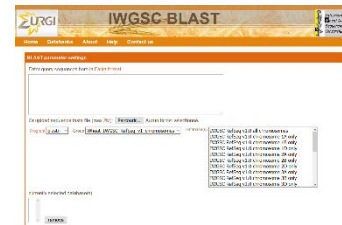# The IWGSC data repository and wheat data resources hosted at URGI: overview and perspectives

Michael Alaux

IWGSC Data Repository Overview

IWGSC Data Repository

Assemblies

- ## IWGSC RefSeq v1.0 Assembly

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

  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 assembly 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 .

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

Michael Alaux

Assemblies

- Other assemblies available:
    - ❑ IWGSC WGA v0.4
    - ❑ IWGSC survey sequence (all versions)
    - ❑ TGAC v1
    - ❑ Other wheat species

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

Michael Alaux

Annotations

- ## IWGSC RefSeq v1.0 Annotation

  - **IWGSC RefSeq v1.0 annotation** is available for download and display in a browser .

  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 annotation includes gene models generated by integrating predictions made by INRA-GDEC using Triannot and PGSB using their customised pipeline (previously MIPS pipeline). The integration was undertaken by the Earlham institute (EI), who have also added UTRs to the gene models where supporting data are available. Gene models have been assigned to high confidence (HC) or low confidence (LC) classes based on completeness, similarity to genes represented in protein and DNA databases and repeat content. The automated assignment of functional annotation to genes has been generated by PGSB based on AHRD parameters. In addition, annotated transposable elements (TEs) and non-coding RNAs are available. More information about the annotation data is provided in the README file .

  How to access the 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/Annotations

Michael Alaux

Annotations

- Data available:
    - ❑ Genes: HighConf and LowConf
    - ❑ Functional annotation
    - ❑ Transposable elements
    - ❑ Markers: ISBP, SNP, DArT, SSR, EST, etc.
    - ❑ ncRNAs: miRNA, lncRNA
    - ❑ RH maps
    - ❑ GBS maps
    - ❑ Optical maps

Michael Alaux

Annotations

- IWGSC RefSeq v1.1 Annotation
    - Refers to the same assembly: the IWGSC RefSeq v1.0 Assembly
    - Data will be available upon publication:
        - Genes: HighConf and LowConf
        - RNA-seq mapping

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

Michael Alaux

Access to the IWGSC RefSeq v1.0 data

- Under Toronto agreement (IWGSC general access agreement)

