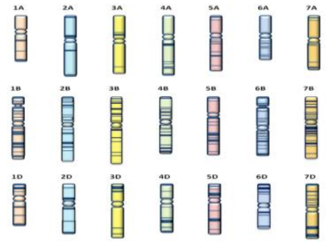
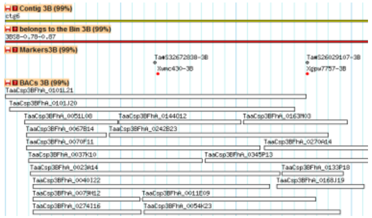
IWGSC Sequence Repository:

How to facilitate pseudomolecule assembly ?

Michael Alaux





IWGSC S&P, San Diego, 13 January 2015

Wheat@URGI website

Michael Alaux

Wheat@URGI

Xwmc430

Examples: Xwmc430, QTL, Tae_1272327

EVENTS & PUBLICATIONS

02 Jan 2014
Talks at PAG 2014

Computer demo 1: Saturday 11 at 1.50 PM **Wheat 3B Annotation Viewer: Browse ...**

02 Jan 2014
IWGSC Sequence Repository : new data and browsers ...
COM (talks)
International Wheat Genetics Symposium (IWGS) 12th 2013, Yokohama, Japan
...

WHAT'S NEW ?

21 Nov 2014
2A physical map is now available
2AS and 2AL physical maps are now available for IWGSC partners.
...

22 Oct 2014
Wheat3BMine is now available
Wheat3BMine, a data warehouse dedicated to wheat chromosome 3B is now available. ...

11 Oct 2014
GnpIS version 14.2 availability
GnpIS version 14.2 is available.
Major new functionalities include:
- Sequence Polymorphisms:Genotyping : The ...

03 Sep 2014
New survey sequence gene models (v2.2) available
New version of the survey sequence gene models (v2.2 July 2014) is ...

02 Sep 2014
GBrowse Wheat Survey Sequence Annotation is available
The GBrowse Wheat Survey Sequence Annotation (Markers, SNPs, Gene Models) is now available
...

02 Sep 2014
7DS physical map is publicly available
7DS physical map is publicly available

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

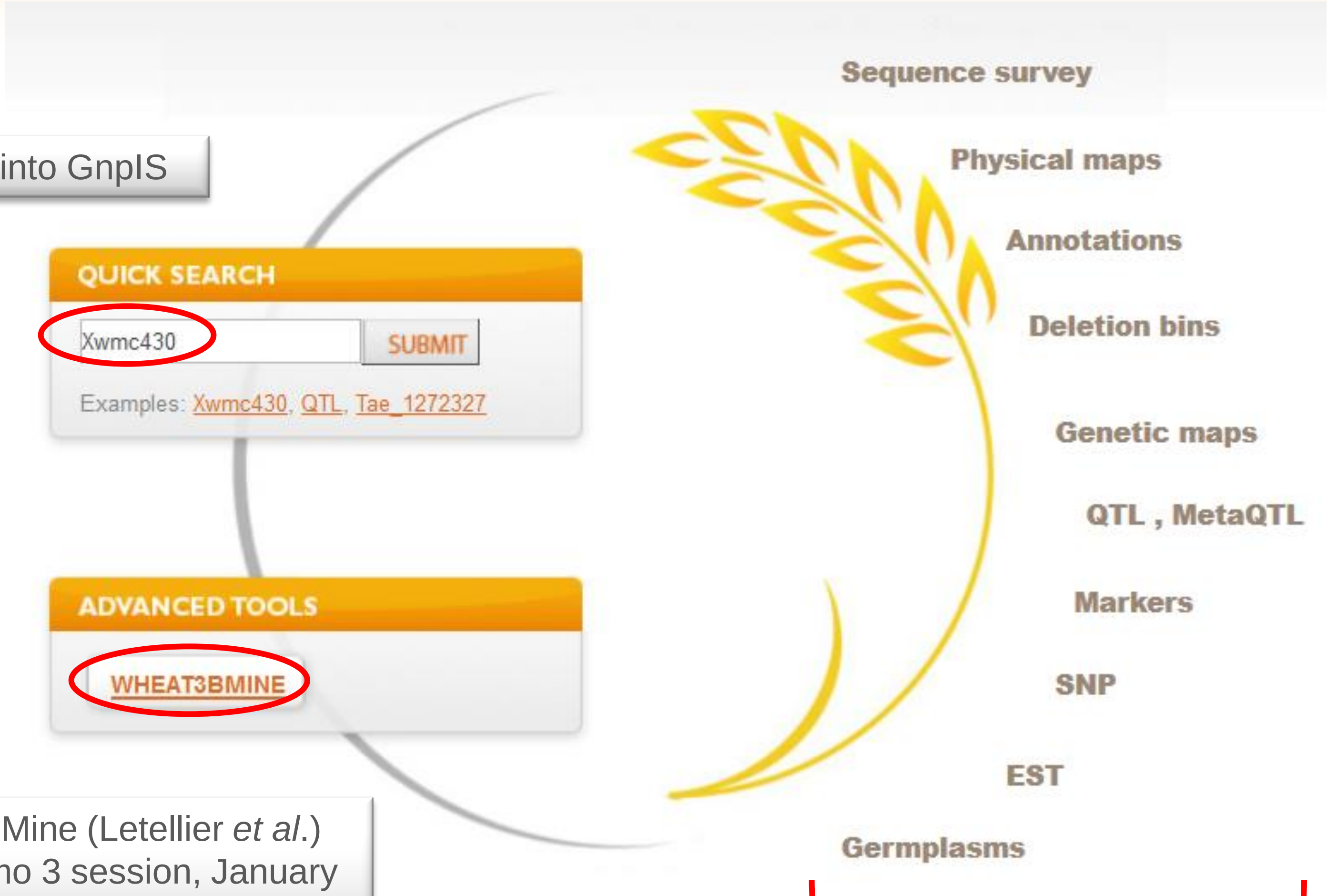
Michael Alaux

Wheat@URGI

Search wheat data into GnpIS

Google-like

Intermine



Look at Wheat3BMine (Letellier *et al.*)
- computer demo 3 session, January the 13th at 11AM
- poster P0654

Quick Links to data

Wheat@URGI

Projects Data Tools Seq Repository

Projects

- 3BSeq
- BreedWheat
- TriticeaeGenome
- Whealbi
- Wheat Initiative

You are here : Home / Home Wheat / Projects

Projects

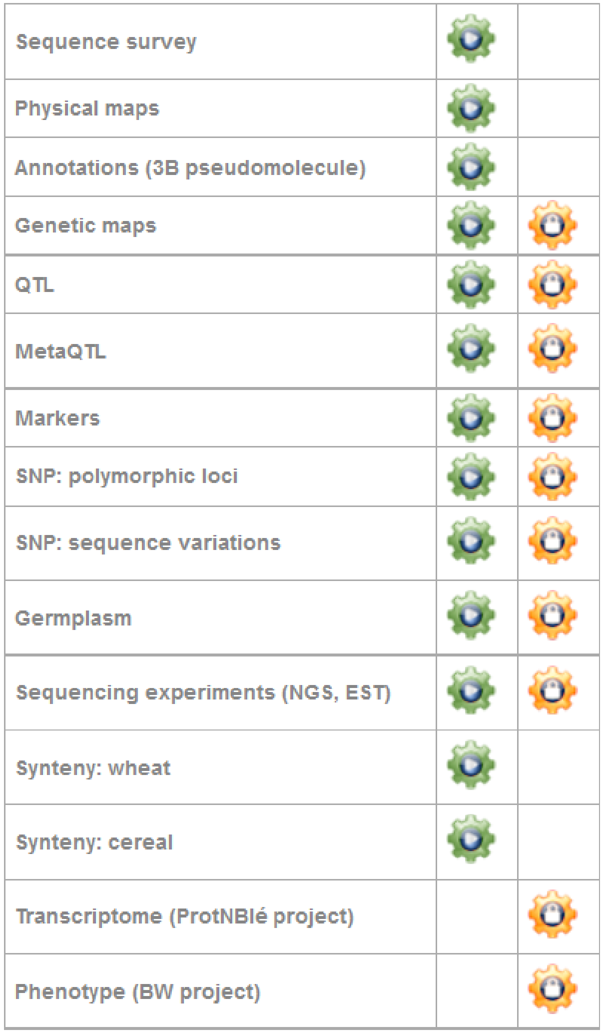
No one can be a statesman who is entirely ignorant of the problems of wheat (Socrates, 470-399BC).

Why develop Wheat genomics?

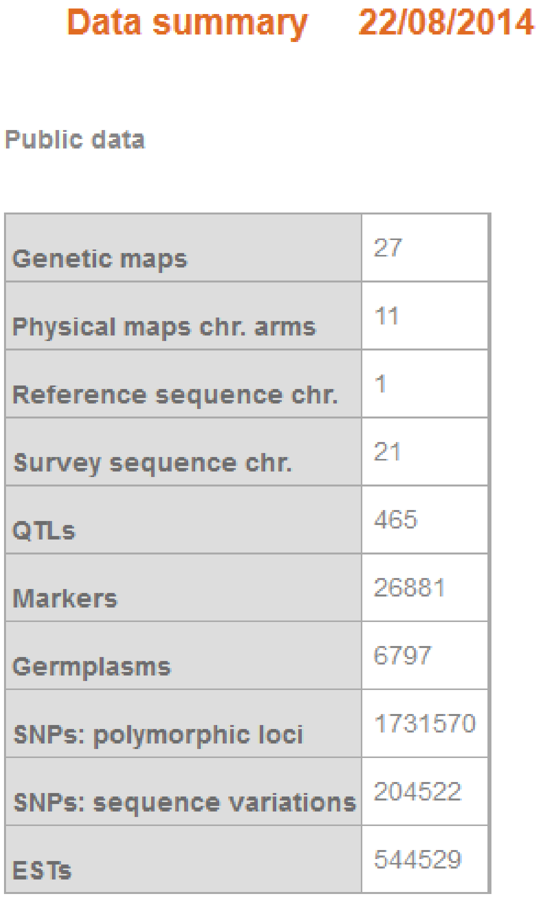
Today, agriculture is facing one of its greatest challenges since it began 10 000 years ago, that of producing an ample supply of high quality food and feed for a growing world population in a sustainable and environment friendly manner. Bread wheat (Triticum aestivum L.) is the staple food for more than 35% of the people and is grown on over 95% of the wheat growing area. Its sequence holds the key to genetic improvements that will allow growers to meet the increasing demands for high quality food and feed produced in an environmentally sensitive, sustainable, and profitable manner. Further, because of its recent history, hexaploid wheat is a very good model to study polyploidy, a driving evolutionary force for more than 90% of all plants.

The International Wheat Genome Sequencing Consortium (IWGSC) was established by a group of plant scientists, breeders, and growers dedicated to sequencing the wheat genome to enhance our knowledge of the structure and function of the wheat genome. By gaining increased understanding of the biology of agronomically important traits and deploying state-of-the-art molecular tools, plant scientists and breeders will be able to accelerate wheat improvement to meet the challenges of the 21st century. The IWGSC is committed to ensuring that the sequence of the wheat genome and the resulting DNA-based tools are available for all to use without restriction.

Wheat@URGI



Sequence survey	⚙	
Physical maps	⚙	
Annotations (3B pseudomolecule)	⚙	
Genetic maps	⚙	🔒
QTL	⚙	🔒
MetaQTL	⚙	🔒
Markers	⚙	🔒
SNP: polymorphic loci	⚙	🔒
SNP: sequence variations	⚙	🔒
Germplasm	⚙	🔒
Sequencing experiments (NGS, EST)	⚙	🔒
Synteny: wheat	⚙	
Synteny: cereal	⚙	
Transcriptome (ProtNBlé project)		🔒
Phenotype (BW project)		🔒



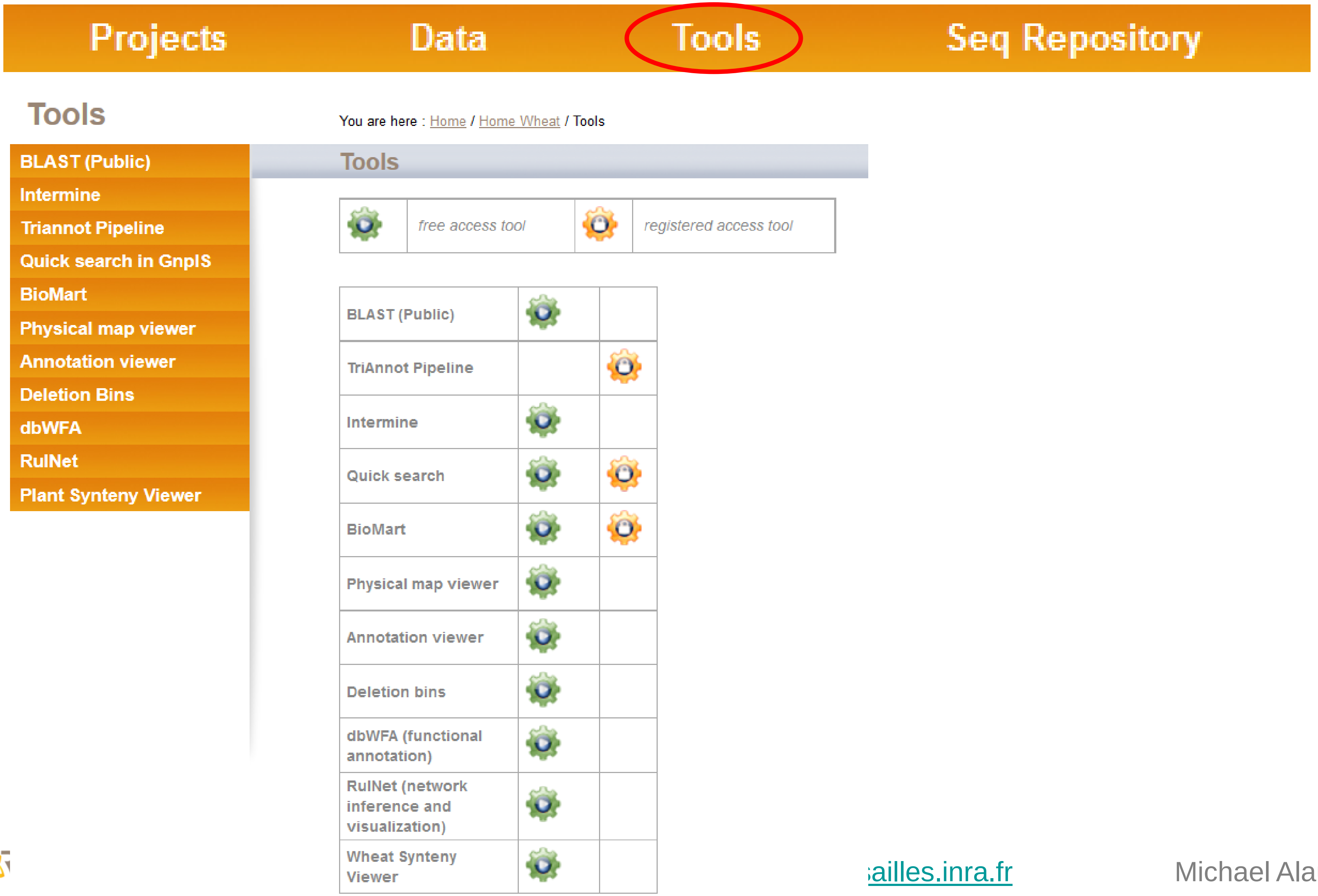
Data summary 22/08/2014

Public data

Genetic maps	27
Physical maps chr. arms	11
Reference sequence chr.	1
Survey sequence chr.	21
QTLs	465
Markers	26881
Germplasms	6797
SNPs: polymorphic loci	1731570
SNPs: sequence variations	204522
ESTs	544529

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

Wheat@URGI



Tools

- BLAST (Public)
- Intermine
- Triannot Pipeline
- Quick search in GnpIS
- BioMart
- Physical map viewer
- Annotation viewer
- Deletion Bins
- dbWFA
- RulNet
- Plant Synteny Viewer

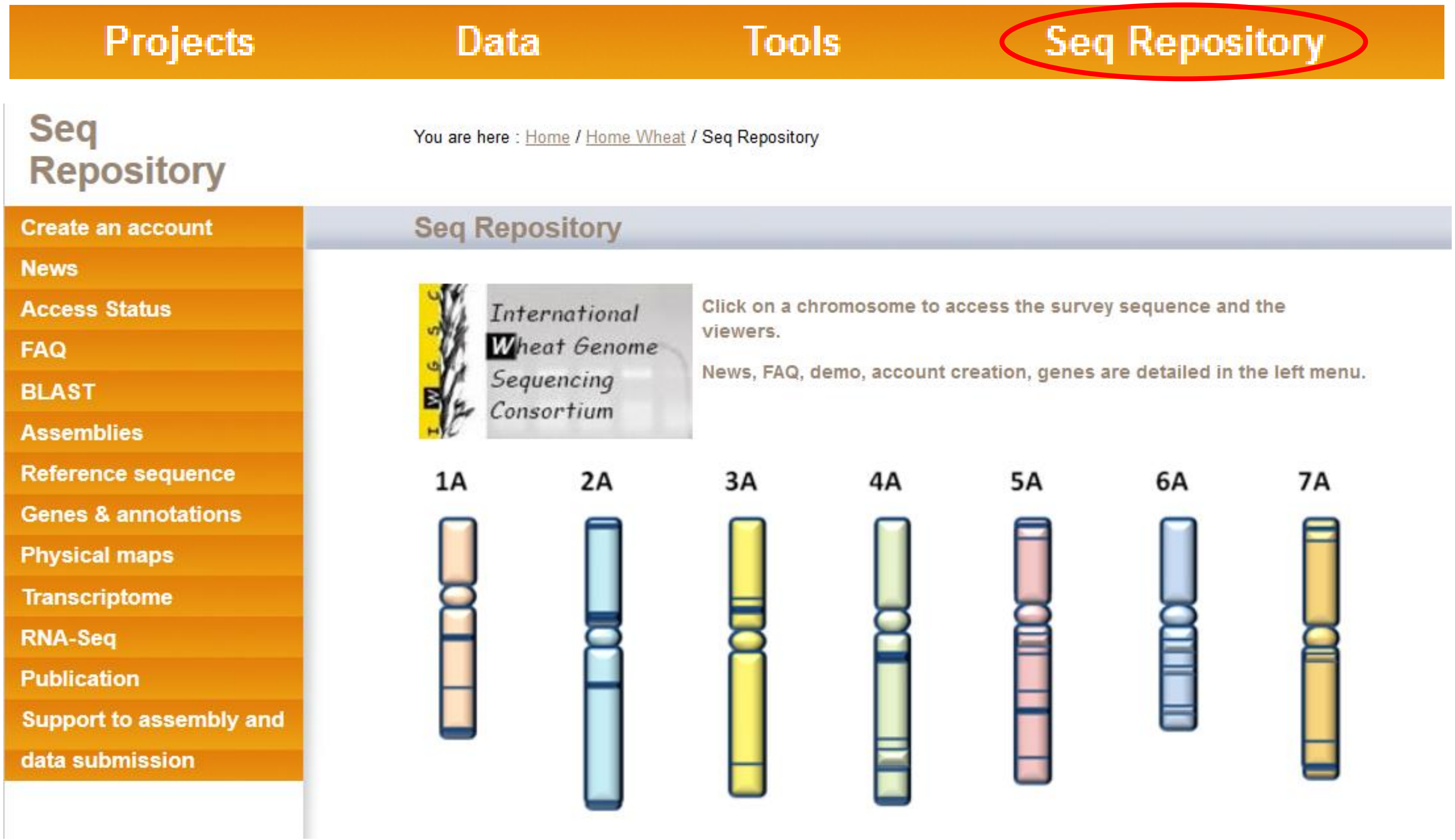
You are here : Home / Home Wheat / Tools

