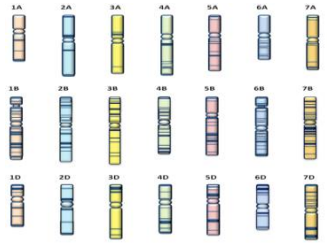
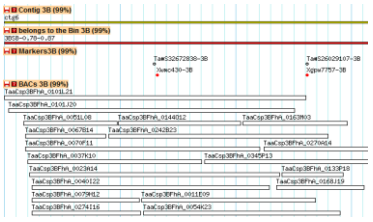
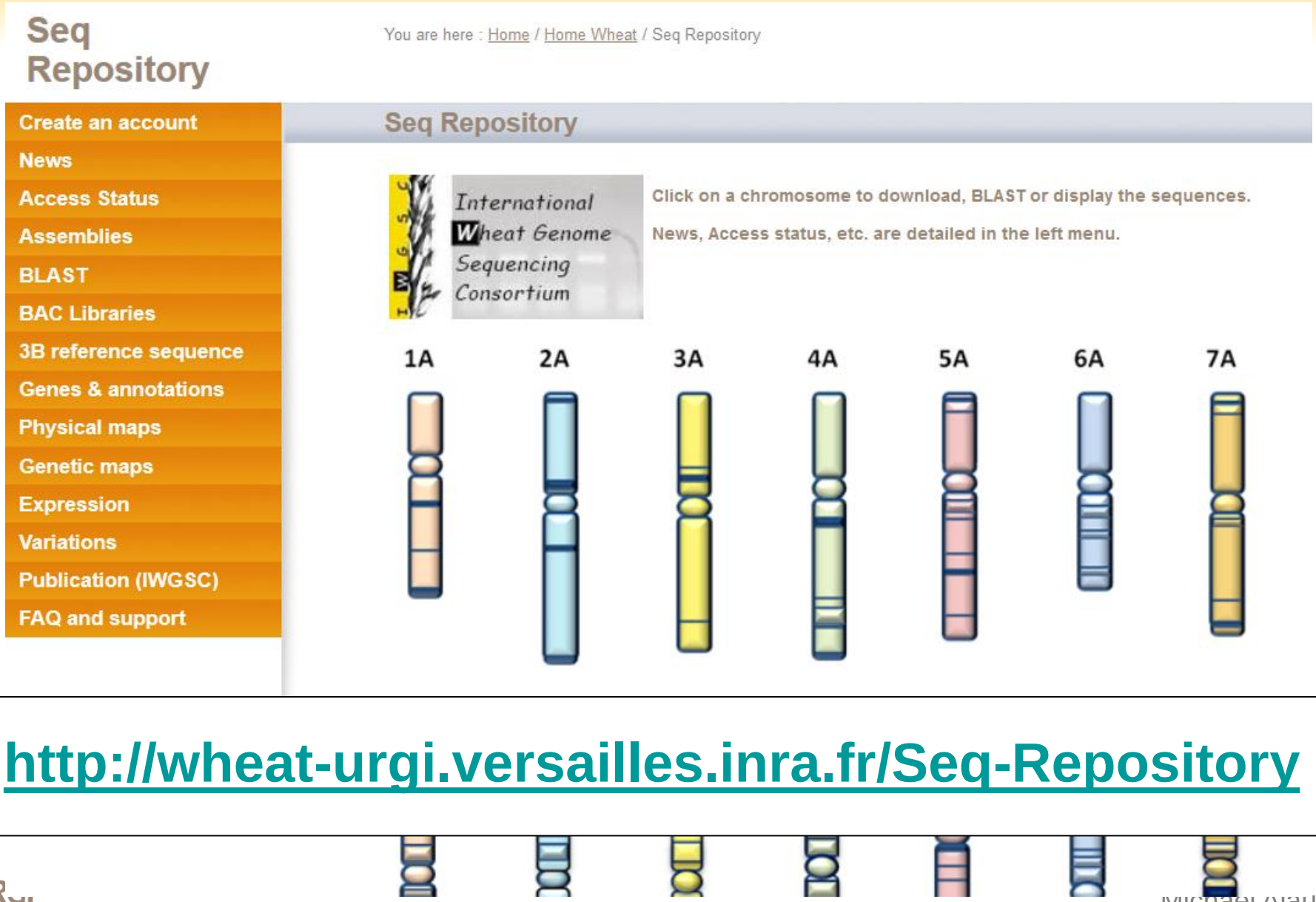
The IWGSC Reference Genome Browser, Data Mining and Beyond

Michael Alaux





IWGSC Sequence Repository



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

New data available

Major News in 2016

- Create an account
- News
- Access Status
- Assemblies
- BLAST
- BAC Libraries
- Genes & annotations
- Physical maps
- Genetic maps
- 3B reference sequence
- Expression
- Variations
- Publication (IWGSC)
- FAQ and support

Jan 2017:

- The **IWGSC RefSeq v1.0** is now available for download and BLAST (under IWGSC General data access agreement).

June 2016:

- The IWGSC WGA v0.4 (Illumina short sequence reads assembled with NRGene's DeNovoMAGIC^{TM} software) is available to BLAST and download (under IWGSC General data access agreement).

May 2016:

- 3AL, 3AS v2 and 3DS v2 physical maps are available to display (under IWGSC Project data access agreement).

April 2016:

- TGACv1 assembly is publicly available for download at EnsemblPlants and in our BLAST .