http://www.wheatgenome.org/Tools-and-Resources

→ will be in open access once published

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

Michael Alaux

Tools to Browse and Mine the IWGSC RefSeq

Download

IWGSC RefSeq 1.0 assembly

[ ]	iwgsc_refseqv1.0_all_chromosomes.zip	04-May-2017 12:31	8.0G
[TXT]	iwgsc_refseqv1.0_all_chromosomes.zip.md5.txt	05-Dec-2017 15:00	71
[ ]	iwgsc_refseqv1.0_chr1A.fsa.zip	04-May-2017 12:14	169M
[TXT]	iwgsc_refseqv1.0_chr1A.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr1B.fsa.zip	04-May-2017 12:15	197M
[TXT]	iwgsc_refseqv1.0_chr1B.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr1D.fsa.zip	04-May-2017 12:15	140M
[TXT]	iwgsc_refseqv1.0_chr1D.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr2A.fsa.zip	04-May-2017 12:16	222M
[TXT]	iwgsc_refseqv1.0_chr2A.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr2B.fsa.zip	04-May-2017 12:17	229M
[TXT]	iwgsc_refseqv1.0_chr2B.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr2D.fsa.zip	04-May-2017 12:18	185M
[TXT]	iwgsc_refseqv1.0_chr2D.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr3A.fsa.zip	04-May-2017 12:18	214M
[TXT]	iwgsc_refseqv1.0_chr3A.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr3B.fsa.zip	04-May-2017 12:19	237M
[TXT]	iwgsc_refseqv1.0_chr3B.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr3D.fsa.zip	04-May-2017 12:20	175M
[TXT]	iwgsc_refseqv1.0_chr3D.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr4A.fsa.zip	04-May-2017 12:21	212M
[TXT]	iwgsc_refseqv1.0_chr4A.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr4B.fsa.zip	04-May-2017 12:22	192M
[TXT]	iwgsc_refseqv1.0_chr4B.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr4D.fsa.zip	04-May-2017 12:22	144M
[TXT]	iwgsc_refseqv1.0_chr4D.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr5A.fsa.zip	04-May-2017 12:23	202M
[TXT]	iwgsc_refseqv1.0_chr5A.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr5B.fsa.zip	04-May-2017 12:24	203M
[TXT]	iwgsc_refseqv1.0_chr5B.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr5D.fsa.zip	04-May-2017 12:25	160M
[TXT]	iwgsc_refseqv1.0_chr5D.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr6A.fsa.zip	04-May-2017 12:25	176M
[TXT]	iwgsc_refseqv1.0_chr6A.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr6B.fsa.zip	04-May-2017 12:26	205M
[TXT]	iwgsc_refseqv1.0_chr6B.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr6D.fsa.zip	04-May-2017 12:27	134M
[TXT]	iwgsc_refseqv1.0_chr6D.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr7A.fsa.zip	04-May-2017 12:27	209M
[TXT]	iwgsc_refseqv1.0_chr7A.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr7B.fsa.zip	04-May-2017 12:28	214M