Tools

🟢	*free access tool*	🟠	*registered access tool*

BLAST (Public)	🟢	
TriAnnot Pipeline		🟠
Intermine	🟢	
Quick search	🟢	🟠
BioMart	🟢	🟠
Physical map viewer	🟢	
Annotation viewer	🟢	
Deletion bins	🟢	
dbWFA (functional annotation)	🟢	
RulNet (network inference and visualization)	🟢	
Wheat Synteny Viewer	🟢	

Wheat@URGI



You are here : Home / Home Wheat / Seq Repository

Seq Repository

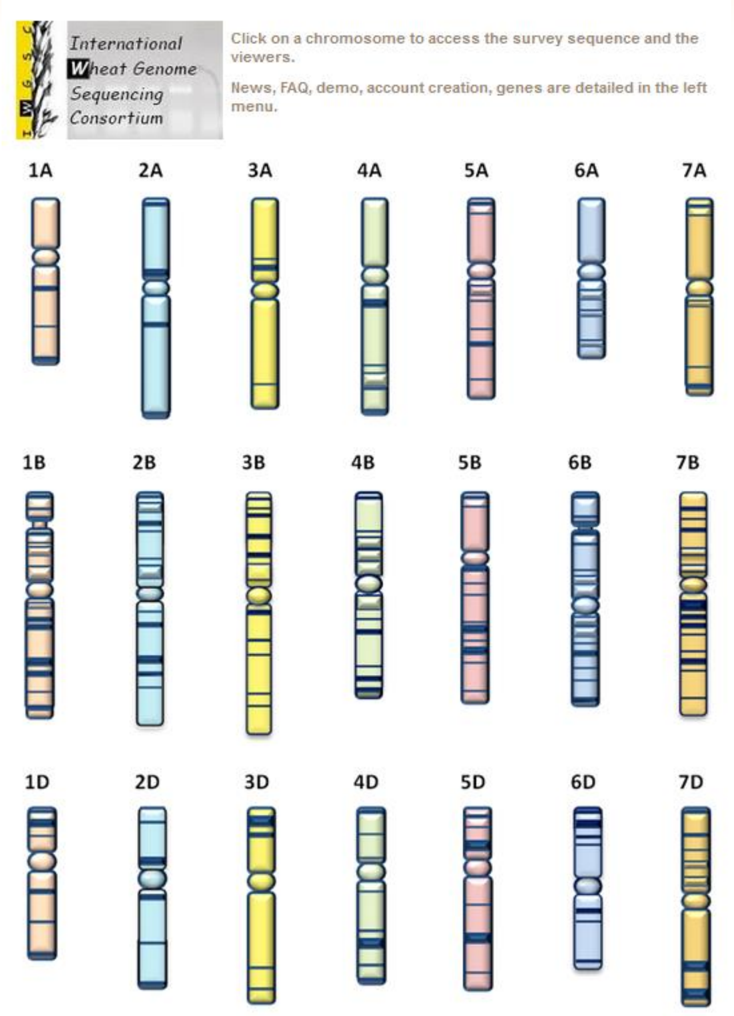
Click on a chromosome to access the survey sequence and the viewers.

News, FAQ, demo, account creation, genes are detailed in the left menu.

Sequence Repository

Michael Alaux

Sequence Repository



International Wheat Genome Sequencing Consortium

Click on a chromosome to access the survey sequence and the viewers.

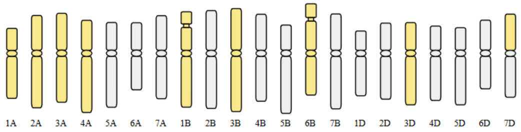
News, FAQ, demo, account creation, genes are detailed in the left menu.

Click on a chromosome to access the survey sequence chromosome arm assembly for:

BLAST (Public)

Download (Public)

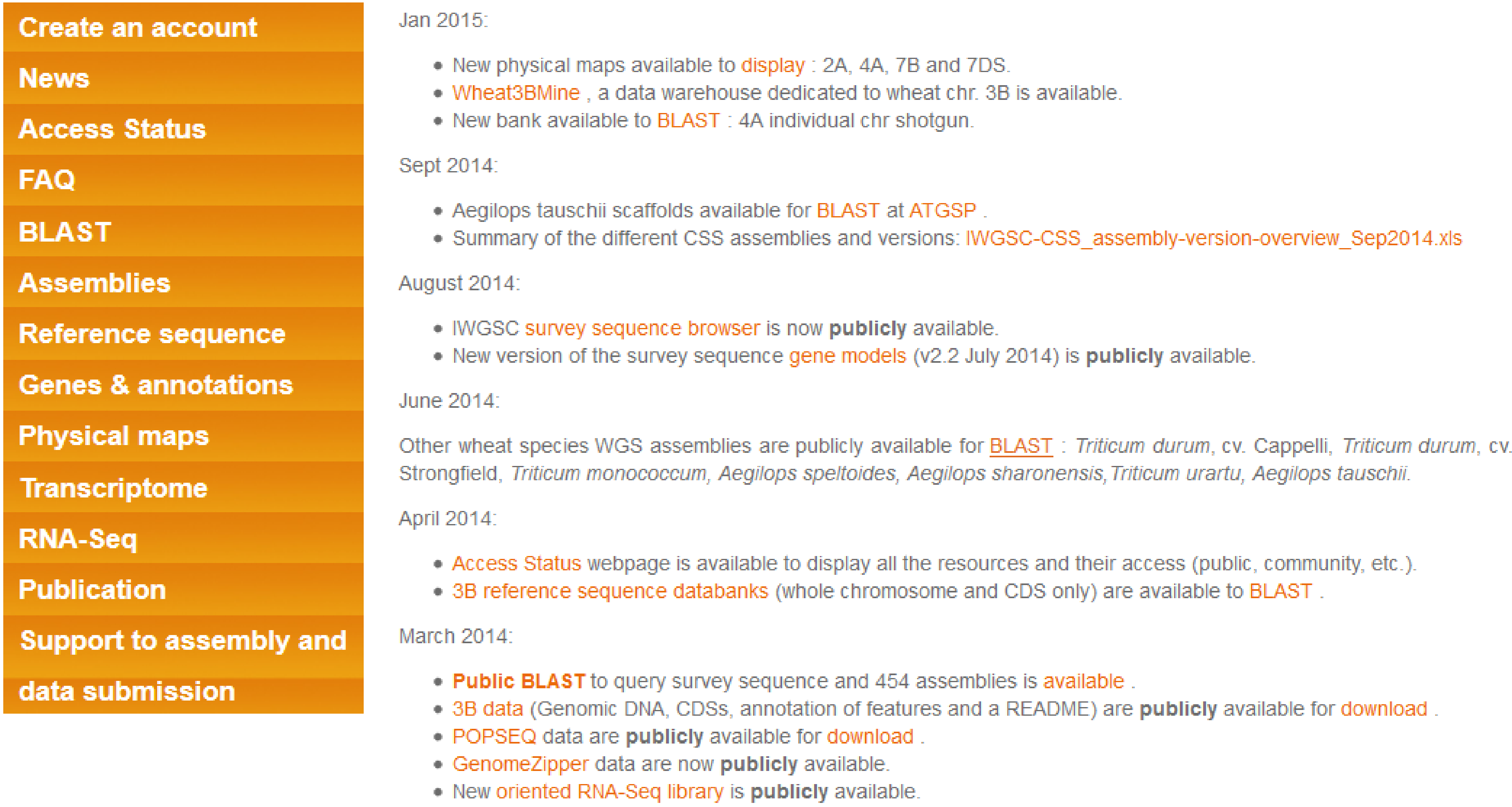
Browsers:



- Physical maps
- Annotated reference sequence 3B (Public)
- Annotated survey sequence (Public)

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

Sequence Repository



Create an account

✓ News

Access Status

FAQ

BLAST

Assemblies

Reference sequence

Genes & annotations

Physical maps

Transcriptome

RNA-Seq

Publication

Support to assembly and data submission

Jan 2015:

- New physical maps available to display : 2A, 4A, 7B and 7DS.
- Wheat3BMine , a data warehouse dedicated to wheat chr. 3B is available.
- New bank available to BLAST : 4A individual chr shotgun.

Sept 2014:

- Aegilops tauschii scaffolds available for BLAST at ATGSP .
- Summary of the different CSS assemblies and versions: IWGSC-CSS_assembly-version-overview_Sep2014.xls

August 2014:

- IWGSC survey sequence browser is now **publicly** available.
- New version of the survey sequence gene models (v2.2 July 2014) is **publicly** available.

June 2014:

Other wheat species WGS assemblies are publicly available for BLAST : *Triticum durum*, cv. Cappelli, *Triticum durum*, cv. Strongfield, *Triticum monococcum*, *Aegilops speltoides*, *Aegilops sharonensis*, *Triticum urartu*, *Aegilops tauschii*.

April 2014:

- Access Status webpage is available to display all the resources and their access (public, community, etc.).
- 3B reference sequence databanks (whole chromosome and CDS only) are available to BLAST .

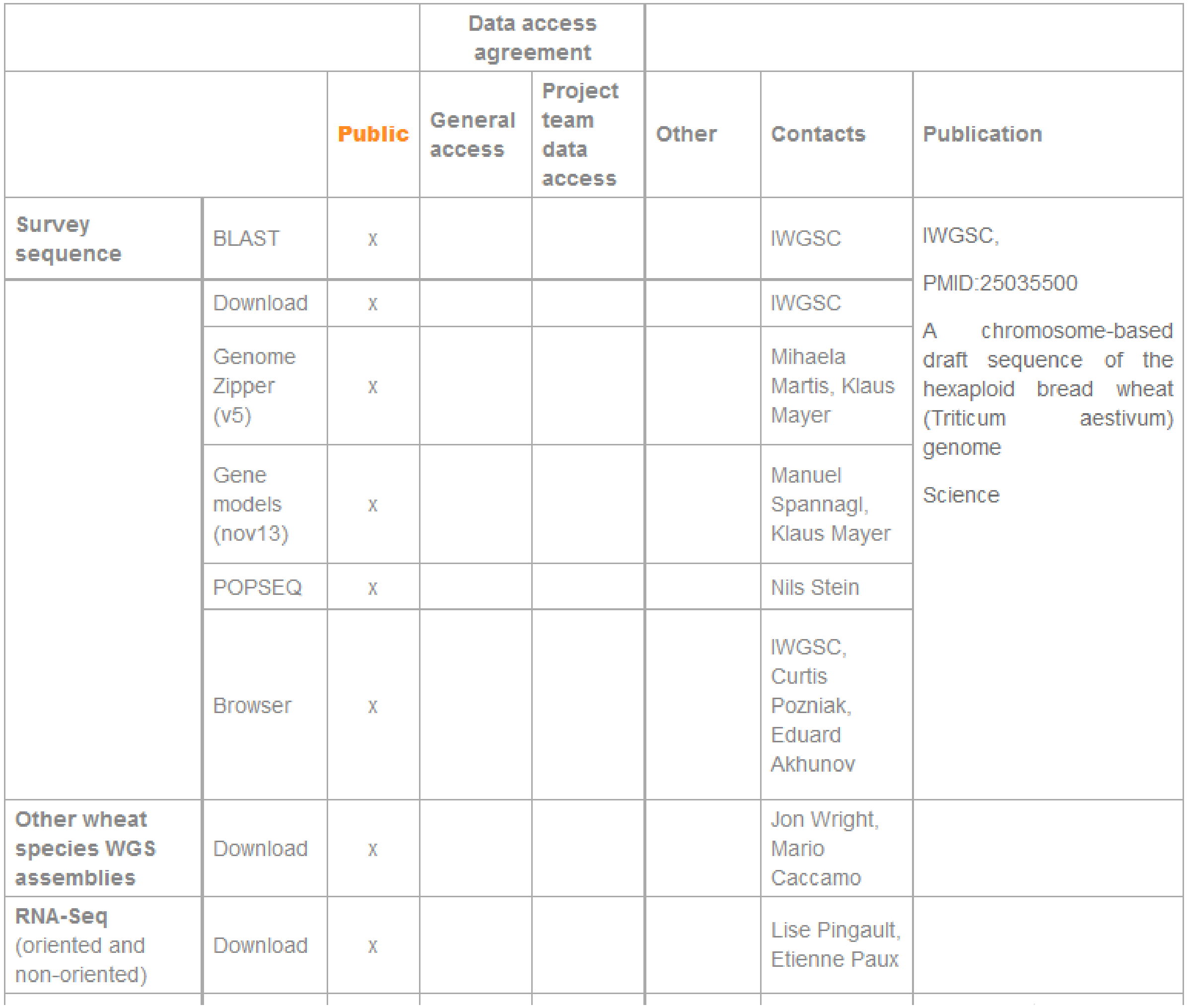
March 2014:

- **Public BLAST** to query survey sequence and 454 assemblies is available .
- 3B data (Genomic DNA, CDSs, annotation of features and a README) are **publicly** available for download .
- POPSEQ data are **publicly** available for download .
- GenomeZipper data are now **publicly** available.
- New oriented RNA-Seq library is **publicly** available.

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

Sequence Repository

- Create an account
- News
- ✓ Access Status
- FAQ
- BLAST
- Assemblies
- Reference sequence
- Genes & annotations
- Physical maps
- Transcriptome
- RNA-Seq
- Publication
- Support to assembly and data submission



			Data access agreement				
		Public	**General access**	**Project team data access**	**Other**	**Contacts**	**Publication**
Survey sequence	BLAST	x				IWGSC	IWGSC, PMID:25035500 A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome Science
	Download	x				IWGSC	
	Genome Zipper (v5)	x				Mihaela Martis, Klaus Mayer	
	Gene models (nov13)	x				Manuel Spannagl, Klaus Mayer	
	POPSEQ	x				Nils Stein	
	Browser	x				IWGSC, Curtis Pozniak, Eduard Akhunov	
Other wheat species WGS assemblies	Download	x				Jon Wright, Mario Caccamo	
RNA-Seq (oriented and non-oriented)	Download	x				Lise Pingault, Etienne Paux	

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

Sequence Repository

- Create an account
- News
- Access Status
- FAQ
- ✓ BLAST
- Assemblies
- Reference sequence
- Genes & annotations
- Physical maps
- Transcriptome
- RNA-Seq
- Publication
- Support to assembly and data submission

- **Public BLAST**

A BLAST is now **publicly available** to query the Triticum aestivum **survey sequence** assemblies, the **3B reference sequence** (whole chromosome and CDS only), **454 assemblies** and **other wheat species** assemblies.

- SRS BLAST

A short demonstration video of the SRS BLAST is available here .

A direct link is available to perform a SRS BLAST (registered access).

- External BLAST

Aegilops tauschii scaffolds available for BLAST at ATGSP .

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

Sequence Repository

Triticum aestivum (Chinese Spring) survey sequence chromosomes assemblies are **publicly available** for download , BLAST and in a browser .

Summary of the different CSS assemblies and versions (TGAC): IWGSC-CSS_assembly-version-overview_Sep2014.xls (30.50 kB)

Some individual chromosome shotgun (454 data) are also **publicly available** for BLAST .

Other wheat species WGS assemblies (TGAC, TSL) are **publicly available** for download and BLAST :

- *Triticum durum*, cv. Cappelli (listed as durum _v1)
- *Triticum durum*, cv. Strongfield
- *Triticum monococcum*
- *Aegilops speltoides*
- *Aegilops. sharonensis*
- *Triticum urartu*
- *Aegilops tauschii*

Triticum durum cv. Cappelli

The cultivar Senatore Cappelli (know as Cappelli) is an historical durum wheat genotype selected from a North African landrace by N. Strampelli and registered in 1915. Cappelli is one of the founders of the Italian durum wheat breeding program and it is present in the pedigree of many durum wheat cultivars released in southern Europe in the XXth century. Cappelli has been largely grown in Mediterranean regions since the late 1950s.
Cappelli is characterized by an elevated water use efficiency (Rizza et al., 2012 Field Crops Research 125, 49–60).

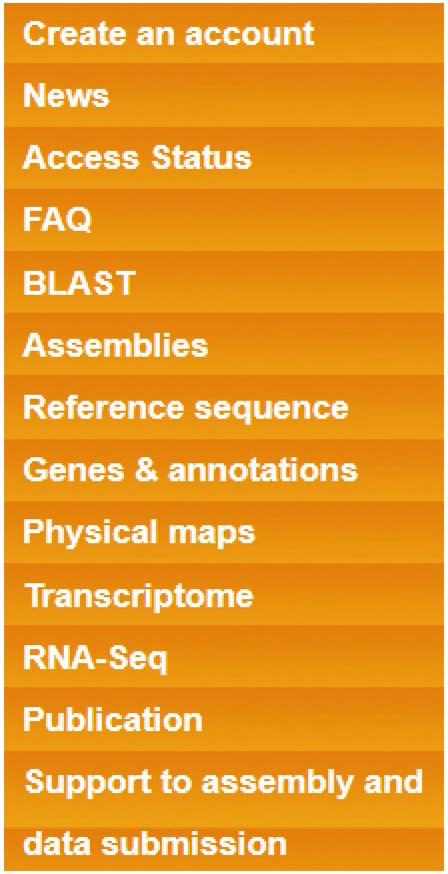
Triticum durum cv. Strongfield

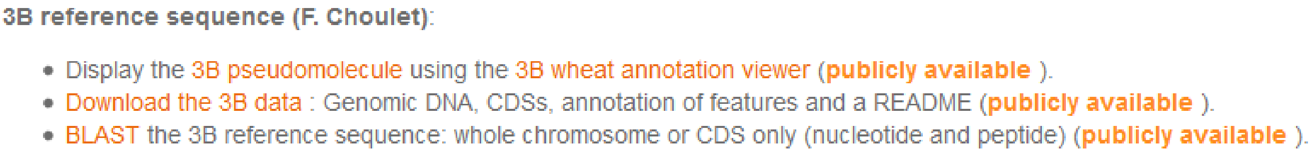
Strongfield durum wheat was developed by Dr. John Clarke during his tenure at Agriculture and Agri-Food Canada. "Strongfield" expresses high grain yield, high grain protein concentration coupled with low grain cadmium concentration. Canadian durum what production represents 60% of durum wheat traded globally and Strongfield currently occupies 65% of the total planted area of durum wheat in Canada. A detailed description of the variety has been published (Clarke et al. 2005; Can J. Plant Sci., 83: 651-654 see http://pubs.aic.ca/doi/abs/10.4141/P04-119).