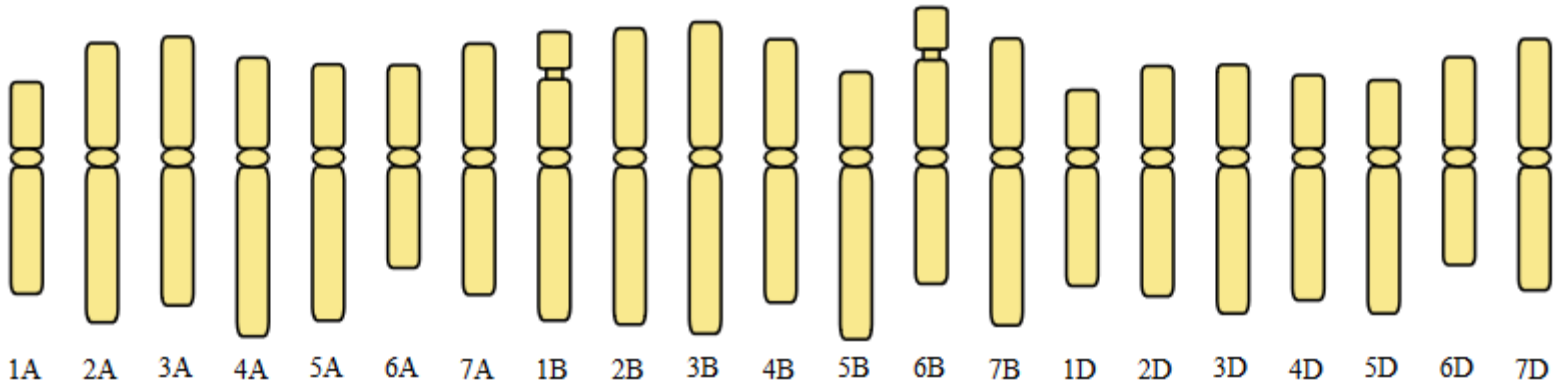
Jan 2016:

- IWGSC survey sequence version 3 assembly (*A. Sharpe, D. Konkin and C. Pozniak*) is available for download and in a browser (under IWGSC General data access agreement).

Physical maps

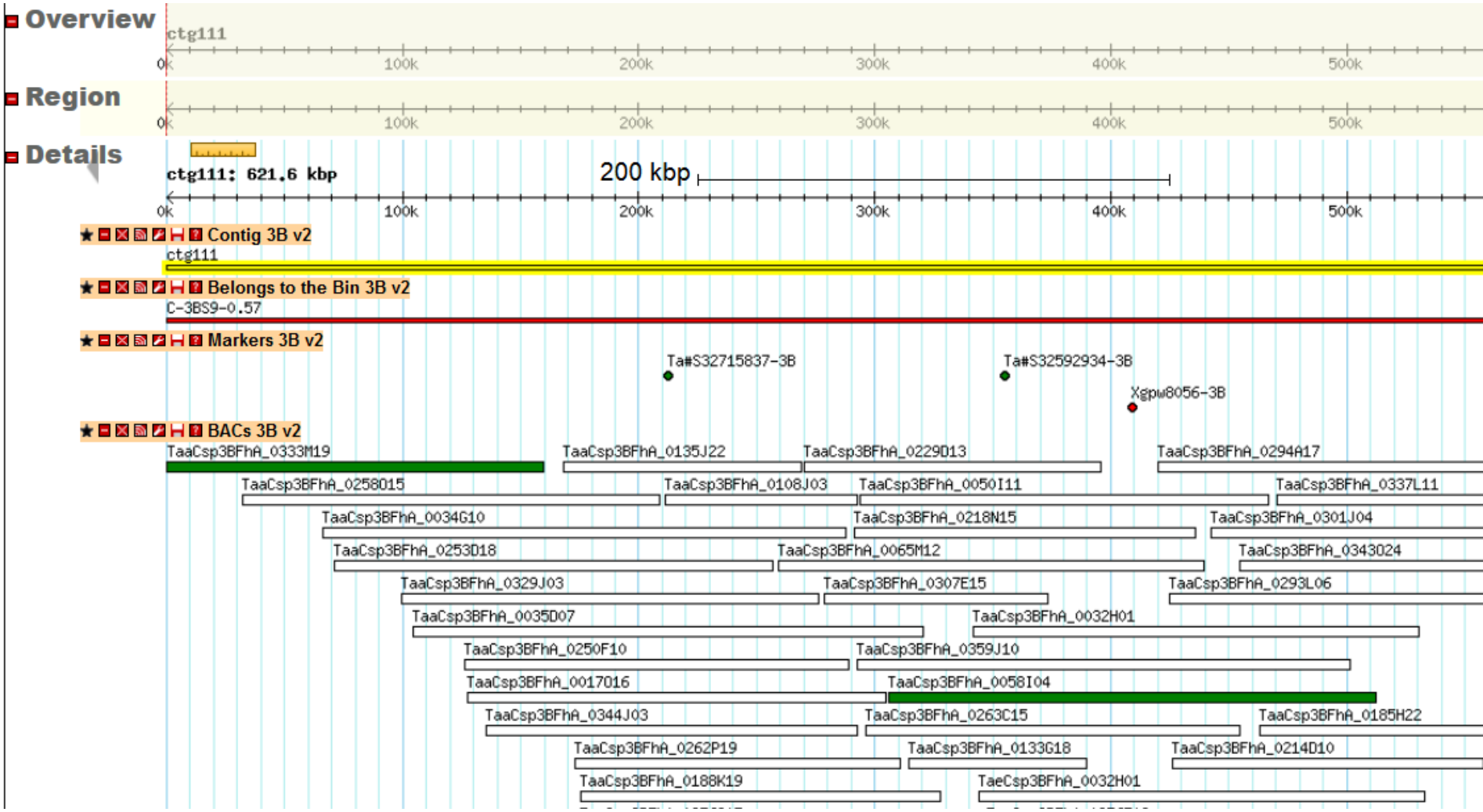
- 3AL added
- New versions of 3AS, 3DL and 4AL

→ All the physical maps are available to display



https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_pub/

Physical maps



https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_pub/

New Assemblies

- **IWGSC Survey sequence chromosomes**

  - **Version 1** and **version 2** assemblies are **publicly available** for download , BLAST and in a browser .

Version 2 assembly is the version 1 assembly cleaned i.e. from which duplicates were removed. Fasta by A- B- and D-genomes are available for download at MIPS.
Summary of the different CSS assemblies and versions (TGAC):
IWGSC-CSS_assembly-version-overview_Sep2014.xls 30.50 kB

  - **Version 3** assembly is available for download , in a browser and BLAST (access under the IWGSC general access agreement) .