[TXT]	iwgsc_refseqv1.0_chr7B.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chr7D.fsa.zip	04-May-2017 12:29	181M
[TXT]	iwgsc_refseqv1.0_chr7D.fsa.zip.md5.txt	05-Dec-2017 15:00	65
[ ]	iwgsc_refseqv1.0_chrUn.fsa.zip	04-May-2017 12:29	118M
[TXT]	iwgsc_refseqv1.0_chrUn.fsa.zip.md5.txt	05-Dec-2017 15:00	65

URGI

https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Assemblies/v1.0/

Michael Alaux

Download

IWGSC RefSeq v1.0 annotation

[ ]	iwgsc_refseqv1.0_FunctionalAnnotation_v1.zip	06-Dec-2017 14:49	19M
[TXT]	iwgsc_refseqv1.0_FunctionalAnnotation_v1.zip.md5.txt	06-Dec-2017 14:49	82
[ ]	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
[ ]	iwgsc_refseqv1.0_PGSB_annotation_files.zip	04-Dec-2017 18:04	252M
[TXT]	iwgsc_refseqv1.0_PGSB_annotation_files.zip.md5.txt	05-Dec-2017 15:01	75
[ ]	iwgsc_refseqv1.0_README.pdf	05-Jan-2018 11:58	311K
[ ]	iwgsc_refseqv1.0_RH_map.zip	27-Nov-2017 17:38	2.1M
[TXT]	iwgsc_refseqv1.0_RH_map.zip.md5.txt	05-Dec-2017 15:01	63
[ ]	iwgsc_refseqv1.0_SynOpRIL993_GBS.zip	09-Oct-2017 11:19	485K
[TXT]	iwgsc_refseqv1.0_SynOpRIL993_GBS.zip.md5.txt	05-Dec-2017 15:01	69
[ ]	iwgsc_refseqv1.0_TransposableElements_2017Mar13.gff3.zip	24-Mar-2017 17:27	139M
[TXT]	iwgsc_refseqv1.0_TransposableElements_2017Mar13.gff3.zip.md5.txt	05-Dec-2017 15:01	91
[ ]	iwgsc_refseqv1.0_Whealbi_GWAS.zip	06-Dec-2017 11:10	226M
[TXT]	iwgsc_refseqv1.0_Whealbi_GWAS.zip.md5.txt	06-Dec-2017 11:10	66
[ ]	iwgsc_refseqv1.0_lncRNA_2017Dec13.gff3.zip	19-Dec-2017 16:53	112M
[TXT]	iwgsc_refseqv1.0_lncRNA_2017Dec13.gff3.zip.md5.txt	19-Dec-2017 16:53	78
[ ]	iwgsc_refseqv1.0_miRNA_2017May26.gff3.zip	29-May-2017 11:17	347K
[TXT]	iwgsc_refseqv1.0_miRNA_2017May26.gff3.zip.md5.txt	05-Dec-2017 15:01	76
[ ]	iwgsc_refseqv1.0_miRNA_mature-miRNA_sequences_2017May26.fsa.zip	29-May-2017 11:17	57K
[TXT]	iwgsc_refseqv1.0_miRNA_mature-miRNA_sequences_2017May26.fsa.zip.md5.txt	05-Dec-2017 15:01	98
[ ]	iwgsc_refseqv1.0_miRNA_pre-miRNA_sequences_2017May26.fsa.zip	29-May-2017 11:17	287K
[TXT]	iwgsc_refseqv1.0_miRNA_pre-miRNA_sequences_2017May26.fsa.zip.md5.txt	05-Dec-2017 15:01	95
[ ]	iwgsc_refseqv1.0_optical_maps_group7.zip	27-Nov-2017 17:38	4.6M
[TXT]	iwgsc_refseqv1.0_optical_maps_group7.zip.md5.txt	05-Dec-2017 15:01	76
[ ]	iwgsc_refseqv1.0_vs_TGACv1.refmap.zip	26-Jun-2017 13:39	8.9M
[TXT]	iwgsc_refseqv1.0_vs_TGACv1.refmap.zip.md5.txt	05-Dec-2017 15:01	74
[ ]	iwgsc_refseqv1.0_vs_TGACv1.stats.zip	26-Jun-2017 13:39	1.0K
[TXT]	iwgsc_refseqv1.0_vs_TGACv1.stats.zip.md5.txt	05-Dec-2017 15:01	71
[ ]	iwgsc_refseqv1.0_vs_TGACv1.tmap.zip	26-Jun-2017 13:39	8.9M
[TXT]	iwgsc_refseqv1.0_vs_TGACv1.tmap.zip.md5.txt	05-Dec-2017 15:01	70