Aegilops tauschii scaffolds available for BLAST at ATGSP

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

Sequence Repository




3B reference sequence (F. Choulet):

- Display the 3B pseudomolecule using the 3B wheat annotation viewer (**publicly available**).
- Download the 3B data : Genomic DNA, CDSs, annotation of features and a README (**publicly available**).
- BLAST the 3B reference sequence: whole chromosome or CDS only (nucleotide and peptide) (**publicly available**).

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

Sequence Repository

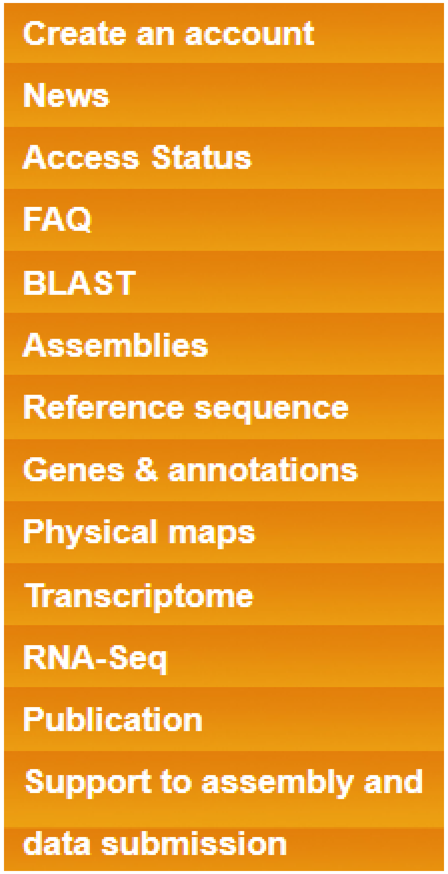
- **Gene models** performed by MIPS plant group (K. Mayer) are **publicly available**
  - Feb. 2013 version
  - Nov. 2013 version (README file included)
  - Jul. 2014 (version v2.2)

Major changes are:
a.) we re-named the genome assembly scaffolds from the old e.g ">10" identifiers to ">ta_iwgsc_1al_v2_10" identifiers for the fasta files of CLEANED and repeat-masked genome sequences and adapted the ids in the annotation GTF files accordingly.
b.) we fixed an issue with missing stop codons in the gene prediction fasta and GTF files
NO structural changes were made between v2.1 and v2.2 annotation, all gene identifiers remain stable, so this update can be considered cosmetic and mainly attributed to better user convenience.
Re-named genome assembly: genome_assembly/genome_arm_assemblies_CLEANED/ and genome_assembly/genome_arm_assemblies_CLEANED_REPMASKED/
gene predictions incl. changelog, README, ...: genePrediction_v2.2/

- **Genome Zipper** performed by MIPS plant group (K. Mayer) is **publicly available**
  - GenomeZipper_output_README.doc (35.00 kB)
  - v2: excel , tab formats
  - v3: csv , excel , tab formats
  - v4: excel , tab formats
  - v4.2: excel , tab formats
  - v5: excel , tab formats
- **POPSEQ** performed by IPK (N. Stein) is **publicly available**
  - Mar. 2014 version (README file included)
- 1AS sequence model (T. Wicker) is publicly available for download.

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

Sequence Repository



- Create an account
- News
- Access Status
- FAQ
- BLAST
- Assemblies
- Reference sequence
- Genes & annotations
- Physical maps
- Transcriptome
- ✓ RNA-Seq
- Publication
- Support to assembly and data submission

RNA-Seq data from **INRA GDEC (L. Pingault, E. Paux)** are **publicly** available to download .

Deep transcriptome sequencing was conducted on two RNA-Seq library types:

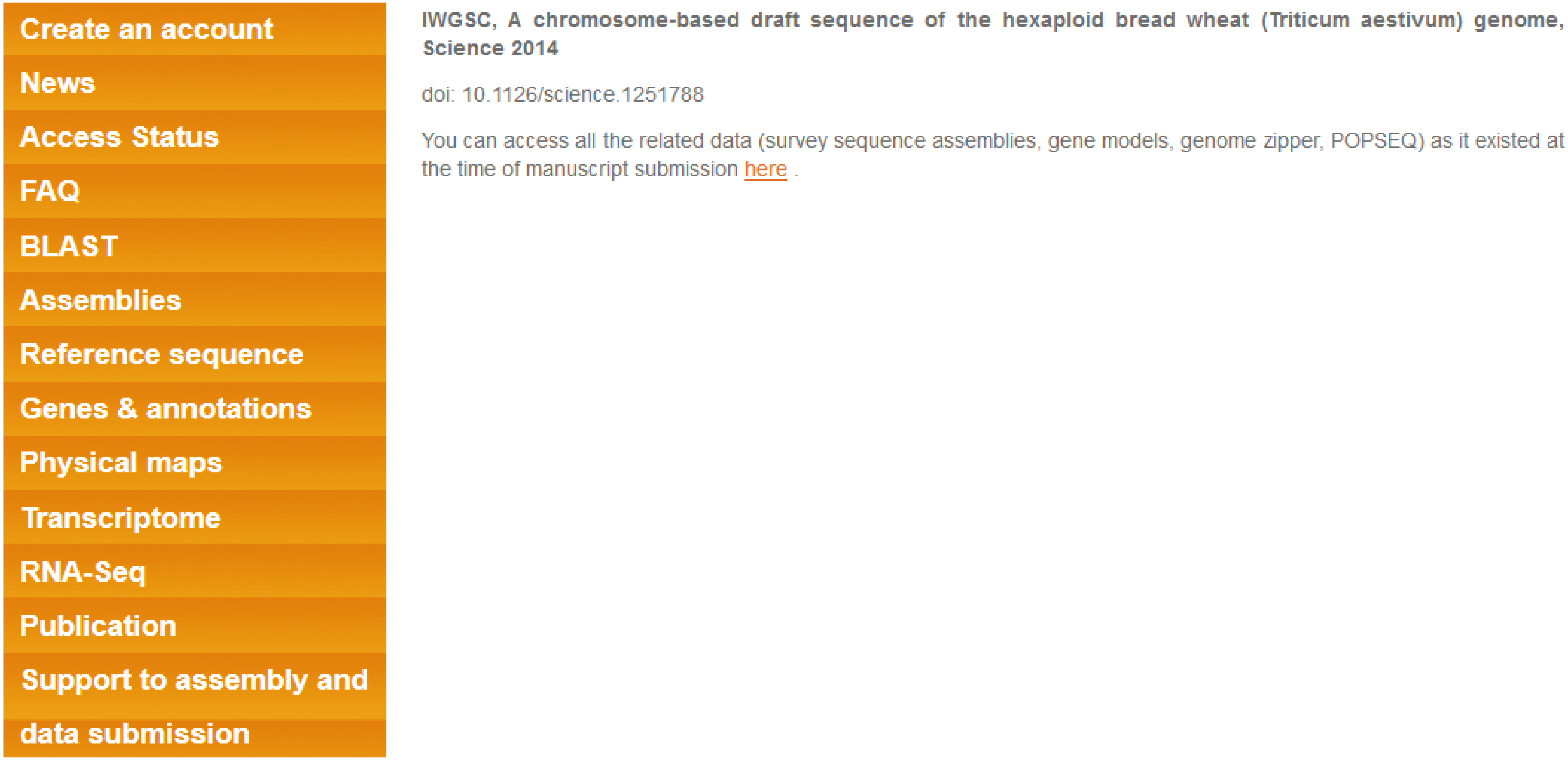
(i) **non-oriented library** (TruSeq, Illumina) sequenced on Illumina HiSeq2000 2x100bp (PE) for 15 different conditions corresponding to five wheat organs (root, leaf, stem, spike, grain) at three developmental stages each in duplicates

(ii) **oriented library** (ScriptSeq, Epicentre) sequenced on Illumina HiSeq2000 1x100bp (SE) for five conditions corresponding to five wheat organs (root, leaf, stem, spike, grain) without duplicates.

Mapping RNA-Seq reads to the chromosome 3B reference sequence allowed to validate gene prediction, as well as to identify a significant number of novel transcribed regions in which no gene structure was predicted and that might correspond to long non-coding RNAs. Transcription profiles, alternative splicing and expression breadth were also studied, providing new insights into the structural and functional compartmentalization of the wheat chromosome 3B. In combination with the IWGSC survey sequences, these data also proved to be extremely useful to decipher the specific evolutionary history of wheat chromosomes as well as to study the relative expression of homoeologous and paralogous copies of wheat genes.

```
[ ]   READ_ME
[DIR] grain_Z71_rep1/
[DIR] grain_Z71_rep2/
[DIR] grain_Z75_rep1/
[DIR] grain_Z75_rep2/
[DIR] grain_Z85_rep1/
[DIR] grain_Z85_rep2/
[DIR] leaf_Z10_rep1/
[DIR] leaf_Z10_rep2/
[DIR] leaf_Z23_rep1/
[DIR] leaf_Z23_rep2/
[DIR] leaf_Z71_rep1/
[DIR] leaf_Z71_rep2/
[DIR] root_Z10_rep1/
[DIR] root_Z10_rep2/
[DIR] root_Z13_rep1/
[DIR] root_Z13_rep2/
[DIR] root_Z39_rep1/
[DIR] root_Z39_rep2/
[DIR] spike_Z32_rep1/
[DIR] spike_Z32_rep2/
[DIR] spike_Z39_rep1/
[DIR] spike_Z39_rep2/
[DIR] spike_Z65_rep1/
[DIR] spike_Z65_rep2/
[DIR] stem_Z30_rep1/
[DIR] stem_Z30_rep2/
[DIR] stem_Z32_rep1/
[DIR] stem_Z32_rep2/
[DIR] stem_Z65_rep1/
[DIR] stem_Z65_rep2/
```

Sequence Repository

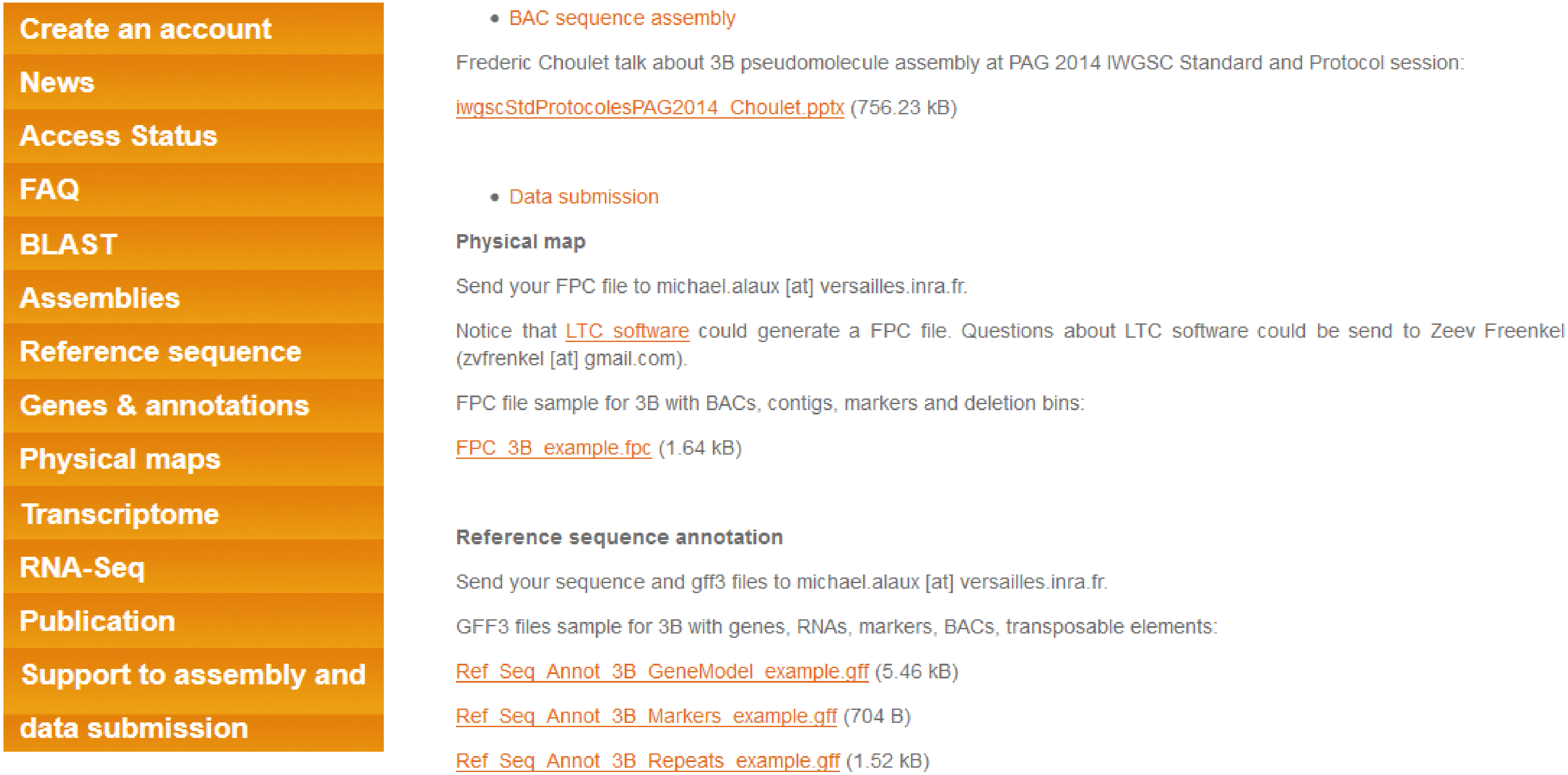


IWGSC, A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome, Science 2014

doi: 10.1126/science.1251788

You can access all the related data (survey sequence assemblies, gene models, genome zipper, POPSEQ) as it existed at the time of manuscript submission here .

Sequence Repository



- Create an account
- News
- Access Status
- FAQ
- BLAST
- Assemblies
- Reference sequence
- Genes & annotations
- Physical maps
- Transcriptome
- RNA-Seq
- Publication
- ✓ Support to assembly and data submission

- BAC sequence assembly

Frederic Choulet talk about 3B pseudomolecule assembly at PAG 2014 IWGSC Standard and Protocol session:

iwgscStdProtocolesPAG2014_Choulet.pptx (756.23 kB)

- Data submission

Physical map

Send your FPC file to michael.alaux [at] versailles.inra.fr.

Notice that LTC software could generate a FPC file. Questions about LTC software could be send to Zeev Freenkel (zvfrenkel [at] gmail.com).

FPC file sample for 3B with BACs, contigs, markers and deletion bins:

FPC_3B_example.fpc (1.64 kB)

Reference sequence annotation

Send your sequence and gff3 files to michael.alaux [at] versailles.inra.fr.

GFF3 files sample for 3B with genes, RNAs, markers, BACs, transposable elements:

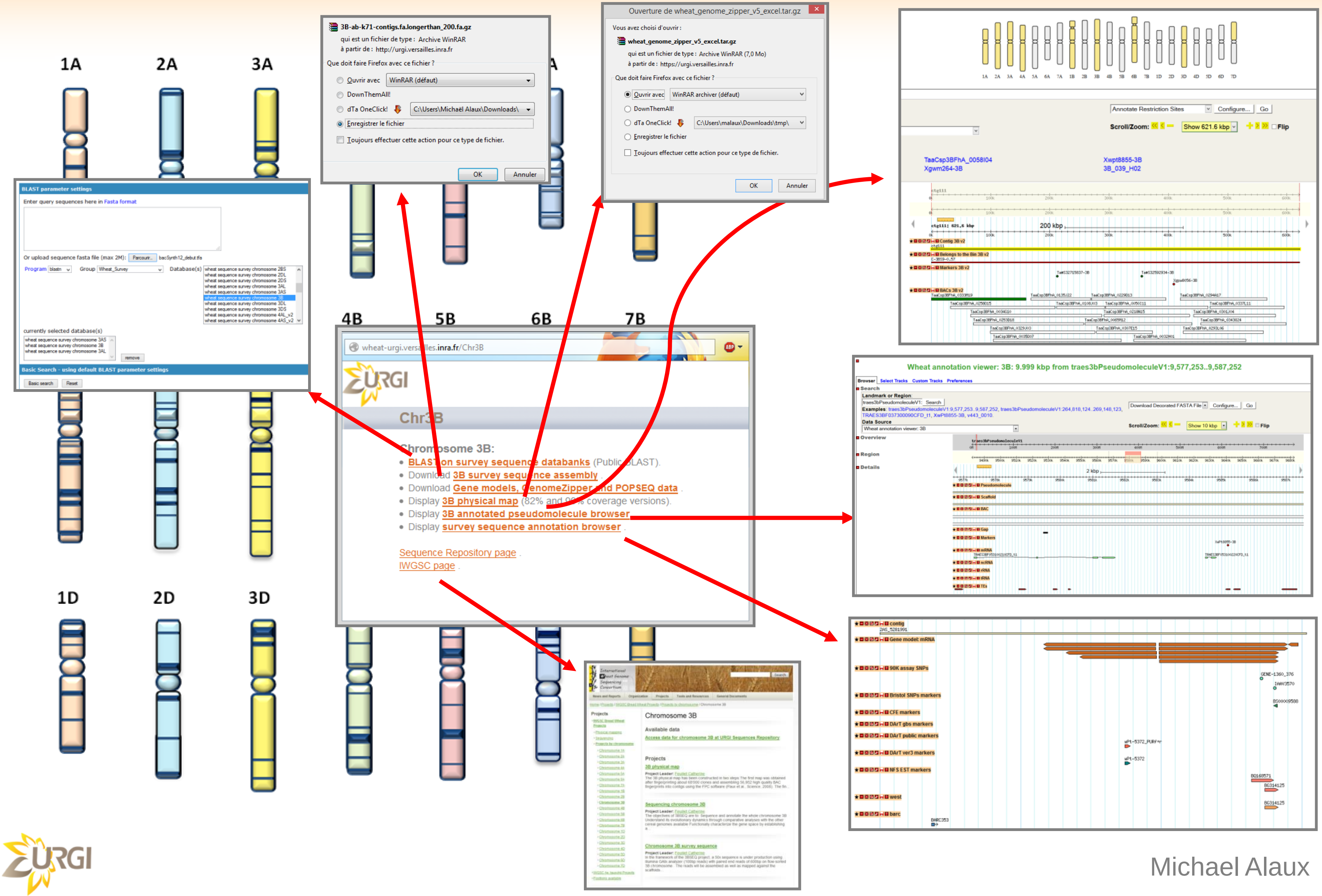
Ref_Seq_Annot_3B_GeneModel_example.gff (5.46 kB)

Ref_Seq_Annot_3B_Markers_example.gff (704 B)

Ref_Seq_Annot_3B_Repeats_example.gff (1.52 kB)

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

Sequence Repository: 3B example



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_Survey Database(s):
- wheat sequence survey chromosome 2BS
- wheat sequence survey chromosome 2DL
- wheat sequence survey chromosome 2DS
- wheat sequence survey chromosome 3AL
- wheat sequence survey chromosome 3AS
- wheat sequence survey chromosome 3B
- wheat sequence survey chromosome 3DL
- wheat sequence survey chromosome 3DS
- wheat sequence survey chromosome 4AL_v2
- wheat sequence survey chromosome 4AS_v2

currently selected database(s)
- wheat sequence survey chromosome 3AS
- wheat sequence survey chromosome 3B
- wheat sequence survey chromosome 3AL

Basic Search - using default BLAST parameter settings

3B-ab-k71-contigs.fa.longerthan_200.fa.gz

qui est un fichier de type : Archive WinRAR
à partir de : http://urgi.versailles.inra.fr

Que doit faire Firefox avec ce fichier ?