This new version of the IWGSC CSS wheat survey sequence has been generated by the incorporation of ca. 185 Gbp of mate pair sequence data produced from libraries ranging in size from 1-40kb from a Chinese Spring + 7EL addition line. The assembly has been produced by A. Sharpe, D. Konkin and C. Pozniak, at the National Research Council Canada and the U. of Saskatchewan, Canada.

New Assemblies

- **TGACv1** whole genome assembly is publicly available for download at EnsemblPlants and in our BLAST .

the first release of the TGACv1 genome assembly of *Triticum aestivum* cv. Chinese Spring, generated by The Genome Analysis Centre, Norwich , as part of the BBSRC-funded project, Triticeae Genomics for Sustainable Agriculture. The assembly has an N50 of 88 Kb and a total length of 13.4 Gb in contigs greater than 500 bp. A total of 98,974 genes (99% of the total) annotated on the previously released assembly have been located on the new assembly. Alignments of RNA-seq data from 3 different studies across 18 samples have additionally been located on the new assembly.

New Assemblies

- The **IWGSC WGA v0.4** (Whole Genome Assembly), comprised of Illumina short sequence reads assembled with NRGene's DeNovoMAGIC^{TM} software, produced scaffolds totaling 14.5Gb with a L50 of 7.1Mb that have been assigned to chromosomal locations using POPSEQ data and a HiC map.

Over 99% of chromosome survey contigs map to the **IWGSC WGA v0.4** assembly. We hope that the more contiguous sequences of the new assembly will help users accelerate the identification of genes associated with important traits. The data are being made available before publication in accordance with the Toronto Agreement under which the IWGSC reserves the right to publish the first global analyses of the data. This includes descriptions of whole chromosome or genome-level analyses of genes, gene families, repetitive elements, and comparisons with other organisms.
Project leaders interested in contributing to these analyses are encouraged to contact the executive director of the IWGSC, Kellye Eversole (eversole@eversoleassociates.com) to discuss potential collaboration.
A BLAST server has been set up to facilitate rapid access to single or small numbers of queries.The data can also be downloaded from the URGI IWGSC repository.

How to access the IWGSC WGA v0.4 data? Access does require registration and agreeing to respect the right of the IWGSC to publish first. For specific access terms, see the IWGSC General Data Access agreement .

- Individuals who have not sign the IWGSC Data Access Agreement should **FIRST** register on the IWGSC website and sign the Agreement; URGI login details will be provided subsequently for access to the data.

- Individuals who have already signed the IWGSC Data Access Agreement can go directly to the URGI website to access the data using their URGI login details.

New Assemblies

- The **IWGSC RefSeq v1.0** (the first version of the reference sequence of the bread wheat variety *Chinese Spring*) is now available for download and BLAST .

The pre-publication data are being made available under the IWGSC General Data Access Agreement which is consistent with the Toronto Agreement and that grants the IWGSC the right to publish the first global analyses of the data. This includes descriptions of whole chromosome or genome-level analyses of genes, gene families, repetitive elements, and comparisons with other organisms.

The **IWGSC RefSeq v1.0** is an integration of the IWGSC WGA v0.4 – made available in June 2016 – with IWGSC chromosome-based and other resources, including but not limited to:

- Physical maps for all chromosomes;
- Sequenced BACs for 8 chromosomes (1A, 1B, 3B, 3D, 6B, 7A, 7B, 7D) and partial MTP BAC sequences for 2 chromosome arms (4AL, 5BS);
- MTP BAC WGP^{TM} sequence tags for all chromosomes, except 3B;
- BioNano optical maps (7A, 7B, 7DS);
- Alignment to RH maps (D chromosomes); and
- GBS map of the SynOp RIL population CsxRn genetic map (INRA).

With the addition of the resources that have been developed by IWGSC members over the past few years, the quality of the assembly increased substantially. When compared with IWGSC WGA v0.4, the chromosomal scaffold/ superscaffold N50 increased from 7.0 Mb to 22.8 Mb.

The data are available for BLAST searches and can be downloaded .

Access to the IWGSC RefSeq v1.0

How to access IWGSC RefSeq v1.0 data?

Access does require registration and agreeing to respect the right of the IWGSC to publish first. For specific access terms, see the IWGSC General Data Access agreement .

- Individuals who have not signed the IWGSC Data Access Agreement should **FIRST** register on the IWGSC website and sign the Agreement; URGI login details will be provided subsequently by email for access to the data. Typically, this it will take no more than 2 business days for your URGI account to be established but occasionally it make take up to a week.

- Individuals who have already signed the IWGSC Data Access Agreement can go directly to the URGI website to access the data using their URGI login details.