https://urgi.versailles.inra.fr/download/iwgsc/IWGSC_RefSeq_Annotations/v1.0/

Michael Alaux

BLAST

- BLAST dedicated to IWGSC general access agreement:

  https://urgi.versailles.inra.fr/blast_iwgsc/?dbgroup=wheat_iwgsc_refseq_v1_chromosomes&program=blastn

- Allow to BLAST all the available assemblies in one time including the **IWGSC RefSeq v1.0**

- 476k BLAST searches performed in 2017

  476k

Michael Alaux

BLAST

BLAST

- Link to download the matching sequence
- Link to display the JBrowse zoomed in the matching region

JBrowse

- IWGSC RefSeq v1.0 JBrowse available under general access agreement:
  https://urgi.versailles.inra.fr/jbrowseiwgsc/gmod_jbrowse/?data=myData%2FIWGSC_RefSeq_v1.0
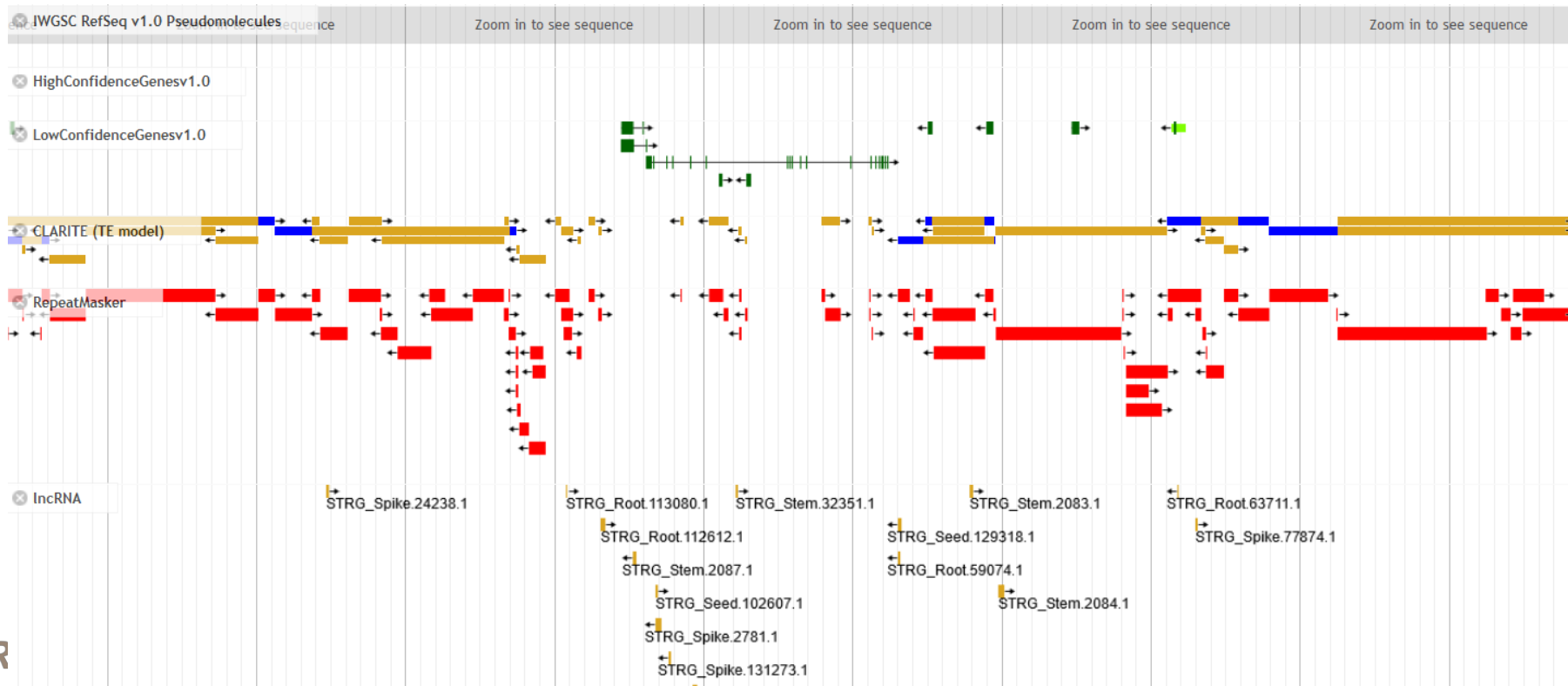
WheatMine

- IWGSC RefSeq v1.0 InterMine available under general access agreement:
https://urgi.versailles.inra.fr/WheatMine

WheatMine

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

filter by text

IWGSC RefSeq v1.0   File   View   Help   Share

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

- ▼ IWGSC Reference Sequence   3
  - ☐ IWGSC RefSeq v1.0 Pseudomolecules
  - ☐ IWGSC RefSeq v1.0 Superscaffolds
  - ☑ IWGSC RefSeq v1.0 Scaffolds
- ▼ IWGSC Official Annotation   6
  - ▼ Gene Model   2
    - ☑ HighConfidenceGenesv1.0
    - ☑ LowConfidenceGenesv1.0
  - ▼ Transposable Elements   2
    - ☐ CLARITE (TE model)
    - ☐ RepeatMasker

miRNA

IWGSC RefSeq v1.0 Scaffolds

HighConfidenceGenesv1.0

LowConfidenceGenesv1.0

3 Data Sets

Name
High confidence genes
Functionnal annotation
Genes and markers sequences

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
IPR001611	Leucine-rich repeat
IPR003591	Leucine-rich repeat, typical subtype
IPR008266	Tyrosine-protein kinase, active site