○ Ouvrir avec WinRAR (défaut)
○ DownThemAll!
○ dTa OneClick! C:\Users\Michaël Alaux\Downloads\
● Enregistrer le fichier
☐ Toujours effectuer cette action pour ce type de fichier.

Ouverture de wheat_genome_zipper_v5_excel.tar.gz

Vous avez choisi d'ouvrir :

wheat_genome_zipper_v5_excel.tar.gz

qui est un fichier de type : Archive WinRAR (7,0 Mo)
à partir de : https://urgi.versailles.inra.fr

Que doit faire Firefox avec ce fichier ?

● Ouvrir avec WinRAR archiver (défaut)
○ DownThemAll!
○ dTa OneClick! C:\Users\malaux\Downloads\tmp\
○ Enregistrer le fichier
☐ Toujours effectuer cette action pour ce type de fichier.

wheat-urgi.versailles.inra.fr/Chr3B

Chr3B

Chromosome 3B:
- **BLAST on survey sequence databanks** (Public BLAST).
- Download **3B survey sequence assembly** .
- Download **Gene models, GenomeZipper and POPSEQ data** .
- Display **3B physical map** (82% and 90% coverage versions).
- Display **3B annotated pseudomolecule browser**
- Display **survey sequence annotation browser** .

Sequence Repository page .
IWGSC page .

Wheat annotation viewer: 3B: 9.999 kbp from traes3bPseudomoleculeV1:9,577,253..9,587,252

Browser Select Tracks Custom Tracks Preferences

Search

Landmark or Region: traes3bPseudomoleculeV1: Search

Examples: traes3bPseudomoleculeV1:9,577,253..9,587,252, traes3bPseudomoleculeV1:264,818,124..269,148,123, TRAES3BF037300090CFD_t1, XwPt8855-3B, v443_0010.

Data Source: Wheat annotation viewer: 3B

International Wheat Genome Sequencing Consortium

Home / Projects / IWGSC Bread Wheat Projects / Projects by chromosome / Chromosome 3B

Projects
- IWGSC Bread Wheat Projects
  - Physical mapping
  - Sequencing
  - Projects by chromosome

Chromosome 3B

Available data

Access data for chromosome 3B at URGI Sequences Repository

Projects

3B physical map

Project Leader: Feuillet Catherine
The 3B physical map has been constructed in two steps: The first map was obtained after fingerprinting about 68'000 clones and assembling 56,952 high quality BAC fingerprints into contigs using the FPC software (Paux et al., Science, 2008). The fin...

Sequencing chromosome 3B

Project Leader: Feuillet Catherine
The objectives of 3BSEQ are to: Sequence and annotate the whole chromosome 3B Understand its evolutionary dynamics through comparative analyses with the other cereal genomes available Functionally characterize the gene space by establishing a ...

Chromosome 3B survey sequence

Project Leader: Feuillet Catherine
In the framework of the 3BSEQ project, a 50x sequence is under production using Illumina GAIIx analyzer (100bp reads) with paired end reads of 600bp on flow-sorted 3B chromosome . The reads will be assembled as well as mapped against the scaffolds...

Michael Alaux

BLAST

Michael Alaux

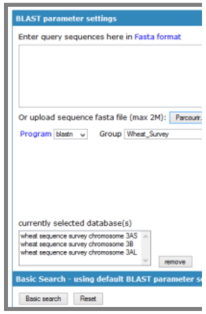
BLAST

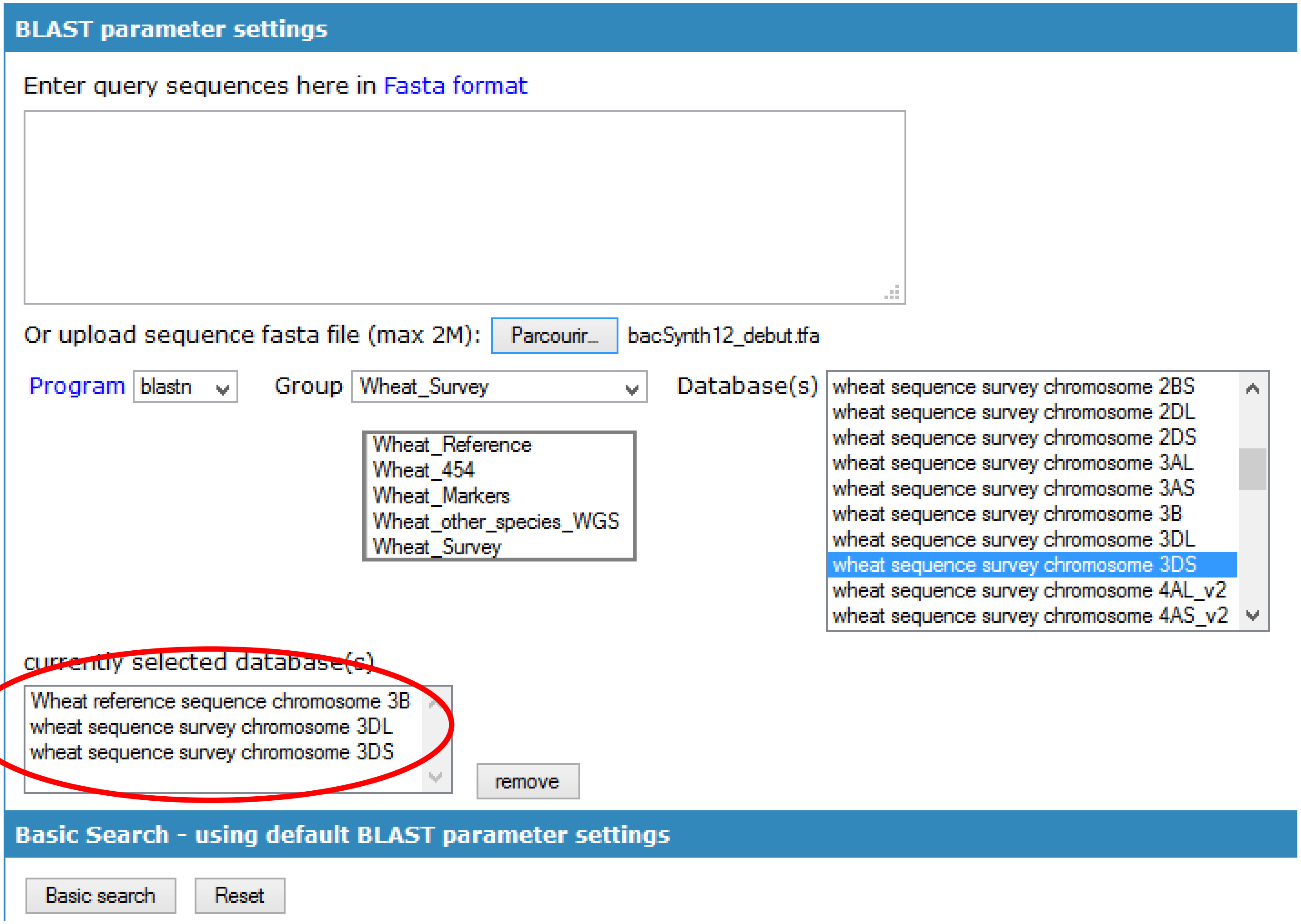
- **Public BLAST**

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

- More than 250 000 BLAST searches the past 2 years !

BLAST





BLAST parameter settings

Enter query sequences here in Fasta format

Or upload sequence fasta file (max 2M): bacSynth12_debut.tfa

Program: blastn

Group: Wheat_Survey

Wheat_Reference
Wheat_454
Wheat_Markers
Wheat_other_species_WGS
Wheat_Survey

Database(s)

wheat sequence survey chromosome 2BS
wheat sequence survey chromosome 2DL
wheat sequence survey chromosome 2DS
wheat sequence survey chromosome 3AL
wheat sequence survey chromosome 3AS
wheat sequence survey chromosome 3B
wheat sequence survey chromosome 3DL
wheat sequence survey chromosome 3DS
wheat sequence survey chromosome 4AL_v2
wheat sequence survey chromosome 4AS_v2

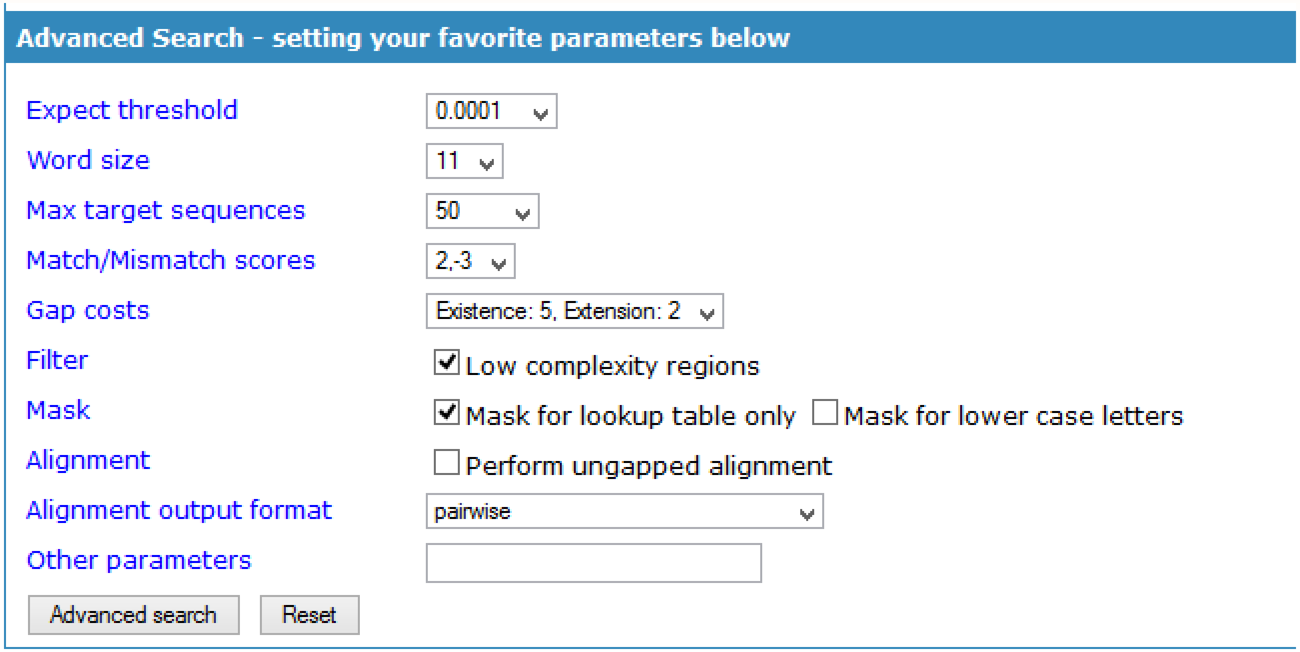
currently selected database(s)

Wheat reference sequence chromosome 3B
wheat sequence survey chromosome 3DL
wheat sequence survey chromosome 3DS

Basic Search - using default BLAST parameter settings

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

BLAST

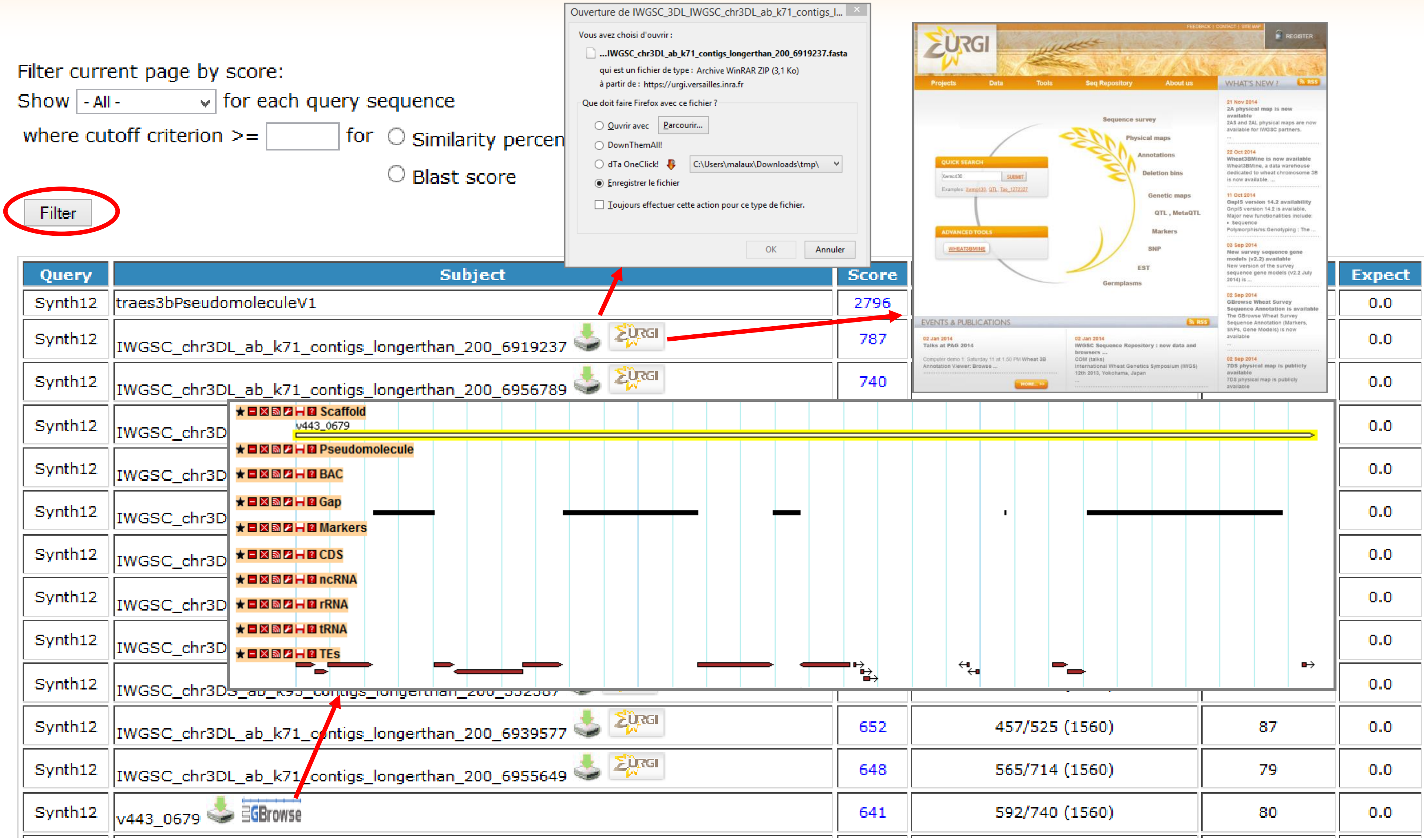


Advanced Search - setting your favorite parameters below

Expect threshold: 0.0001
Word size: 11
Max target sequences: 50
Match/Mismatch scores: 2,-3
Gap costs: Existence: 5, Extension: 2
Filter: ☑ Low complexity regions
Mask: ☑ Mask for lookup table only ☐ Mask for lower case letters
Alignment: ☐ Perform ungapped alignment
Alignment output format: pairwise
Other parameters:

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

BLAST



Filter current page by score:

Show - All - for each query sequence

where cutoff criterion >= ☐ for ○ Similarity percen ○ Blast score

Filter

Query	Subject	Score			Expect
Synth12	traes3bPseudomoleculeV1	2796			0.0
Synth12	IWGSC_chr3DL_ab_k71_contigs_longerthan_200_6919237	787			0.0
Synth12	IWGSC_chr3DL_ab_k71_contigs_longerthan_200_6956789	740			0.0
Synth12	IWGSC_chr3D				0.0
Synth12	IWGSC_chr3D				0.0
Synth12	IWGSC_chr3D				0.0
Synth12	IWGSC_chr3D				0.0
Synth12	IWGSC_chr3D				0.0
Synth12	IWGSC_chr3D				0.0
Synth12	IWGSC_chr3D				0.0
Synth12	IWGSC_chr3DL_ab_k71_contigs_longerthan_200_6939577	652	457/525 (1560)	87	0.0
Synth12	IWGSC_chr3DL_ab_k71_contigs_longerthan_200_6955649	648	565/714 (1560)	79	0.0
Synth12	v443_0679	641	592/740 (1560)	80	0.0

Ouverture de IWGSC_3DL_IWGSC_chr3DL_ab_k71_contigs_l...

Vous avez choisi d'ouvrir :

...IWGSC_chr3DL_ab_k71_contigs_longerthan_200_6919237.fasta

qui est un fichier de type : Archive WinRAR ZIP (3,1 Ko)

à partir de : https://urgi.versailles.inra.fr

Que doit faire Firefox avec ce fichier ?

○ Ouvrir avec [Parcourir...]

○ DownThemAll!

○ dTa OneClick! C:\Users\malaux\Downloads\tmp\

● Enregistrer le fichier

☐ Toujours effectuer cette action pour ce type de fichier.

[OK] [Annuler]

WHAT'S NEW ?

21 Nov 2014
2A physical map is now available
2AS and 2AL physical maps are now available for IWGSC partners.
...

22 Oct 2014
Wheat3BMine is now available
Wheat3BMine, a data warehouse dedicated to wheat chromosome 3B is now available. ...

11 Oct 2014
GnpIS version 14.2 availability
GnpIS version 14.2 is available. Major new functionalities include:
- Sequence Polymorphisms:Genotyping : The ...

03 Sep 2014
New survey sequence gene models (v2.2) available
New version of the survey sequence gene models (v2.2 July 2014) is ...

02 Sep 2014
GBrowse Wheat Survey Sequence Annotation is available
The GBrowse Wheat Survey Sequence Annotation (Markers, SNPs, Gene Models) is now available
...