http://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies

Tools to browse and mine the IWGSC RefSeq v1.0

Michael Alaux

Download

v0.4

[ ]	ALL_DATA_Wheat_IWGSC_WGA_v0.4_All_pseudomolecules.tar.gz	27-May-2016 10:07	12G
[TXT]	ALL_DATA_Wheat_IWGSC_WGA_v0.4_All_pseudomolecules.tar.gz.md5.txt	04-Jul-2016 11:47	112
[ ]	Wheat_IWGSC_WGA_v0.4_1A_pseudomolecule_only.tar.gz	19-May-2016 14:26	169M
[ ]	Wheat_IWGSC_WGA_v0.4_1B_pseudomolecule_only.tar.gz	19-May-2016 14:30	197M
[ ]	Wheat_IWGSC_WGA_v0.4_1D_pseudomolecule_only.tar.gz	19-May-2016 14:31	140M
[ ]	Wheat_IWGSC_WGA_v0.4_2A_pseudomolecule_only.tar.gz	19-May-2016 14:33	223M
[ ]	Wheat_IWGSC_WGA_v0.4_2B_pseudomolecule_only.tar.gz	19-May-2016 14:33	229M
[ ]	Wheat_IWGSC_WGA_v0.4_2D_pseudomolecule_only.tar.gz	19-May-2016 14:33	185M
[ ]	Wheat_IWGSC_WGA_v0.4_3A_pseudomolecule_only.tar.gz	19-May-2016 14:34	214M
[ ]	Wheat_IWGSC_WGA_v0.4_3B_pseudomolecule_only.tar.gz	19-May-2016 14:37	237M
[ ]	Wheat_IWGSC_WGA_v0.4_3D_pseudomolecule_only.tar.gz	19-May-2016 14:36	175M
[ ]	Wheat_IWGSC_WGA_v0.4_4A_pseudomolecule_only.tar.gz	19-May-2016 14:37	212M
[ ]	Wheat_IWGSC_WGA_v0.4_4B_pseudomolecule_only.tar.gz	19-May-2016 14:37	192M
[ ]	Wheat_IWGSC_WGA_v0.4_4D_pseudomolecule_only.tar.gz	19-May-2016 14:38	144M
[ ]	Wheat_IWGSC_WGA_v0.4_5A_pseudomolecule_only.tar.gz	19-May-2016 14:39	202M
[ ]	Wheat_IWGSC_WGA_v0.4_5B_pseudomolecule_only.tar.gz	19-May-2016 14:40	203M
[ ]	Wheat_IWGSC_WGA_v0.4_5D_pseudomolecule_only.tar.gz	19-May-2016 14:40	160M
[ ]	Wheat_IWGSC_WGA_v0.4_6A_pseudomolecule_only.tar.gz	19-May-2016 14:43	176M
[ ]	Wheat_IWGSC_WGA_v0.4_6B_pseudomolecule_only.tar.gz	19-May-2016 14:44	205M
[ ]	Wheat_IWGSC_WGA_v0.4_6D_pseudomolecule_only.tar.gz	19-May-2016 14:43	134M

Download

v1.0

[ ]	Wheat_IWGSC_WGA_v1.0_pseudomolecules.zip	02-Dec-2016 15:12	8.0G
[TXT]	Wheat_IWGSC_WGA_v1.0_pseudomolecules.zip.md5.txt	02-Dec-2016 15:15	96

- MD5 checksum
- We will add the possibility to download individual chromosomes.

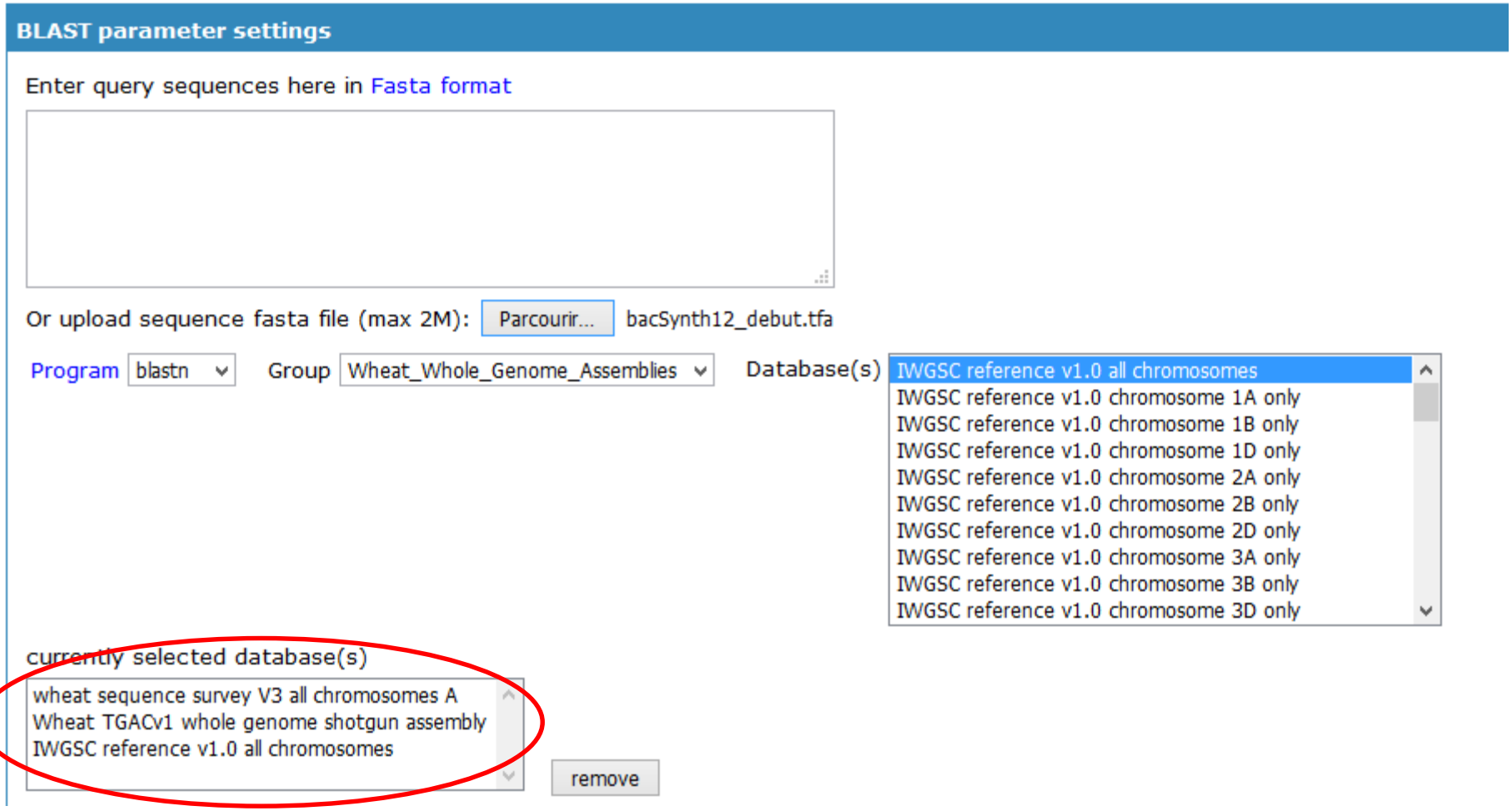
BLAST

- BLAST dedicated to IWGSC general access agreement:

https://urgi.versailles.inra.fr/blast_iwgsc/?dbgroup=wheat_whole_genome_assemblies&program=blastn

