_99
GY_RER-F7_LM_99	grain yield	RER-F7_LM_99
GY_RER-F7_CVIL_99	grain yield	RER-F7_CVIL_99
GY_RER-F7_CHAL_99	grain yield	RER-F7_CHAL_99
GY_RER-F7_CF_99	grain yield	RER-F7_CF_99

ASSOCIATED PHENOTYPES

See phenotypes for trait WIPO:0000217

CC_Qualité Wholemeal Flour color: ColFa-a*

Clermont-F

Geolocation

DATA SETS:

- Network Data Set: INRA Wheat Network BRC accession (A series)
- Network Data Set: INRA Small Grain Cereals Network
- Network Data Set: Leaflet | Map data © OpenStreetMap contributors, CC-BY-SA, Imagery © Mapbox

Phenotyping campaign(s): 2000-2008, 2009-2015

Trial list / Phenotypic data

LEVEL: TRIAL

GENOTYPE ID	Accession Number	Accession Name	TREATMENT	Trial Name	Trial Site
EM99006	37403	EM99006	treated	BTH Colmar 2000 SetA	Colmar
EM00002	23832	EM00002	low inputs	BTH Le Moulon 2000 SetB1	Le Moulon
CF00189	20362	CF00189	treated	BTH Le Moulon 2000 SetB1	Le Moulon
RE99123	29194	RE99123	low inputs	BTH Estrees-Mons 2000 SetB2	Estrees-Mons
CF99306	CF99306	CF99306	treated	BTH Clermont-Ferrand 2000 SetB3	Clermont-Ferrand
CF99270	CF99270	CF99270	treated	BTH Estrees-Mons 2000 SetB3	Estrees-Mons
Tremie	7043	TREMIE	treated	BTH Lusignan 2000 SetB3	Lusignan
RE99134	RE99134	RE99134	treated	BTH Lusignan 2000 SetB3	Lusignan
EM99017	37776	EM99017	low inputs	BTH Le Moulon 2001 SetA1	Le Moulon
EM99003	29160	EM99003	treated	BTH Orgeval 2001 SetA1	Orgeval

Ephesis data export **Ephesis MIAPPE ISA-Tab export**

Acknowledgements



Alaux M.
Letellier T.
Flores R.
Alfama F.
Pommier C.
Mohellibi N.
Durand S.
Kimmel E.
Michotey C.
Guerche C.
Loaec M.
Lainé M.
Steinbach D.
Adam-Blondon A.F.
Quesneville H.

Jamilloux V., Amselem J.,
Charruaud D., Cornut G., Burlot
L., Philippe F., Francillonne N.,
Couderc L., Verdelet D., Brault
B., Chennen K.



Rogers J.
Eversole K.

Caugant I.



Choulet F.
Rimbert H.
Leroy P.
Guilhot N.
Salse J.
Feuillet C.
Paux E.

Le Gouis J.
Charmet G.
Balfourier F.
Sourdille P.
Ravel C.
Oury F.X.
Didier A.



Dzale E.
Aubin S.
Hologne O.

Michael Alaux



Acknowledgements



All data submitters



IWGSC Coordinating Committee
IWGSC Sequencing team
IWGSC Analysis team



WheatIS Expert Working Group



RESEARCH DATA ALLIANCE

RDA Wheat Data Interoperability Group



Michael Alaux

Questions

Alaux et al., Genome Biology 2018

<https://doi.org/10.1186/s13059-018-1491-4>



Linking the International Wheat Genome Sequencin...

The Wheat@URGI portal has been developed to provide the international community of researchers and breeders with access to the bread wheat reference genome se...

genomebiology.biomedcentral.com

Wheat@URGI portal

<https://wheat-urgi.versailles.inra.fr/>



Contact me at michael.alaux@inra.fr



@michaelalaux