02 Sep 2014
7DS physical map is publicly available
7DS physical map is publicly available

EVENTS & PUBLICATIONS

02 Jan 2014
Talks at PAG 2014
Computer demo 1: Saturday 11 at 1.50 PM **Wheat 3B Annotation Viewer: Browse ...**

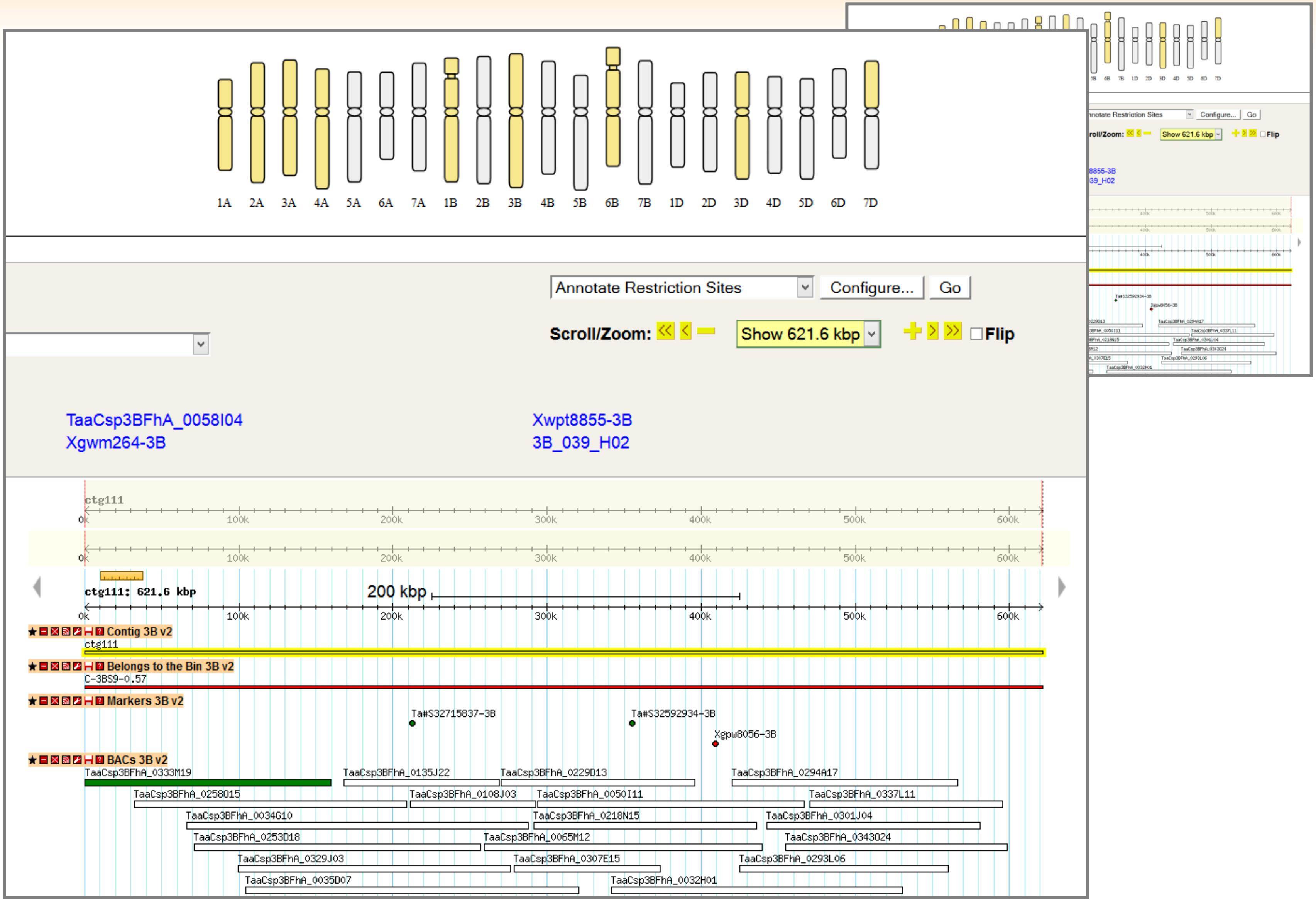
02 Jan 2014
IWGSC Sequence Repository : new data and browsers ...
COM (talks)
International Wheat Genetics Symposium (IWGS) 12th 2013, Yokohama, Japan
...

QUICK SEARCH: Xwmc430 [SUBMIT]
Examples: Xwmc430, QTL, Tae_1272327

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

Physical maps

Physical maps



Annotate Restriction Sites | Configure... | Go

Scroll/Zoom: Show 621.6 kbp ☐ **Flip**

TaaCsp3BFhA_0058I04
Xgwm264-3B

Xwpt8855-3B
3B_039_H02

ctg111

ctg111: 621.6 kbp

200 kbp

Contig 3B v2
ctg111

Belongs to the Bin 3B v2
C-3BS9-0.57

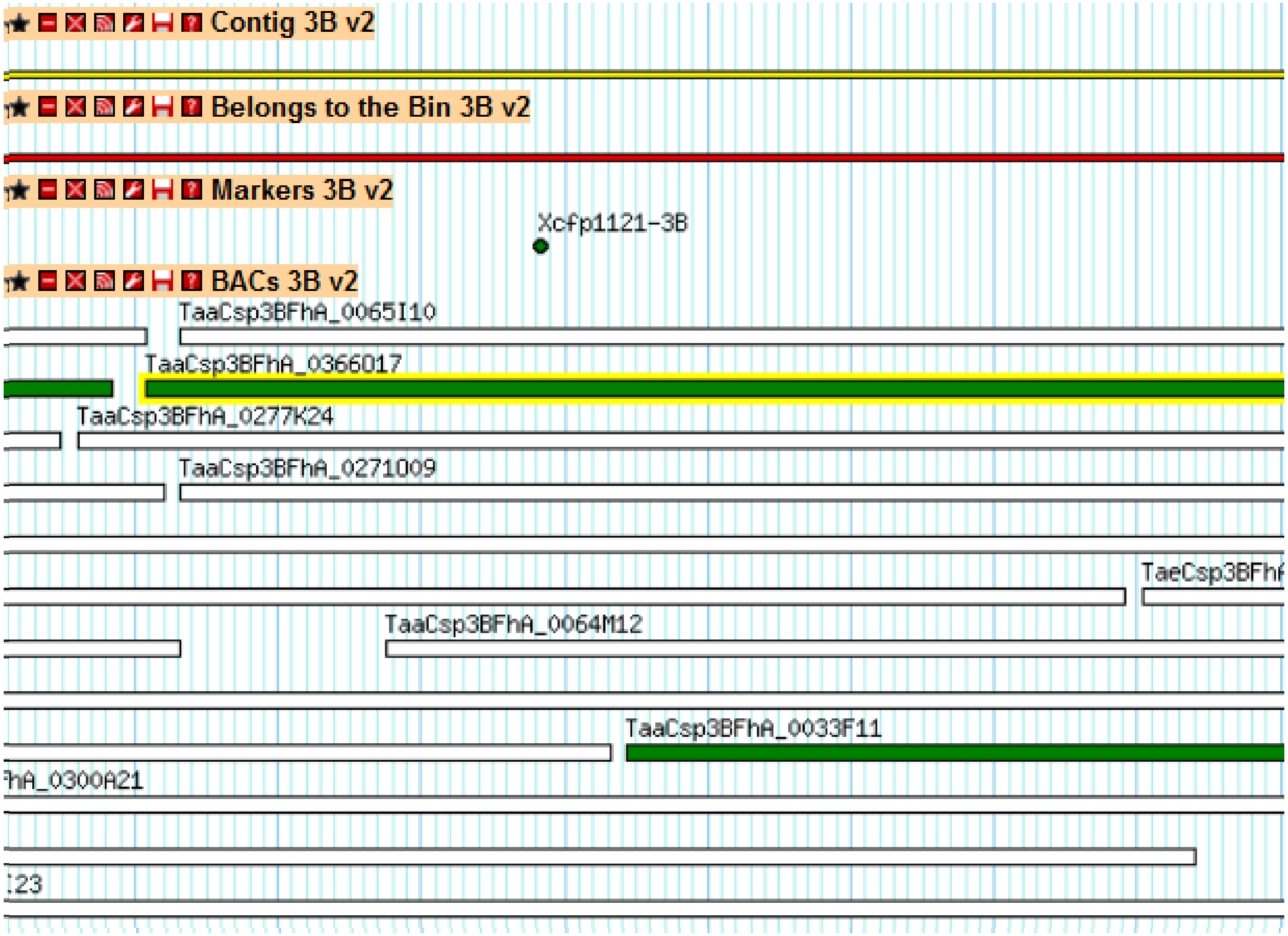
Markers 3B v2
Ta#S32715837-3B
Ta#S32592934-3B
Xgpw8056-3B

BACs 3B v2
TaaCsp3BFhA_0333M19
TaaCsp3BFhA_0135J22
TaaCsp3BFhA_0229D13
TaaCsp3BFhA_0294A17
TaaCsp3BFhA_0258O15
TaaCsp3BFhA_0108J03
TaaCsp3BFhA_0050I11
TaaCsp3BFhA_0337L11
TaaCsp3BFhA_0034G10
TaaCsp3BFhA_0218N15
TaaCsp3BFhA_0301J04
TaaCsp3BFhA_0253D18
TaaCsp3BFhA_0065M12
TaaCsp3BFhA_0343O24
TaaCsp3BFhA_0329J03
TaaCsp3BFhA_0307E15
TaaCsp3BFhA_0293L06
TaaCsp3BFhA_0035D07
TaaCsp3BFhA_0032H01

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

Michael Alaux

Physical maps



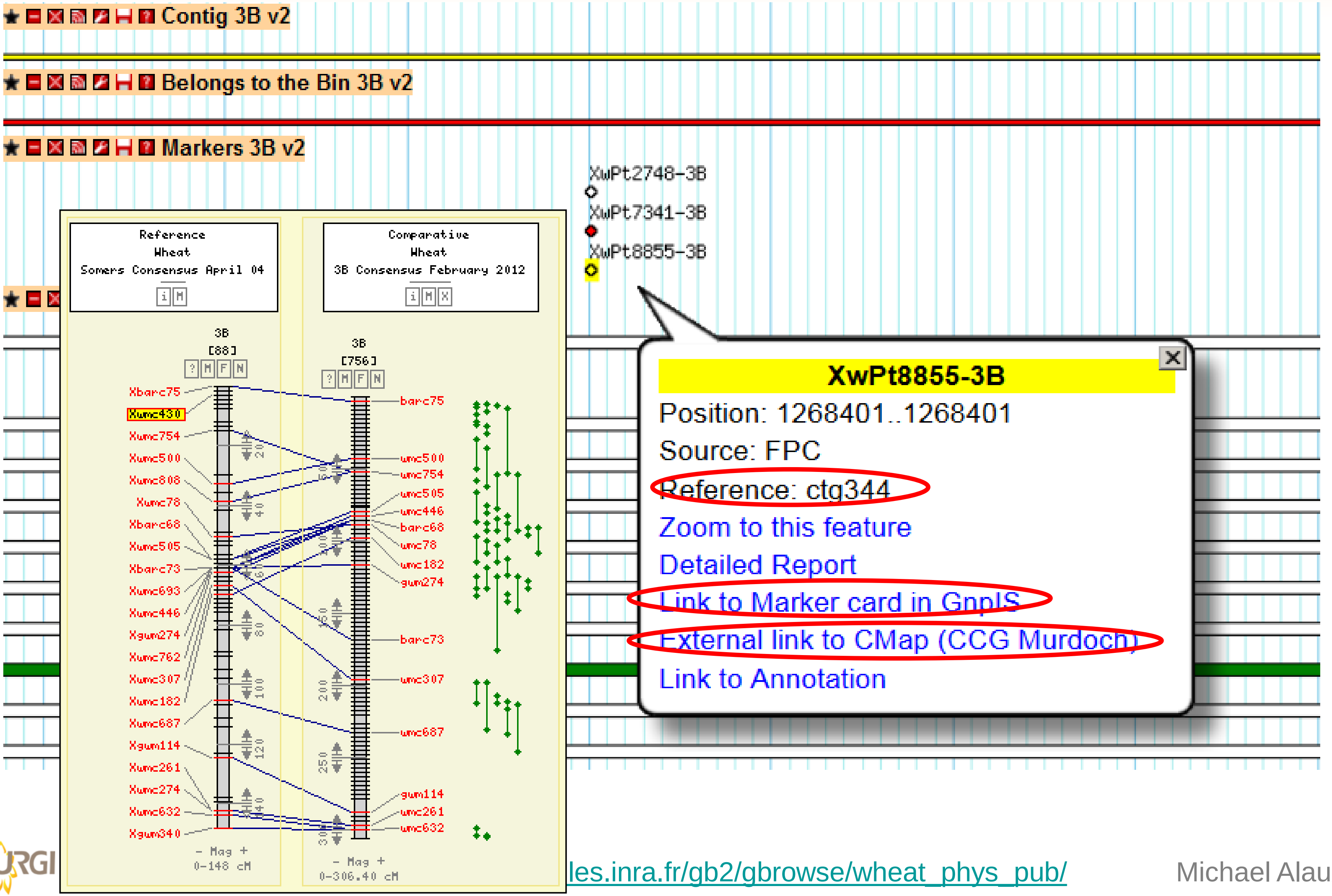
Physical contig

Deletion bin

Markers

BACs

Physical maps



XwPt8855-3B

Position: 1268401..1268401

Source: FPC

Reference: ctg344

Zoom to this feature

Detailed Report

Link to Marker card in GnpIS

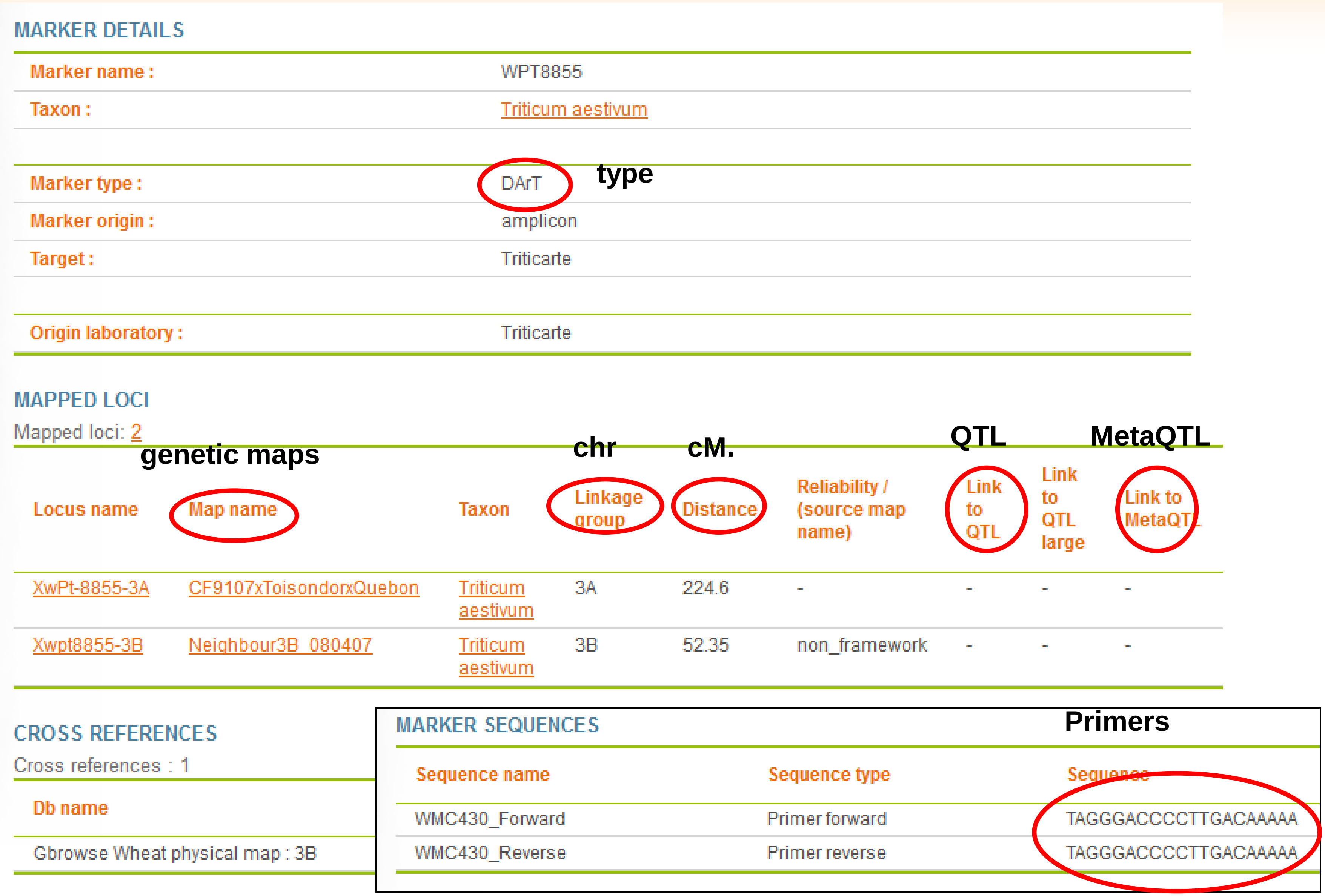
External link to CMap (CCG Murdoch)

Link to Annotation

les.inra.fr/gb2/gbrowse/wheat_phys_pub/

Michael Alaux

Physical maps



type

genetic maps

chr

cM.