- Allow to BLAST all the available assemblies in one time:
  - survey sequence v3
  - TGAC v1
  - **IWGSC RefSeq v1.0**

BLAST


BLAST parameter settings
Enter query sequences here in Fasta format
Or upload sequence fasta file (max 2M):
Parcourir...
bacSynth12_debut.tfa
Program
blastn
Group
Wheat_Whole_Genome_Assemblies
Database(s)
IWGSC reference v1.0 all chromosomes
IWGSC reference v1.0 chromosome 1A only
IWGSC reference v1.0 chromosome 1B only
IWGSC reference v1.0 chromosome 1D only
IWGSC reference v1.0 chromosome 2A only
IWGSC reference v1.0 chromosome 2B only
IWGSC reference v1.0 chromosome 2D only
IWGSC reference v1.0 chromosome 3A only
IWGSC reference v1.0 chromosome 3B only
IWGSC reference v1.0 chromosome 3D only
currently selected database(s)
wheat sequence survey V3 all chromosomes A
Wheat TGACv1 whole genome shotgun assembly
IWGSC reference v1.0 all chromosomes
remove

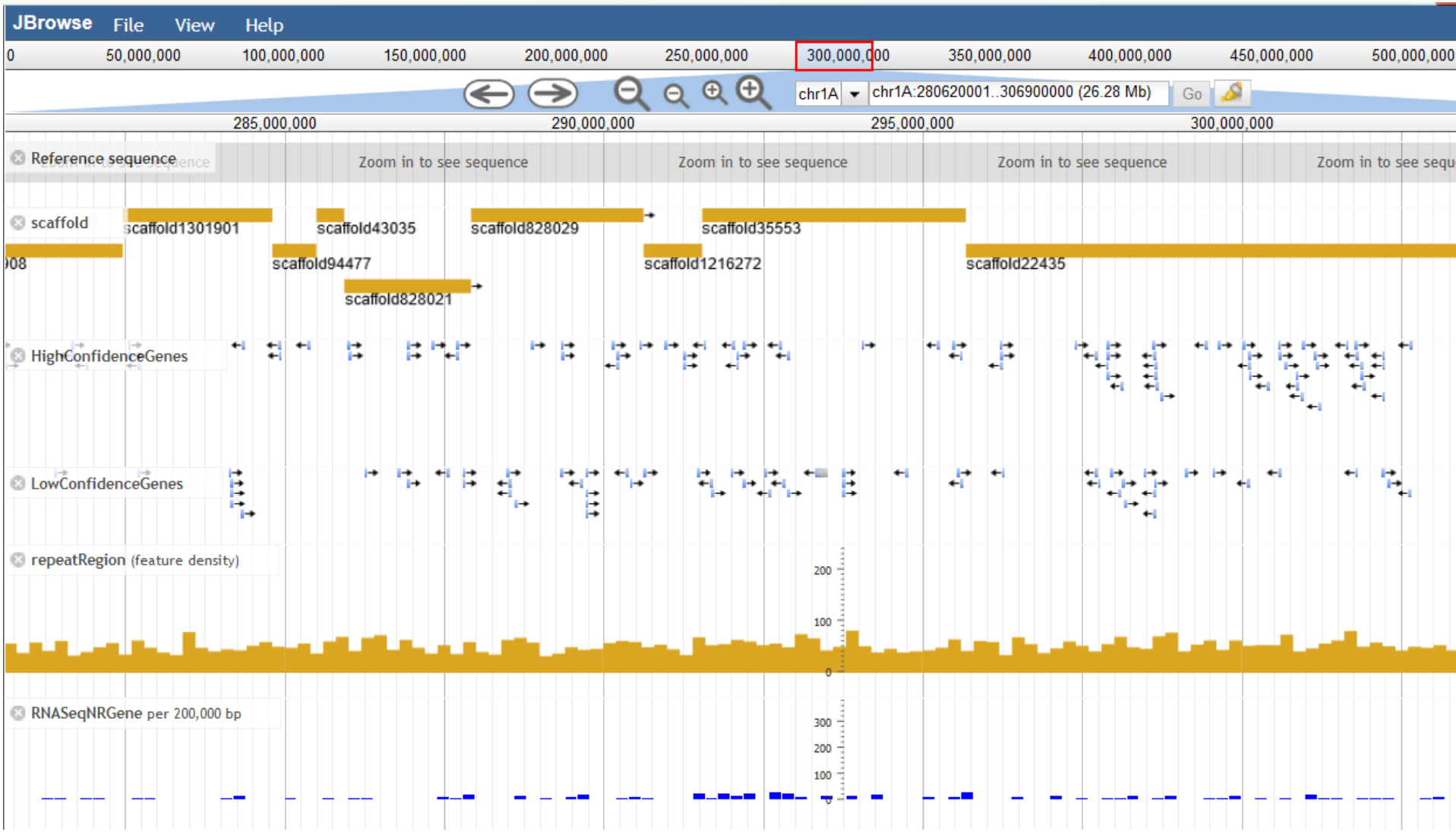
BLAST

Query	Databanks	Subject	Score	Identities (Query length)	Percentage	Expect	Start	
			BEST HSP					
Synth12	IWGSC reference v1.0 chromosome 1A only	chr1A	778	618/740 (1560)	84	0.0	559181689	55
Synth12	IWGSC reference v1.0 chromosome 1D only	chr1D	798	631/750 (1560)	84	0.0	416979484	41
Synth12	IWGSC reference v1.0 chromosome 2A only	chr2A	809	622/730 (1560)	85	0.0	458531602	45
Synth12	IWGSC reference v1.0 chromosome 3A only	chr3A	796	605/710 (1560)	85	0.0	555399983	55
Synth12	IWGSC reference v1.0 chromosome 3B only	chr3B	2805	1559/1560 (1560)	99	0.0	10949433	1
Synth12	IWGSC reference v1.0 chromosome 3D only	chr3D	816	631/752 (1560)	84	0.0	600648937	60
Synth12	IWGSC reference v1.0 chromosome 4A only	chr4A	848	630/731 (1560)	86	0.0	589960558	58
Synth12	IWGSC reference v1.0 chromosome 4D only	chr4D	809	629/749 (1560)	84	0.0	87367538	87
Synth12	IWGSC reference v1.0 chromosome 5A only	chr5A	836	625/731 (1560)	85	0.0	651285927	65
Synth12	IWGSC reference v1.0 chromosome 5B only	chr5B	809	623/735 (1560)	85	0.0	28361821	28
Synth12	IWGSC reference v1.0 chromosome 5D only	chr5D	796	603/706 (1560)	85	0.0	343115315	34
Synth12	IWGSC reference v1.0 chromosome 6A only	chr6A	791	593/693 (1560)	86	0.0	147772773	14
Synth12	IWGSC reference v1.0 chromosome 7A only	chr7A	814	634/744 (1560)	85	0.0	637182879	63
Synth12	IWGSC reference v1.0 chromosome 7B only	chr7B	782	553/634 (1560)	87	0.0	609412936	60
Synth12	IWGSC reference v1.0 chromosome 7D only	chr7D	845	633/732 (1560)	86	0.0	73268459	73
Synth12	Wheat TGACv1 whole genome shotgun assembly	TGACv1_scaffold_221301_3B	2805	1559/1560 (1560)	99	0.0	142977	1
Synth12	Wheat TGACv1 whole genome shotgun assembly	TGACv1_scaffold_290129_4AL	848	630/731 (1560)	86	0.0	1519	
Synth12	Wheat TGACv1 whole genome shotgun assembly	TGACv1_scaffold_210993_3AS	845	633/732 (1560)	86	0.0	103865	1
Synth12	Wheat TGACv1 whole genome shotgun assembly	TGACv1_scaffold_374761_5AL	836	625/731 (1560)	85	0.0	69253	

JBrowse

- JBrowse in development to display all the annotation tracks on genes, TEs, evidences.