15 Overlapping Features

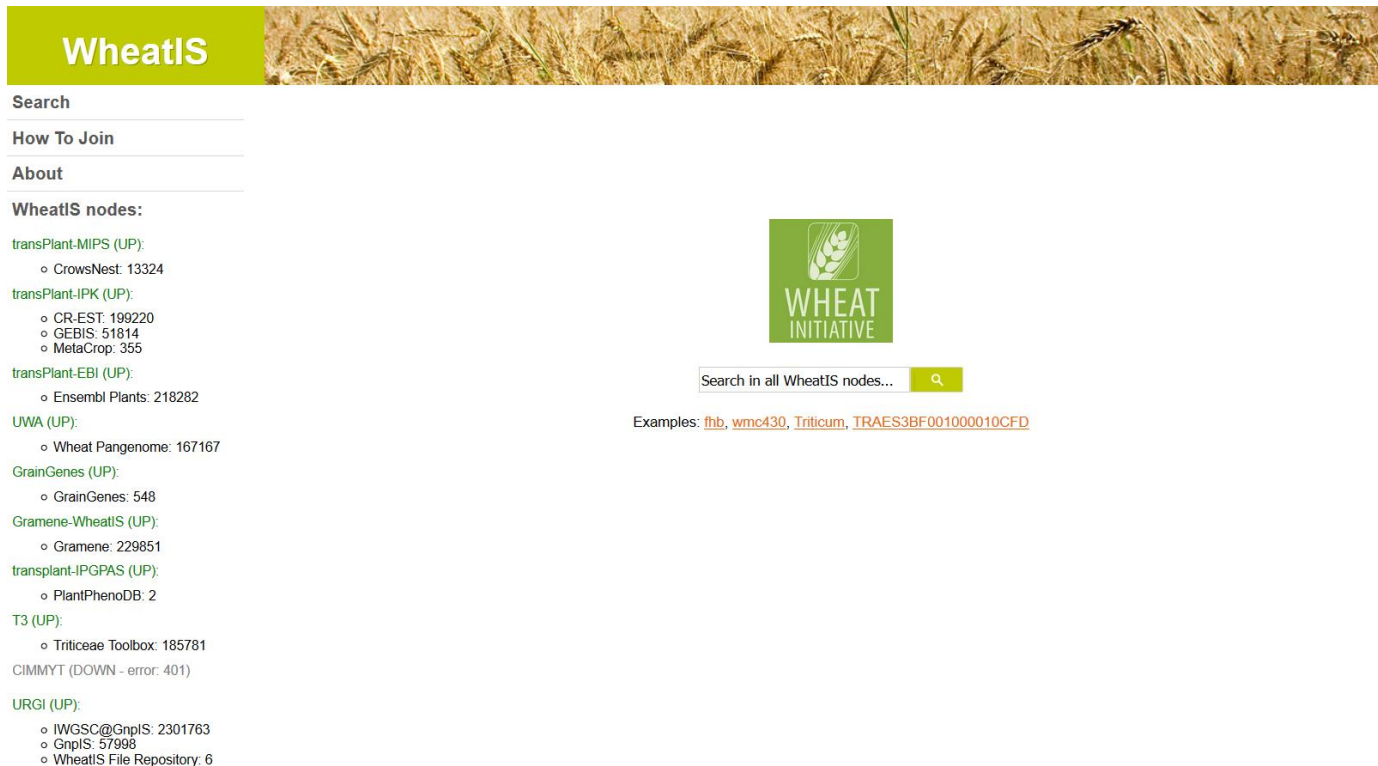| Class	DB identifier	Length	Chromosome Location	Organism . Name
NcRNA	STRG_Root.62029.2.path1	57855	chr1A: 3639864-3697718	
NcRNA	STRG_Root.62029.2.mrna1	57855	chr1A: 3639864-3697718	
NcRNA	STRG_Seedling.59.1.path1	2320	chr1A: 3668200-3670519	
NcRNA	STRG_Seedling.59.1.mrna1	2320	chr1A: 3668200-3670519	
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	Triticum aestivum
Scaffold	scaffold121111	1501578	chr1A: 2943949-4445526	Triticum aestivum

▽ Show more rows

2 Pfams

Identifier	Description
PF00069	Protein kinase domain
PF13855	Leucine rich repeat

URGI

WheatIS search

- IWGSC data can be discovered using the WheatIS search:

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

Michael Alaux

WheatIS search

- IWGSC RefSeq v1.0 data will be added as soon as they are in open access

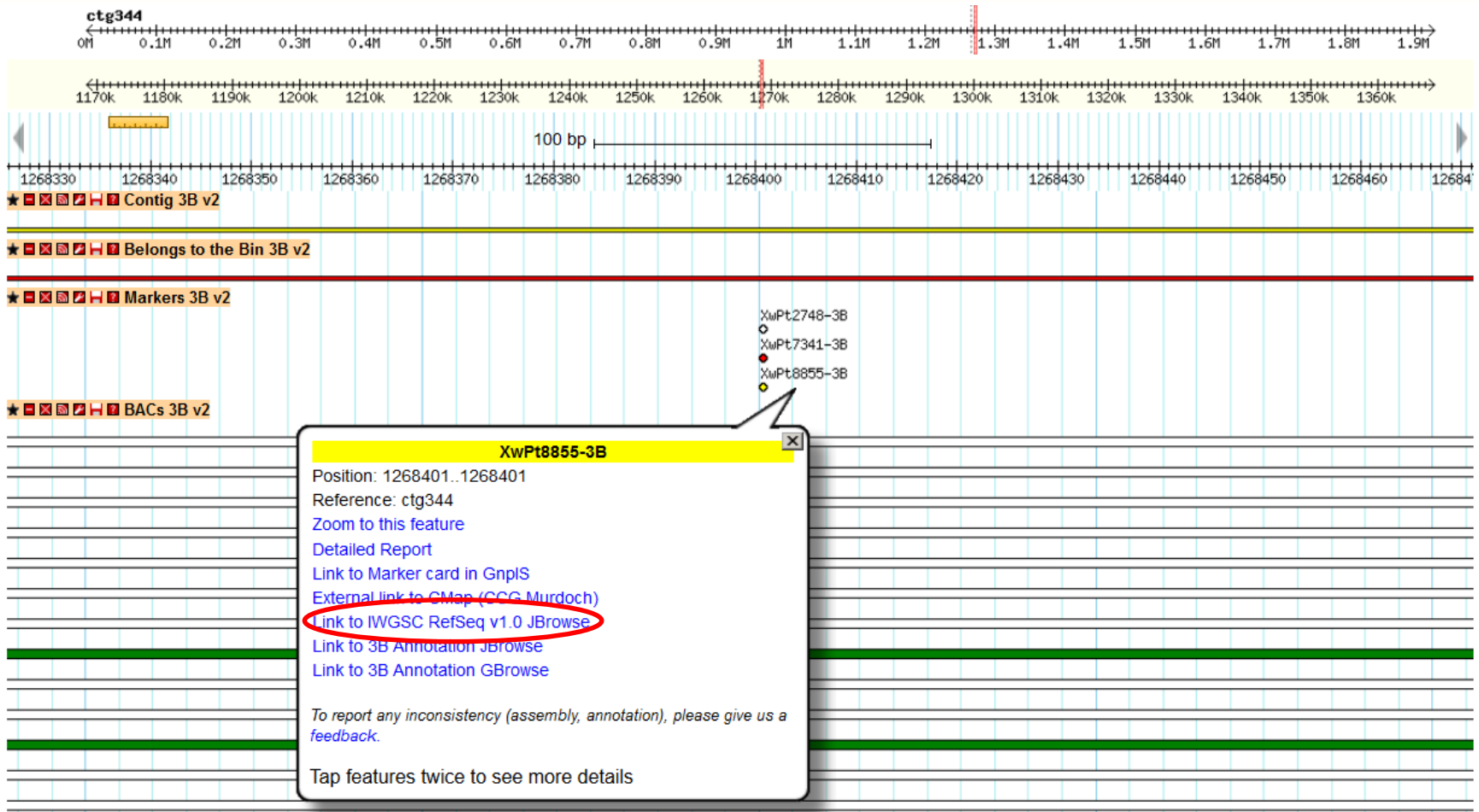| ID	Source	Type	Taxon	Description
JBrowse_chr1A_100233074_100233167_miRNA	IWGSC@GnpIS	SEQUENCE FEATURE	Triticum aestivum	Start = 100233074 [...]
JBrowse_chr1A_101106604_101106809_miRNA	IWGSC@GnpIS	SEQUENCE FEATURE	Triticum aestivum	Start = 101106604 [...]
JBrowse_chr1A_10158696_10158807_miRNA	IWGSC@GnpIS	SEQUENCE FEATURE	Triticum aestivum	Start = 10158696 [...]
JBrowse_chr1A_10231918_10232011_miRNA	IWGSC@GnpIS	SEQUENCE FEATURE	Triticum aestivum	Start = 10231918 [...]
JBrowse_chr1A_102424297_102424428_miRNA	IWGSC@GnpIS	SEQUENCE FEATURE	Triticum aestivum	Start = 102424297 [...]
JBrowse_chr1A_104839251_104839344_miRNA	IWGSC@GnpIS	SEQUENCE FEATURE	Triticum aestivum	Start = 104839251 [...]
JBrowse_chr1A_104839691_104839784_miRNA	IWGSC@GnpIS	SEQUENCE FEATURE	Triticum aestivum	Start = 104839691 [...]
JBrowse_chr1A_105619444_105619537_miRNA	IWGSC@GnpIS	SEQUENCE FEATURE	Triticum aestivum	Start = 105619444 [...]
JBrowse_chr1A_107474939_107475091_miRNA	IWGSC@GnpIS	SEQUENCE FEATURE	Triticum aestivum	Start = 107474939 [...]
JBrowse_chr1A_108071355_108071486_miRNA	IWGSC@GnpIS	SEQUENCE FEATURE	Triticum aestivum	Start = 108071355 [...]