QTL

MetaQTL

Primers

MARKER DETAILS

Marker name :	WPT8855
Taxon :	Triticum aestivum
Marker type :	DArT
Marker origin :	amplicon
Target :	Triticarte
Origin laboratory :	Triticarte

MAPPED LOCI

Mapped loci: 2

Locus name	Map name	Taxon	Linkage group	Distance	Reliability / (source map name)	Link to QTL	Link to QTL large	Link to MetaQTL
XwPt-8855-3A	CF9107xToisondorxQuebon	Triticum aestivum	3A	224.6	-	-	-	-
Xwpt8855-3B	Neighbour3B_080407	Triticum aestivum	3B	52.35	non_framework	-	-	-

CROSS REFERENCES

Cross references : 1

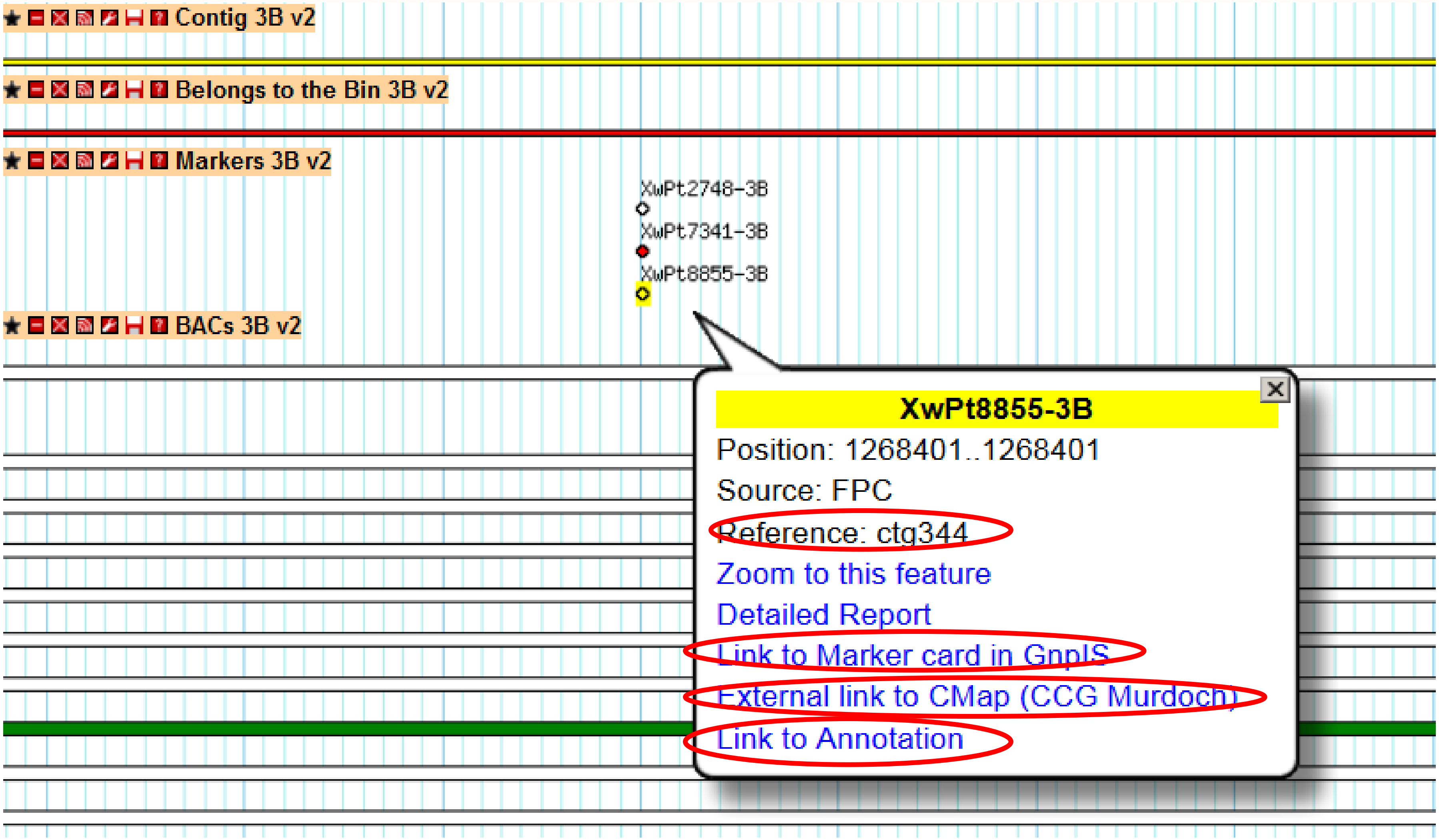
Db name
Gbrowse Wheat physical map : 3B

MARKER SEQUENCES

Sequence name	Sequence type	Sequence
WMC430_Forward	Primer forward	TAGGGACCCCTTGACAAAAA
WMC430_Reverse	Primer reverse	TAGGGACCCCTTGACAAAAA

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

Physical maps



XwPt8855-3B

Position: 1268401..1268401

Source: FPC

Reference: ctg344

Zoom to this feature

Detailed Report

Link to Marker card in GnpIS

External link to CMap (CCG Murdoch)

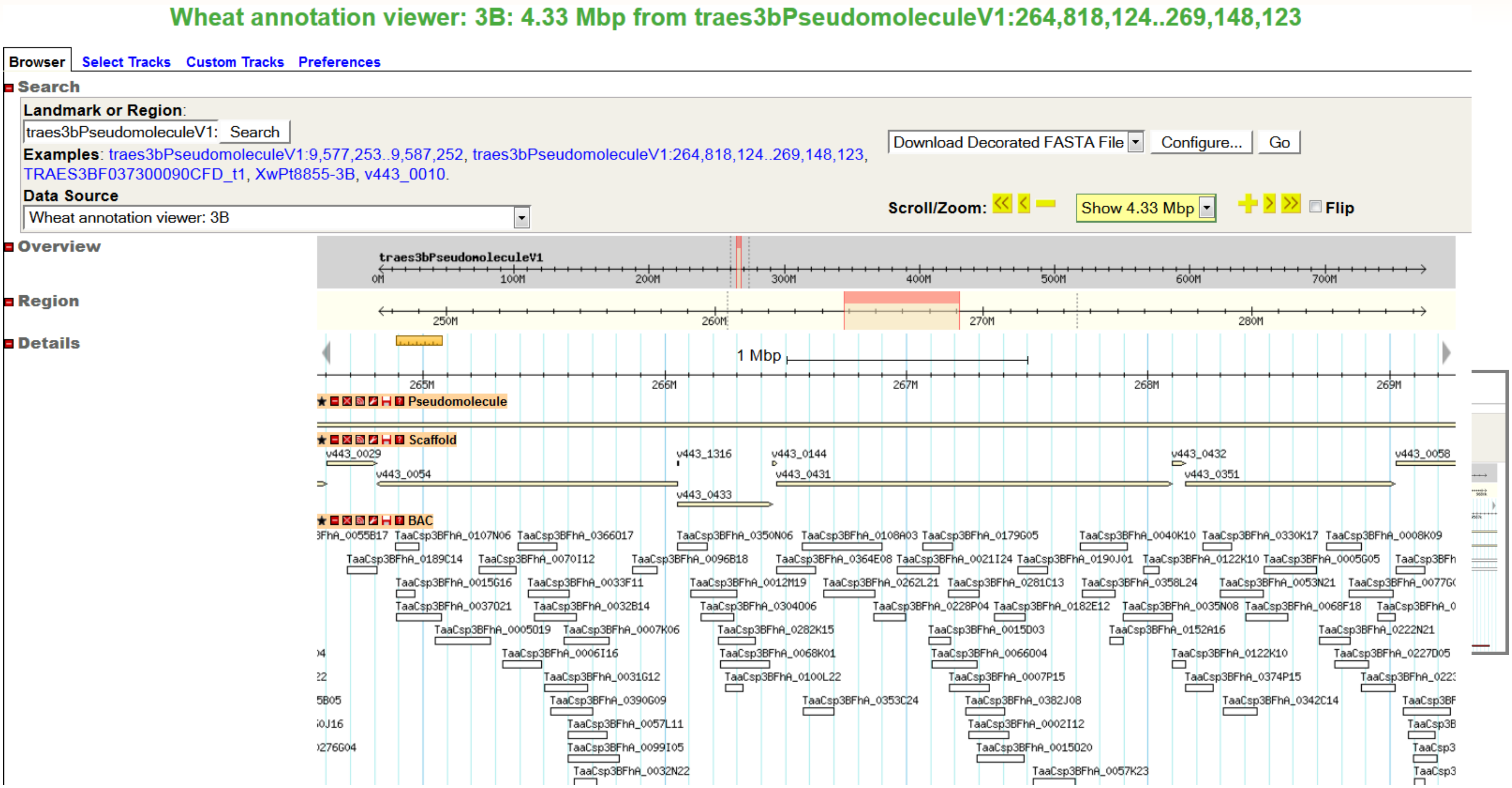
Link to Annotation

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

3B reference sequence annotation

Michael Alaux

3B annotation



Wheat annotation viewer: 3B: 4.33 Mbp from traes3bPseudomoleculeV1:264,818,124..269,148,123

Landmark or Region: traes3bPseudomoleculeV1:

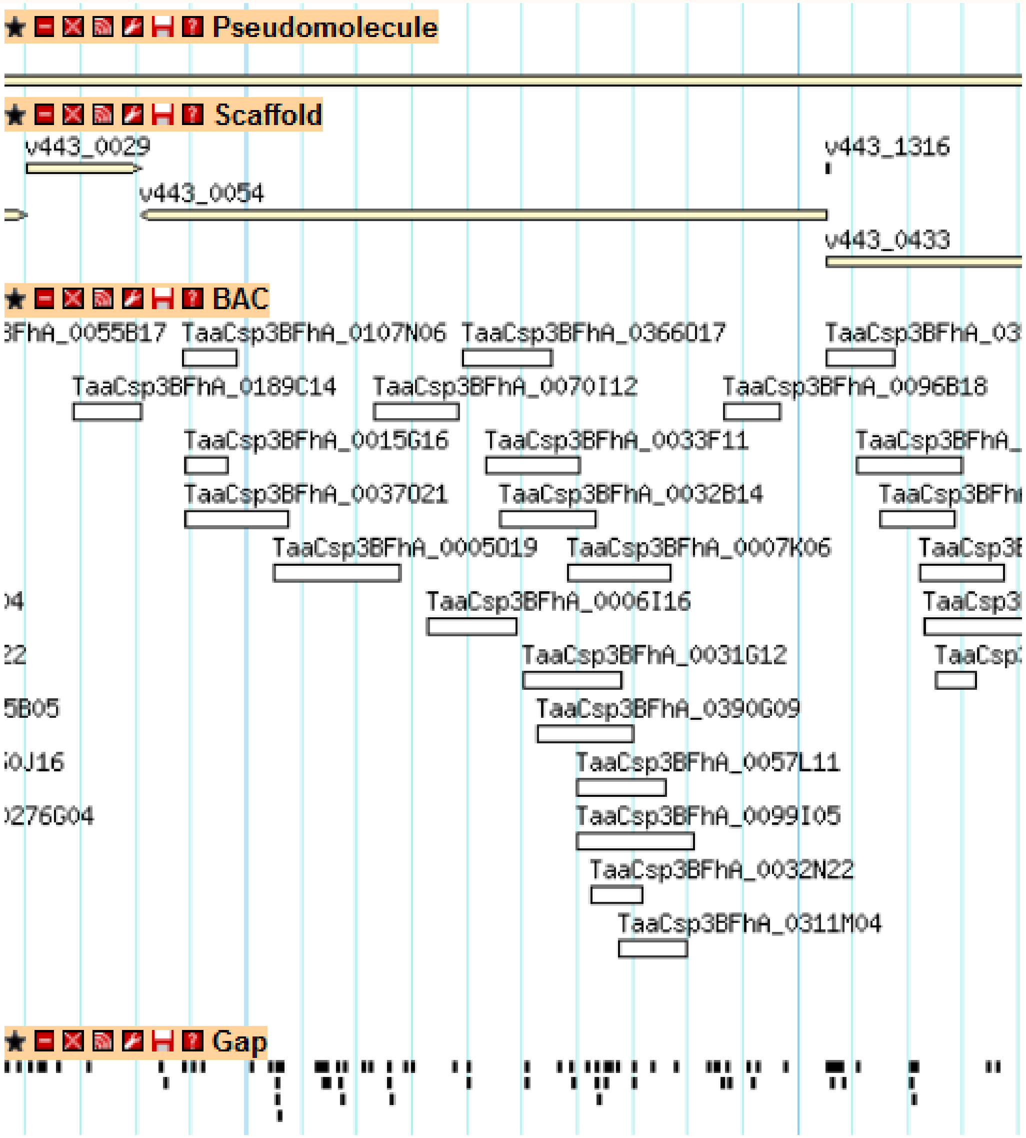
Examples: traes3bPseudomoleculeV1:9,577,253..9,587,252, traes3bPseudomoleculeV1:264,818,124..269,148,123, TRAES3BF037300090CFD_t1, XwPt8855-3B, v443_0010.

Data Source: Wheat annotation viewer: 3B

Scroll/Zoom: Show 4.33 Mbp ☐ Flip

https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_annot_3B

3B annotation



Pseudomolecule

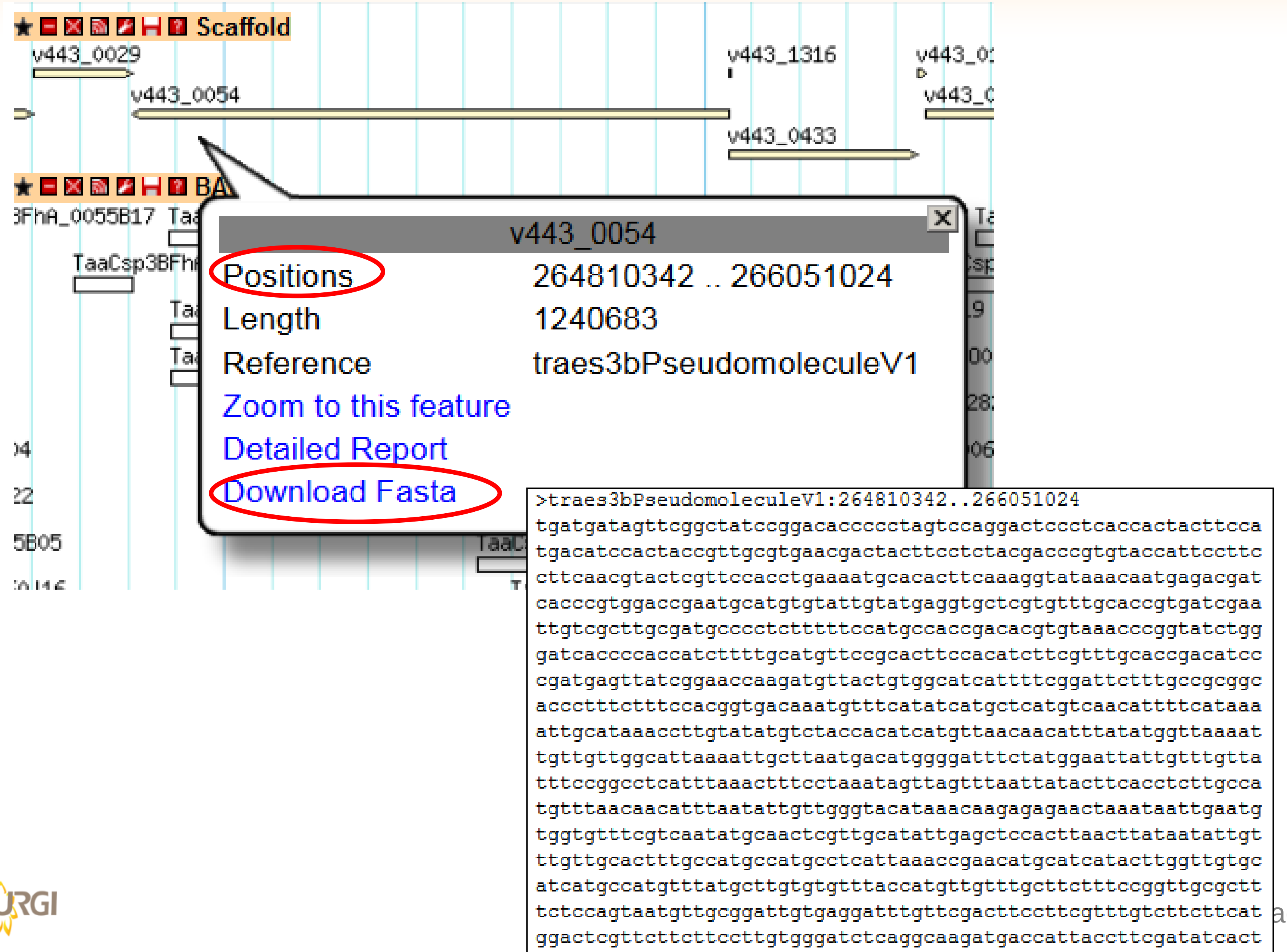
Scaffolds

BACs

Gaps

https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_annot_3B

3B annotation: Scaffold



v443_0054

Positions	264810342 .. 266051024
Length	1240683
Reference	traes3bPseudomoleculeV1

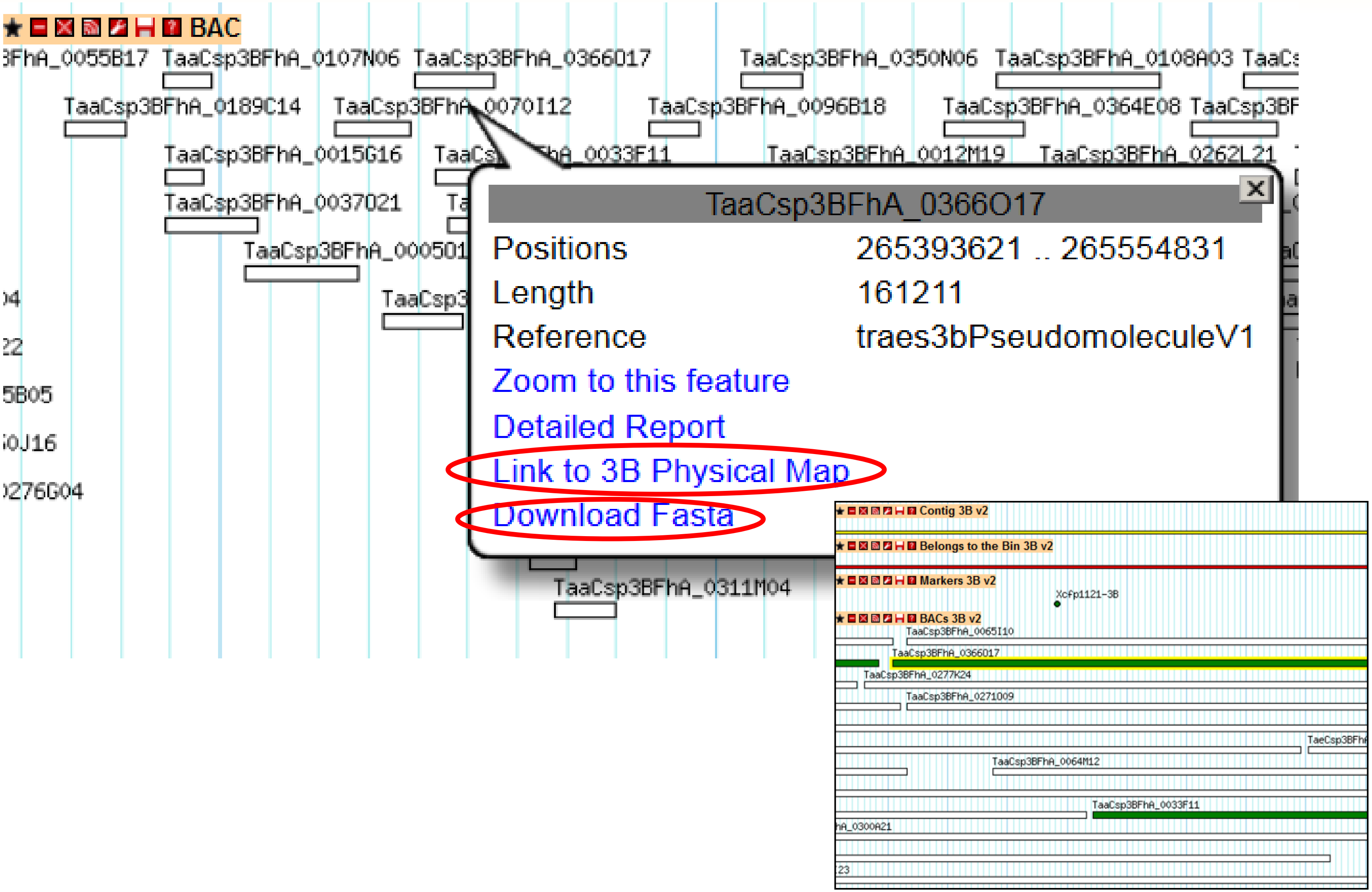
Zoom to this feature

Detailed Report

Download Fasta

```
>traes3bPseudomoleculeV1:264810342..266051024
tgatgatagttcggctatccggacacccccctagtccaggactccctcaccactacttcca
tgacatccactaccgttgcgtgaacgactacttcctctacgacccgtgtaccattccttc
cttcaacgtactcgttccacctgaaaatgcacacttcaaaggtataaacaatgagacgat
cacccgtggaccgaatgcatgtgtattgtatgaggtgctcgtgtttgcaccgtgatcgaa
ttgtcgcttgcgatgcccctctttttccatgccaccgacacgtgtaaacccggtatctgg
gatcaccccaccatcttttgcatgttccgcacttccacatcttcgtttgcaccgacatcc
cgatgagttatcggaaccaagatgttactgtggcatcattttcggattctttgccgcggc
accctttctttccacggtgacaaatgtttcatatcatgctcatgtcaacattttcataaa
attgcataaaccttgtatatgtctaccacatcatgttaacaacatttatatggttaaaat
tgttgttggcattaaaattgcttaatgacatggggatttctatggaattattgtttgtta
tttccggcctcatttaaactttcctaaatagttagtttaattatacttcacctcttgcca
tgtttaacaacatttaatattgttgggtacataaacaagagagaactaaataattgaatg
tggtgtttcgtcaatatgcaactcgttgcatattgagctccacttaacttataatattgt
ttgttgcactttgccatgccatgcctcattaaaccgaacatgcatcatacttggttgtgc
atcatgccatgtttatgcttgtgtgtttaccatgttgtttgcttctttccggttgcgctt
tctccagtaatgttgcggattgtgaggatttgttcgacttccttcgtttgtcttcttcat
ggactcgttcttcttccttgtgggatctcaggcaagatgaccattaccttcgatatcact
```