JBrowse

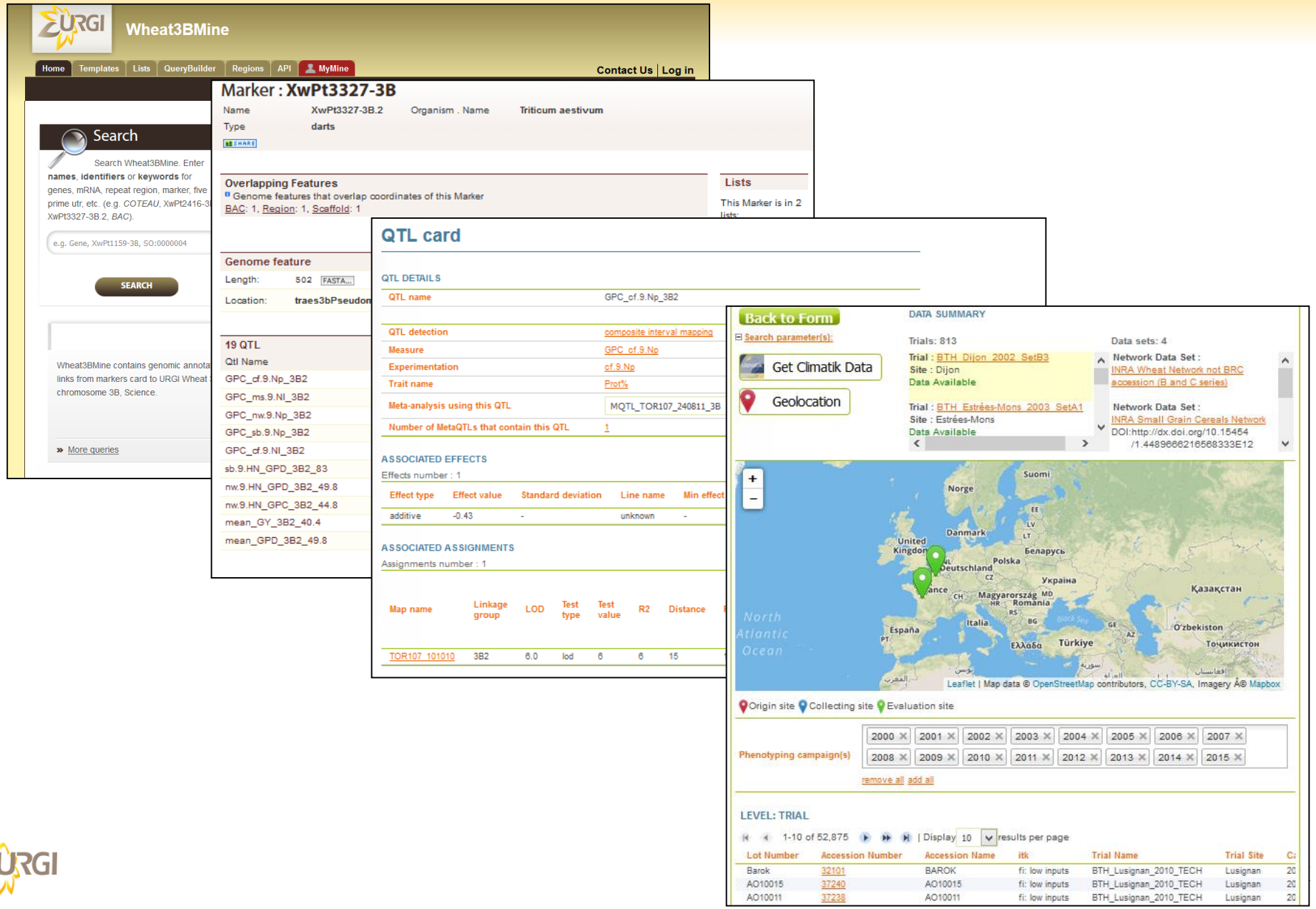


https://urgi.versailles.inra.fr/jbrowse/gmod_jbrowse/

WheatMine

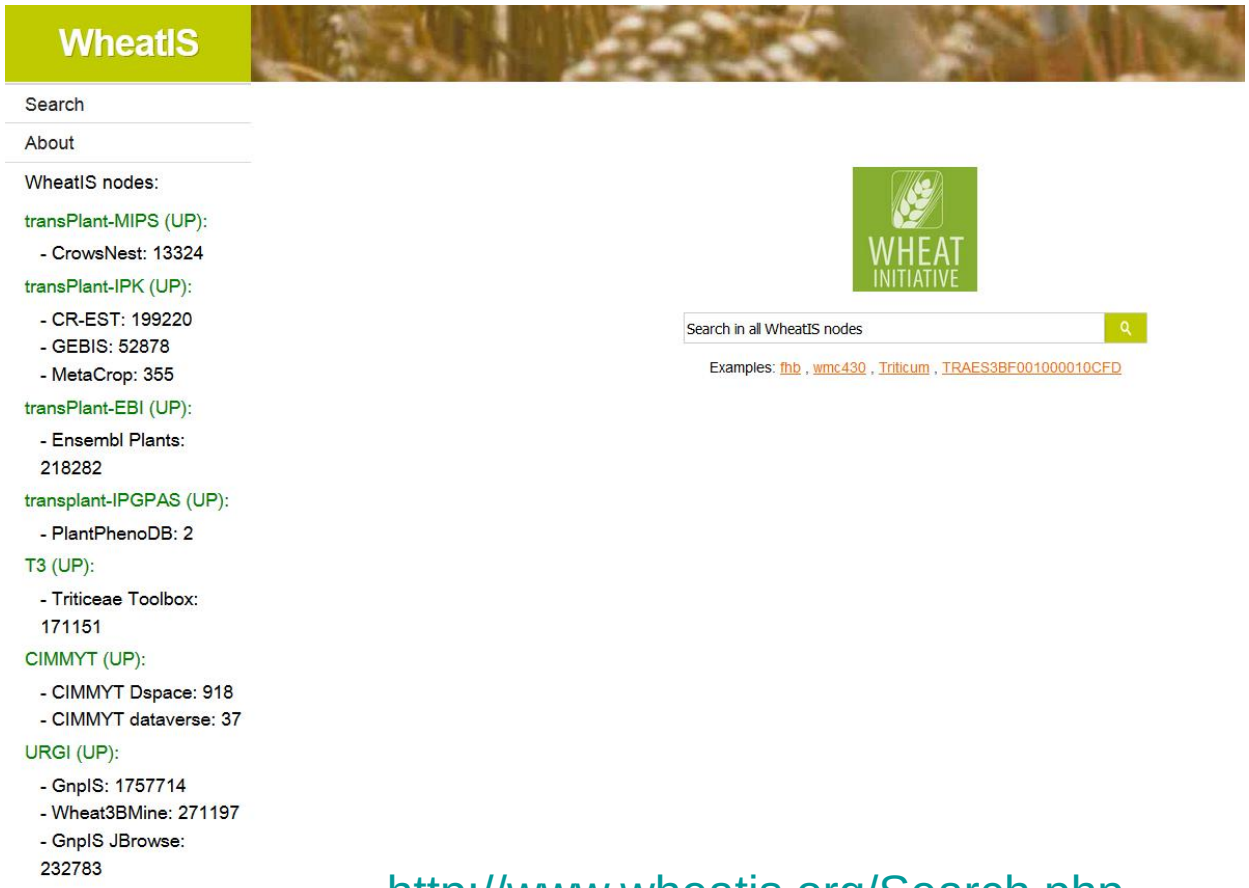
- InterMine tool in development
  - to perform data mining queries though the IWGSC genes,TEs and markers
  - with links to genetics and phenomics data hosted in the URGI Information System (GnpIS).

WheatMine


Wheat3BMine
Home
Templates
Lists
QueryBuilder
Regions
API
MyMine
Contact Us | Log in
Search
Search Wheat3BMine. Enter names, identifiers or keywords for genes, mRNA, repeat region, marker, five prime utr, etc. (e.g. COTEAU, XwPt2416-3, XwPt3327-3B.2, BAC).
e.g. Gene, XwPt1159-3B, SO:0000004
SEARCH
Wheat3BMine contains genomic annota
links from markers card to URGI Wheat
chromosome 3B, Science.
» More queries
Marker : XwPt3327-3B
Name XwPt3327-3B.2
Organism . Name Triticum aestivum
Type darts
SHARE
Overlapping Features
Genome features that overlap coordinates of this Marker
BAC: 1, Region: 1, Scaffold: 1
Lists
This Marker is in 2 lists:
Genome feature
Length: 502 FASTA...
Location: traes3bPseudon
19 QTL
Qtl Name
GPC_cf.9.Np_3B2
GPC_ms.9.NI_3B2
GPC_nw.9.Np_3B2
GPC_sb.9.Np_3B2
GPC_cf.9.NI_3B2
sb.9.HN_GPD_3B2_83
nw.9.HN_GPD_3B2_49.8
nw.9.HN_GPC_3B2_44.8
mean_GY_3B2_40.4
mean_GPD_3B2_49.8
QTL card
QTL DETAILS
QTL name GPC_cf.9.Np_3B2
QTL detection composite interval mapping
Measure GPC_cf.9.Np