URGI (UP):

- ○ IWGSC@GnpIS: 16385772

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

Michael Alaux

Physical maps

- New versions of 7DS, 1BS added

- All the physical maps are available to display

1A 2A 3A 4A 5A 6A 7A 1B 2B 3B 4B 5B 6B 7B 1D 2D 3D 4D 5D 6D 7D
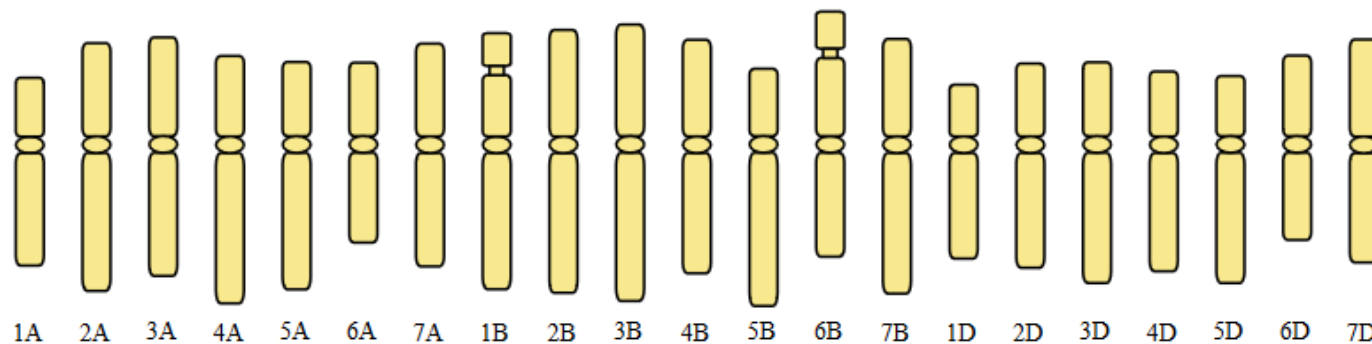
- Link to the IWGSC RefSeq v1.0

Michael Alaux

Physical maps

Michael Alaux

Other Wheat data resources hosted at URGI

GnpIS-coreDB

- Big data generated from french and european projects

- These data are integrated in the GnpIS-coreDB developed at URGI (part of the GnpIS Information System).

Michael Alaux

GnpIS-coreDB

- Wheat data overview:

Thematic	Object	#Total	#Open access	#Restricted access to projects
Genetic Resources	Taxon	56	56	0
	Accession	12839	10016	2823
Genetic Maps	Map	30	29	1
	Marker	704822	34164	670658
	QTL	749	465	284
SNP discovery	In Silico Analysis	11	9	2
	Sequence Variation	134904	55362	79542
	SNP, indel	724132	95	724037
Genotyping (high throughput)	Experiment	22	1	21
	Sample	8229	42	8216
	Marker	668543	0	668543
Phenotyping	Trial	853	821	32
	Sample	3660	2985	901
	Variable	291	91	200
	Observation	1161038	527981	633057
GWAS	Analysis	1555	43	1512
	Sample	2365	1839	526
	Variable	359	37	322
	Marker	123866	4109	119757
	Association	824217	48596	775621

URGI

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

Michael Alaux

Genetic and phenomic data

Michael Alaux

Integration of the sequence
to genetic and phenomic data

- cf. **Alaux *et al.*** companion paper that will be submitted soon.