3B annotation: BAC



TaaCsp3BFhA_0366O17

Positions	265393621 .. 265554831
Length	161211
Reference	traes3bPseudomoleculeV1

Zoom to this feature

Detailed Report

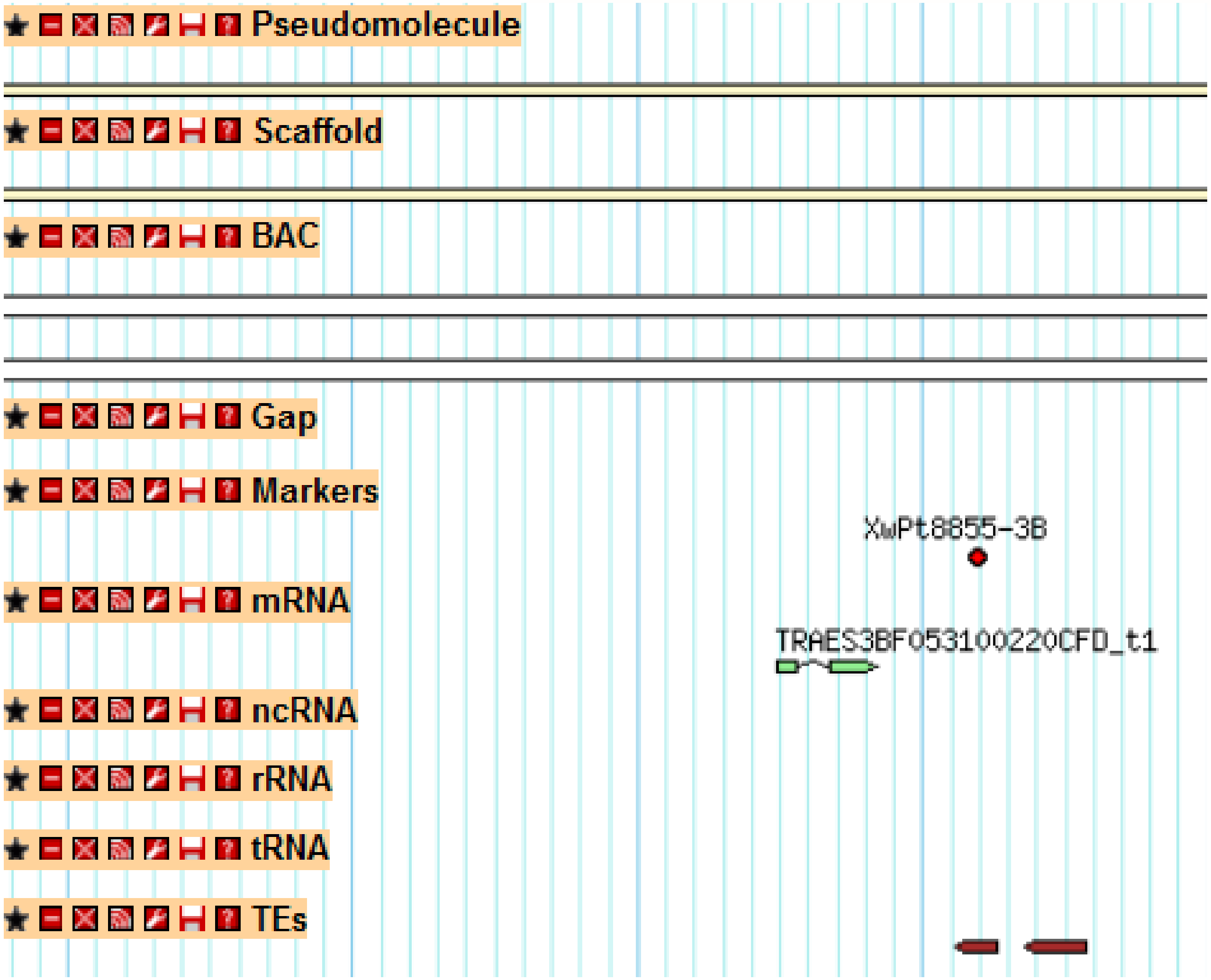
Link to 3B Physical Map

Download Fasta

https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_annot_3B

Michael Alaux

3B annotation



Pseudomolecule

Scaffold

BAC

Gap

Markers

XwPt8855-3B

mRNA

TRAES3BF053100220CFD_t1

ncRNA

rRNA

tRNA

TEs

Markers

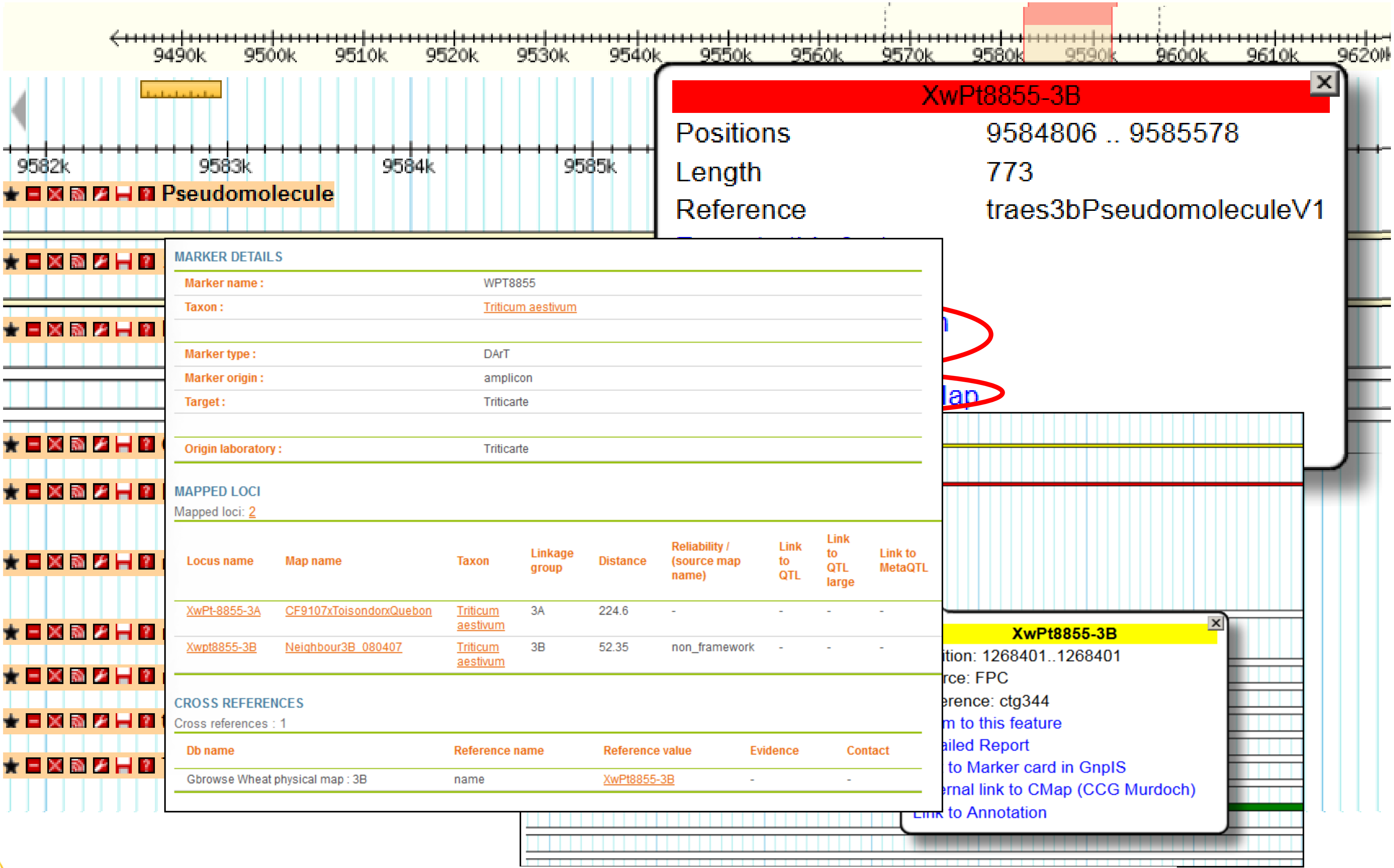
Genes

TE

https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_annot_3B

Michael Alaux

3B annotation: Marker



XwPt8855-3B

Positions	9584806 .. 9585578
Length	773
Reference	traes3bPseudomoleculeV1

MARKER DETAILS

Marker name :	WPT8855
Taxon :	Triticum aestivum
Marker type :	DArT
Marker origin :	amplicon
Target :	Triticarte
Origin laboratory :	Triticarte

MAPPED LOCI

Mapped loci: 2

Locus name	Map name	Taxon	Linkage group	Distance	Reliability / (source map name)	Link to QTL	Link to QTL large	Link to MetaQTL
XwPt-8855-3A	CF9107xToisondorxQuebon	Triticum aestivum	3A	224.6	-	-	-	-
Xwpt8855-3B	Neighbour3B_080407	Triticum aestivum	3B	52.35	non_framework	-	-	-

CROSS REFERENCES

Cross references : 1

Db name	Reference name	Reference value	Evidence	Contact
Gbrowse Wheat physical map : 3B	name	XwPt8855-3B	-	-

XwPt8855-3B

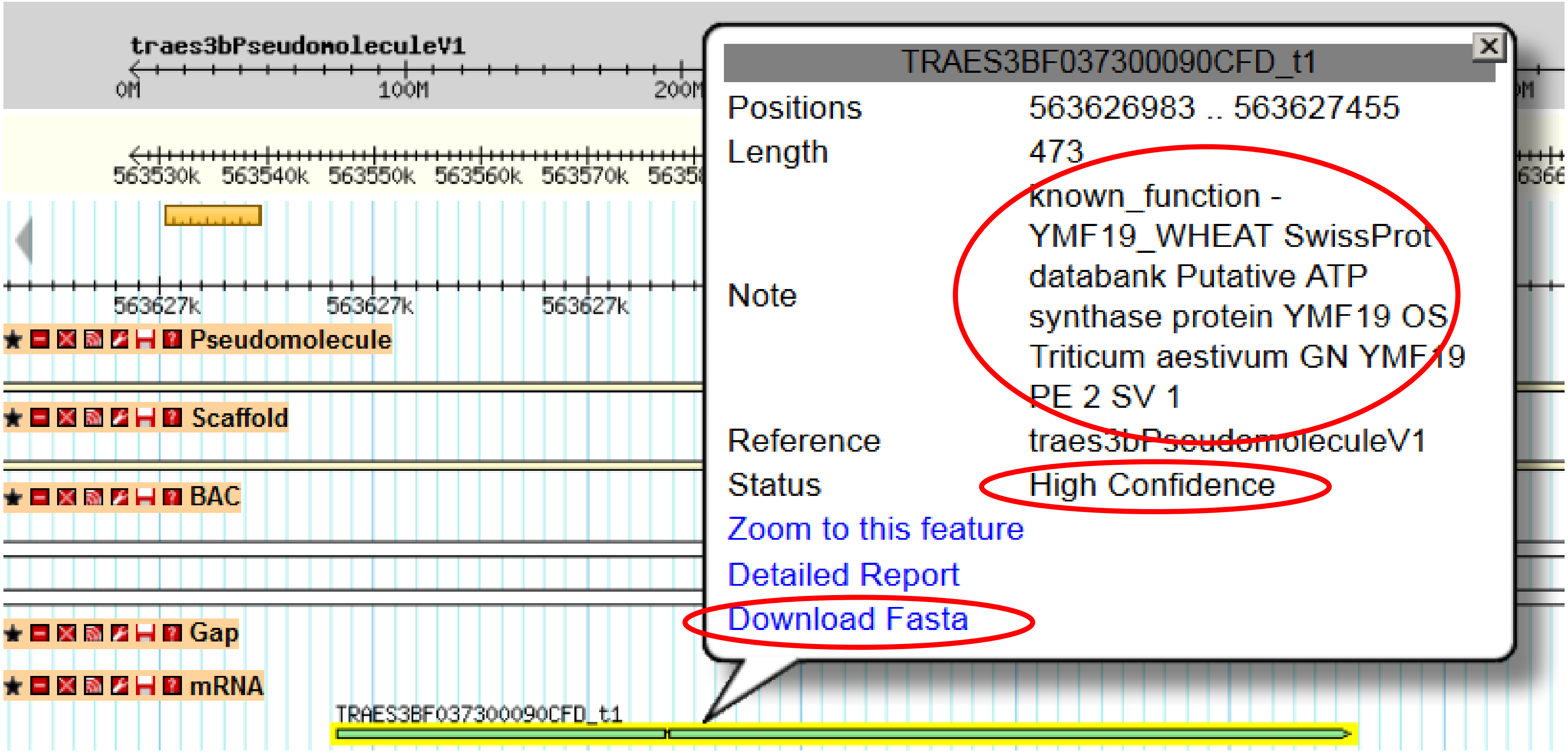
ition: 1268401..1268401
rce: FPC
erence: ctg344
m to this feature
ailed Report
to Marker card in GnpIS
ernal link to CMap (CCG Murdoch)
Link to Annotation

URGI

https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_annot_3B

Michael Alaux

3B annotation: Gene



TRAES3BF037300090CFD_t1

Positions	563626983 .. 563627455
Length	473
Note	known_function - YMF19_WHEAT SwissProt databank Putative ATP synthase protein YMF19 OS Triticum aestivum GN YMF19 PE 2 SV 1
Reference	traes3bPseudomoleculeV1
Status	High Confidence

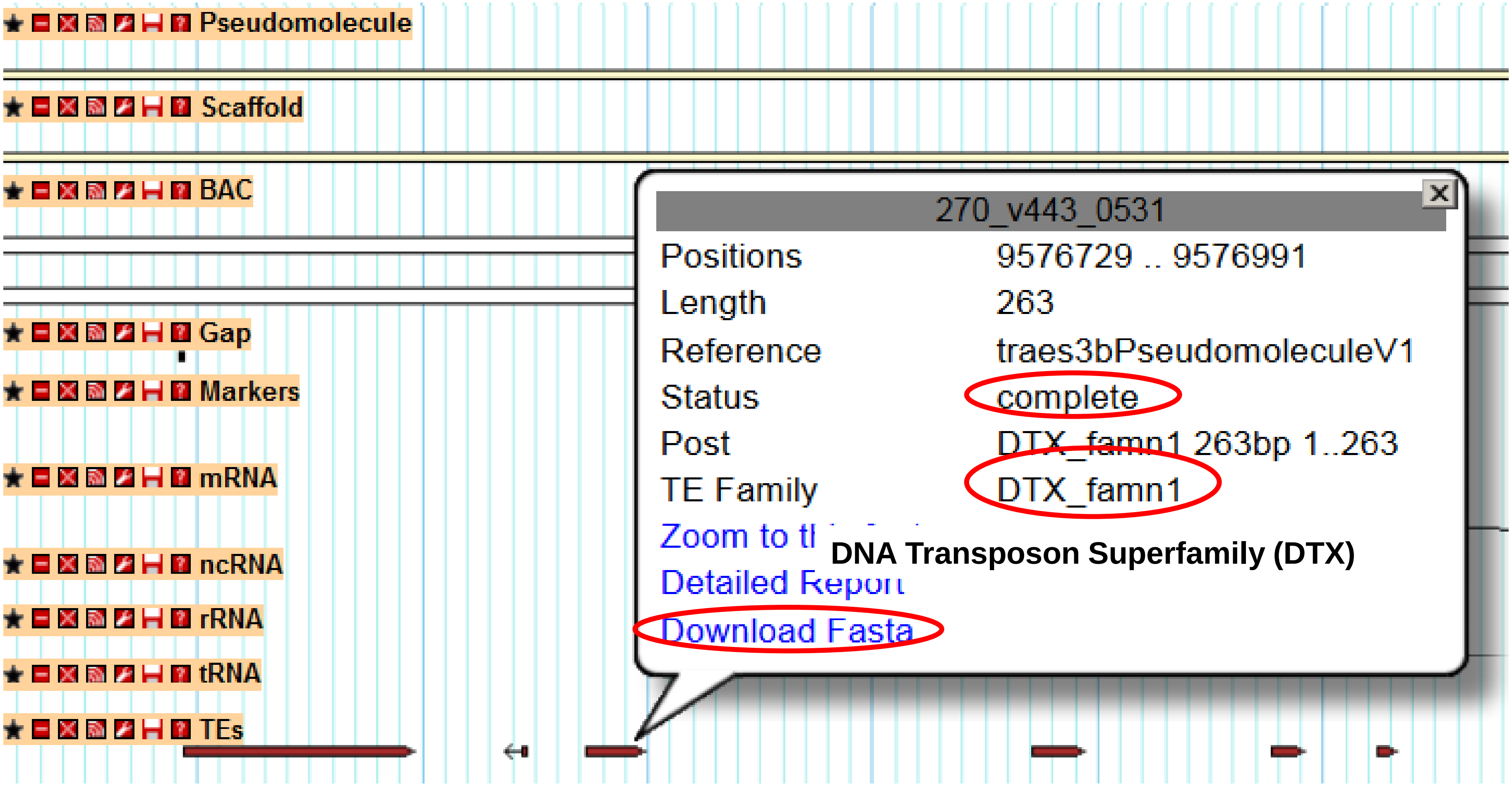
Zoom to this feature

Detailed Report

Download Fasta

https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_annot_3B

3B annotation: TE



270_v443_0531

Positions	9576729 .. 9576991
Length	263
Reference	traes3bPseudomoleculeV1
Status	complete
Post	DTX_famn1 263bp 1..263
TE Family	DTX_famn1