Experimentation cf.9.Np
Trait name Prot%
Meta-analysis using this QTL MQTL_TOR107_240811_3B
Number of MetaQTLs that contain this QTL 1
ASSOCIATED EFFECTS
Effects number : 1
Effect type | Effect value | Standard deviation | Line name | Min effect
additive | -0.43 | - | unknown | -
ASSOCIATED ASSIGNMENTS
Assignments number : 1
Map name | Linkage group | LOD | Test type | Test value | R2 | Distance
TOR107_101010 | 3B2 | 6.0 | lod | 6 | 6 | 15
Back to Form
Search parameter(s):
Get Climatik Data
Geolocation
DATA SUMMARY
Trials: 813
Data sets: 4
Trial : BTH_Dijon_2002_SetB3
Site : Dijon
Data Available
Trial : BTH_Estrées-Mons_2003_SetA1
Site : Estrées-Mons
Data Available
Network Data Set : INRA Wheat Network not BRC accession (B and C series)
Network Data Set : INRA Small Grain Cereals Network
DOI:http://dx.doi.org/10.15454/1.4489666216568333E12
Origin site
Collecting site
Evaluation site
Phenotyping campaign(s)
2000 2001 2002 2003 2004 2005 2006 2007
2008 2009 2010 2011 2012 2013 2014 2015
remove all add all
LEVEL: TRIAL
1-10 of 52,875 | Display 10 results per page
Lot Number | Accession Number | Accession Name | itk | Trial Name | Trial Site
Barok | 32101 | BAROK | fi: low inputs | BTH_Lusignan_2010_TECH | Lusignan
AO10015 | 37240 | AO10015 | fi: low inputs | BTH_Lusignan_2010_TECH | Lusignan
AO10011 | 37238 | AO10011 | fi: low inputs | BTH_Lusignan_2010_TECH | Lusignan