- Use cases:
  - BLAST
  - → Gene (JBrowse)
  - → Marker (GnpIS-coreDB)
  - → QTLs (GnpIS-coreDB)
  - → Phenotyping experiments (GnpIS-coreDB)

Michael Alaux

Integration of the sequence to genetic and phenomic data

❏ WheatIS search

→ Gene (WheatMine)

→ Marker (GnpIS-coreDB)

→ GWAS (GnpIS-coreDB)

→ Phenotyping experiments (GnpIS-coreDB)

Michael Alaux

Perspectives

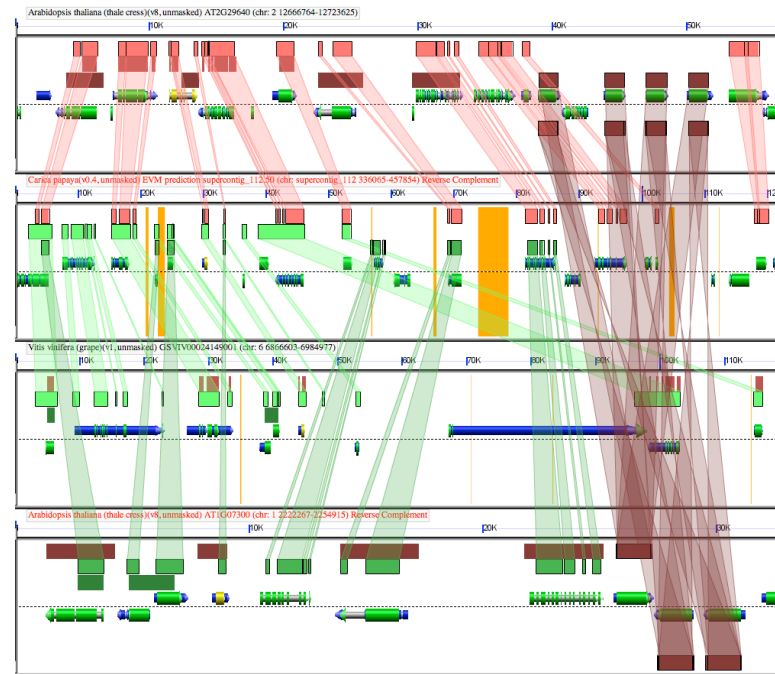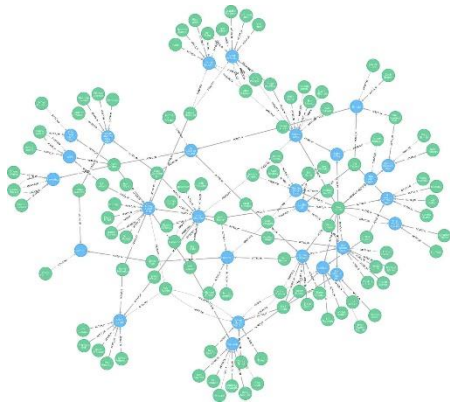- Management of new/curated versions of the IWGSC RefSeq annotation

Perspectives

- Bioinformatics challenges of a Pan-genome
  - ❑ Use new technologies to handle and display large amount of linked data

Michael Alaux

Acknowledgements

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

Choulet F.
Rimbert H.
Leroy P.
Guilhot N.
Paux E.

Rogers J.
Caugant I.
Eversole K.

IWGSC Coordinating
Committee

IWGSC Sequencing and
Analysis team

All data providers

Questions

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

IWGSC BLAST
https://urgi.versailles.inra.fr/blast_iwgsc/?dbgroup=wheat_iwgsc_refseq_v1_chromosomes&program=blastn

Contact me at michael.alaux@inra.fr

@michaelalaux