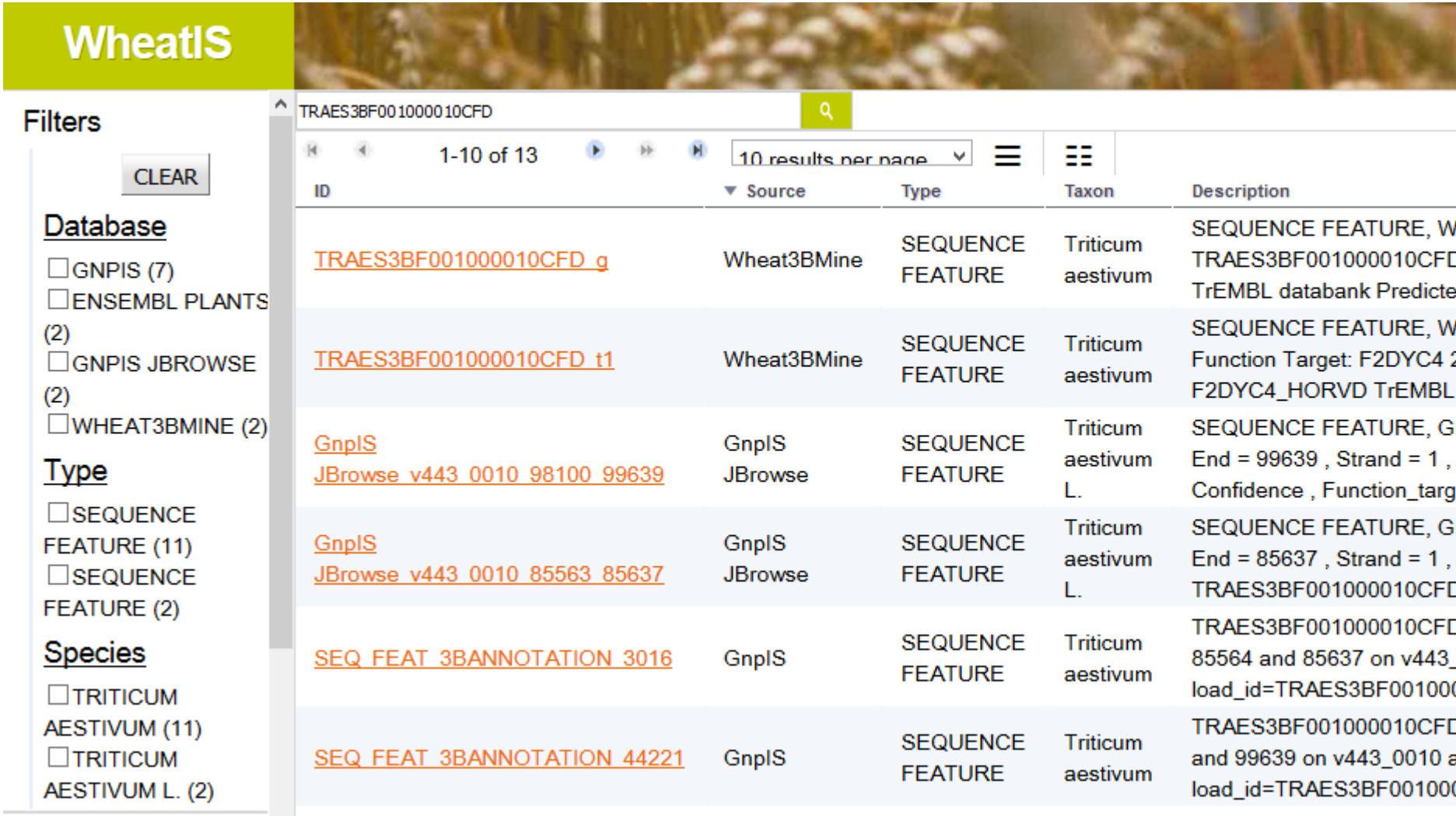
WheatIS search

- IWGSC data can be discovered using the WheatIS search



http://www.wheatis.org/Search.php

WheatIS search



http://www.wheatis.org/Search.php

Acknowledgements

Alaux M.
Letellier T.
Flores R.
Alfama F.
Jamilloux V.
Guerche C.
Loaec M.
Lainé M.
Adam-Blondon A.F.
Quesneville H.

Choulet F.
Rimbert H.
Paux E.

Rogers J.
Caugant I.
Eversole K.

IWGSC Coordinating Committee

IWGSC Sequencing team

All data providers

Michael Alaux

Questions

Sequence Repository
http://wheat-urgi.versailles.inra.fr/Seq-Repository/Assemblies

IWGSC BLAST
https://urgi.versailles.inra.fr/blast_iwgsc/?dbgroup=wheat_whole_genome_assemblies&program=blastn

Contact me at michael.alaux@inra.fr

@michaelalaux