Zoom to tl

Detailed Report

Download Fasta

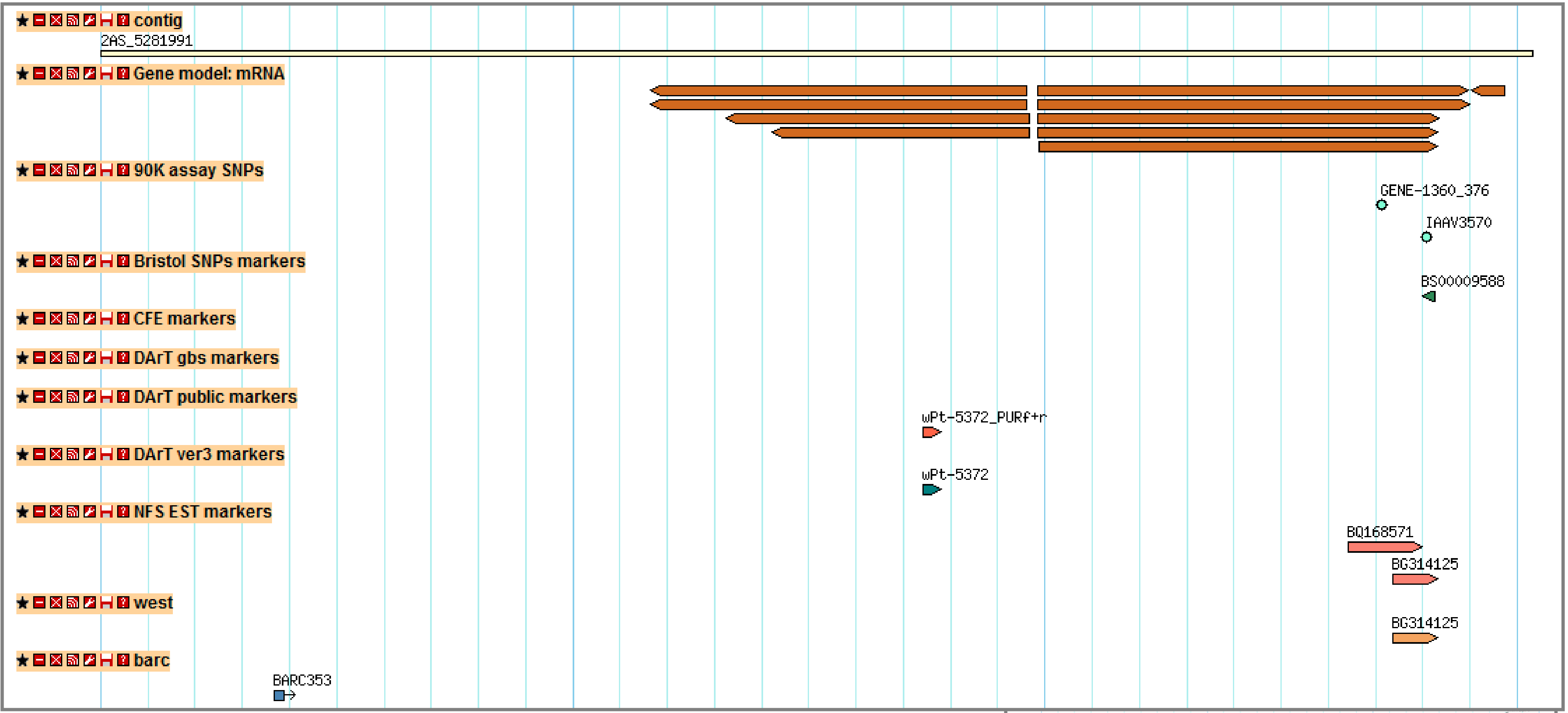
DNA Transposon Superfamily (DTX)

https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_annot_3B

Survey sequence browser

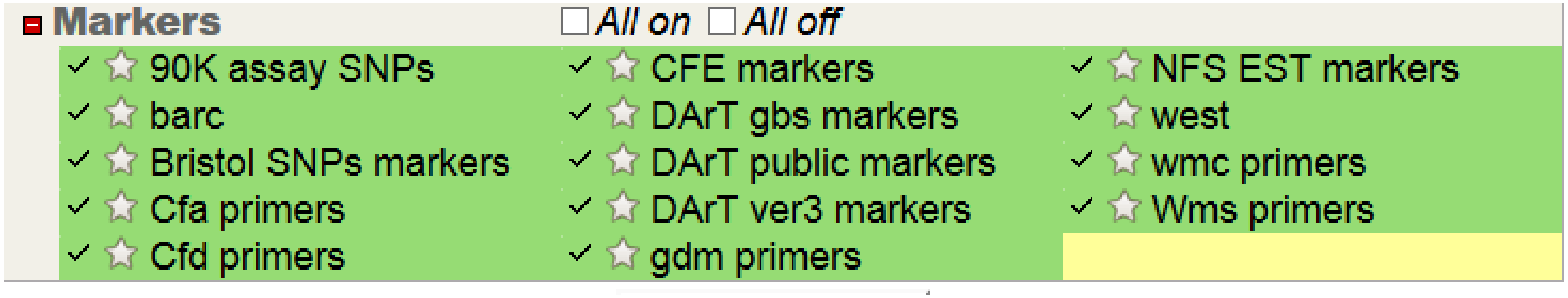
Michael Alaux

Survey sequence



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

Survey sequence



Markers ☐ *All on* ☐ *All off*

- ✓ ☆ 90K assay SNPs
- ✓ ☆ barc
- ✓ ☆ Bristol SNPs markers
- ✓ ☆ Cfa primers
- ✓ ☆ Cfd primers
- ✓ ☆ CFE markers
- ✓ ☆ DArT gbs markers
- ✓ ☆ DArT public markers
- ✓ ☆ DArT ver3 markers
- ✓ ☆ gdm primers
- ✓ ☆ NFS EST markers
- ✓ ☆ west
- ✓ ☆ wmc primers
- ✓ ☆ Wms primers

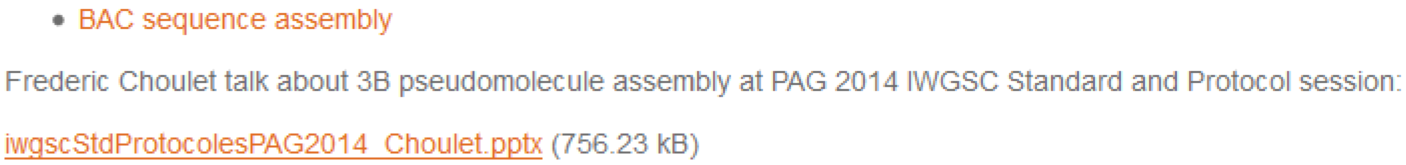
https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_survey_sequence_annotation

Michael Alaux

Support to assembly and data submission

Michael Alaux

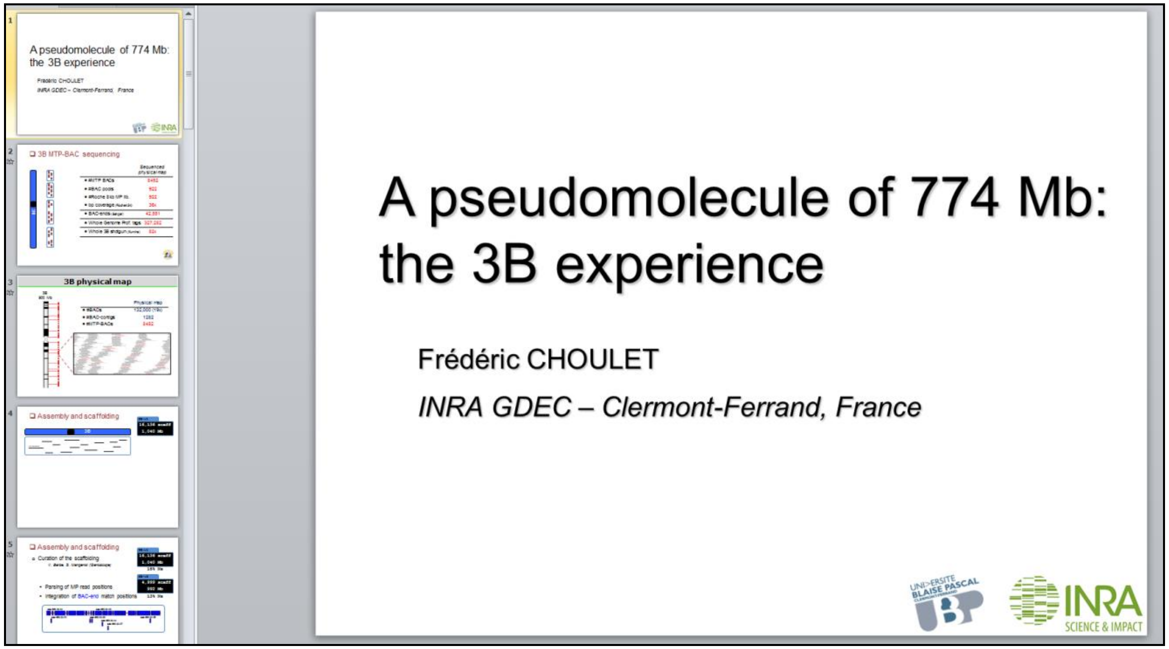
BAC sequence assembly



- BAC sequence assembly

Frederic Choulet talk about 3B pseudomolecule assembly at PAG 2014 IWGSC Standard and Protocol session:

iwgscStdProtocolesPAG2014_Choulet.pptx (756.23 kB)



A pseudomolecule of 774 Mb: the 3B experience

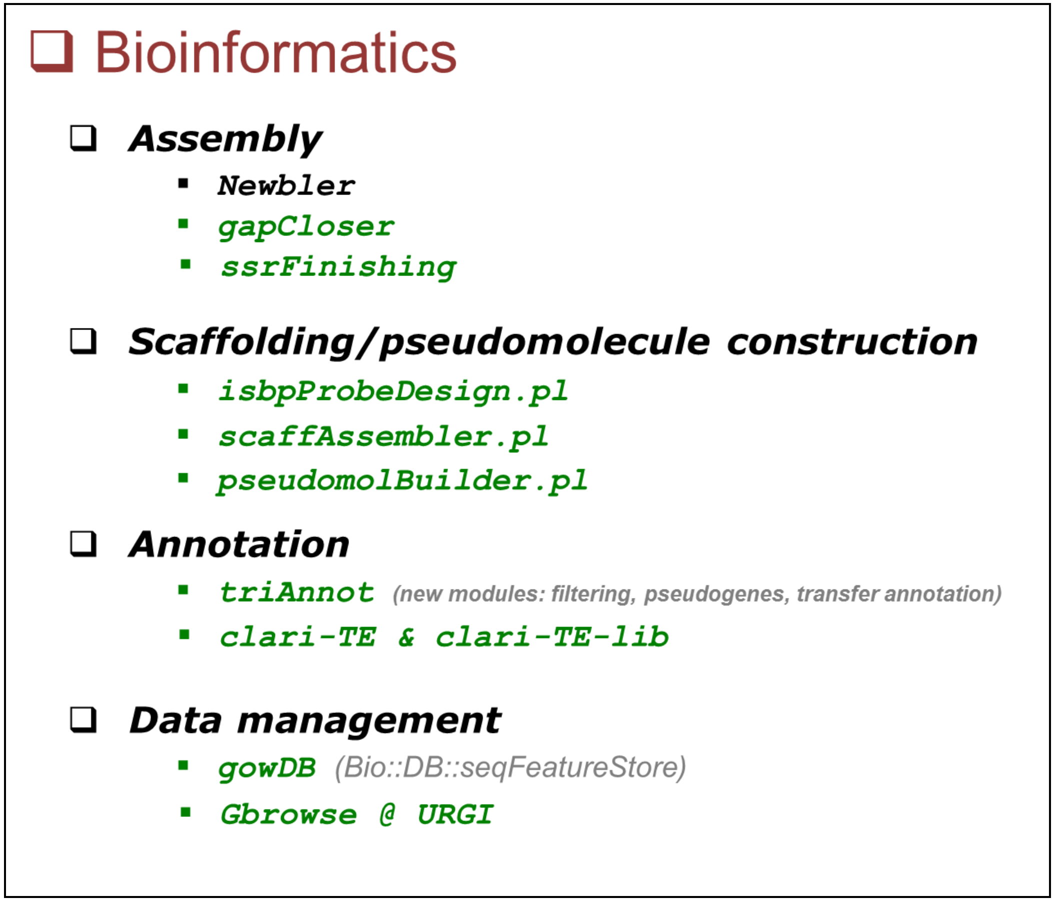
Frédéric CHOULET

INRA GDEC – Clermont-Ferrand, France

- ☐ 3B MTP BAC sequencing
- ☐ 3B physical map
- ☐ Assembly and scaffolding
- ☐ Ordering scaffolds
- ☐ Perform the Annotation
- ☐ Bioinformatic tools

BAC sequence assembly

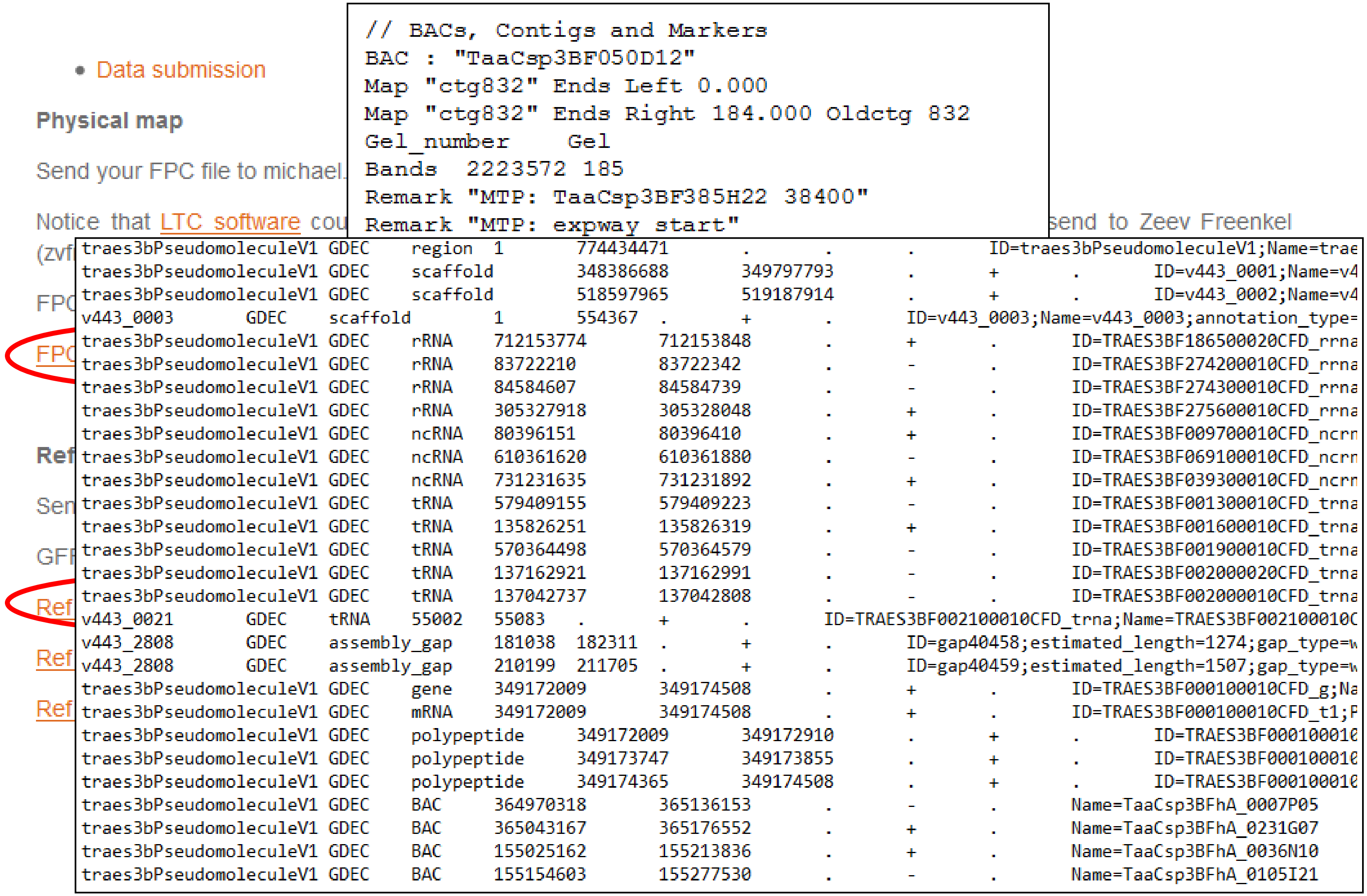
What are the needs ?
i.e. distribute the bioinformatics softwares



❑ Bioinformatics

- ❑ ***Assembly***
  - `Newbler`
  - *`gapCloser`*
  - *`ssrFinishing`*
- ❑ ***Scaffolding/pseudomolecule construction***
  - *`isbpProbeDesign.pl`*
  - *`scaffAssembler.pl`*
  - *`pseudomolBuilder.pl`*
- ❑ ***Annotation***
  - *`triAnnot`* *(new modules: filtering, pseudogenes, transfer annotation)*
  - *`clari-TE & clari-TE-lib`*
- ❑ ***Data management***
  - *`gowDB`* *(Bio::DB::seqFeatureStore)*
  - *`Gbrowse @ URGI`*

Data submission



- Data submission

Physical map

Send your FPC file to michael.

Notice that LTC software cou … send to Zeev Freenkel

(zvf

FPC

FPC

Ref

Sen

GFF

Ref

Ref

Ref

```
// BACs, Contigs and Markers
BAC : "TaaCsp3BF050D12"
Map "ctg832" Ends Left 0.000
Map "ctg832" Ends Right 184.000 Oldctg 832
Gel_number    Gel
Bands  2223572 185
Remark "MTP: TaaCsp3BF385H22 38400"
Remark "MTP: expway start"
```

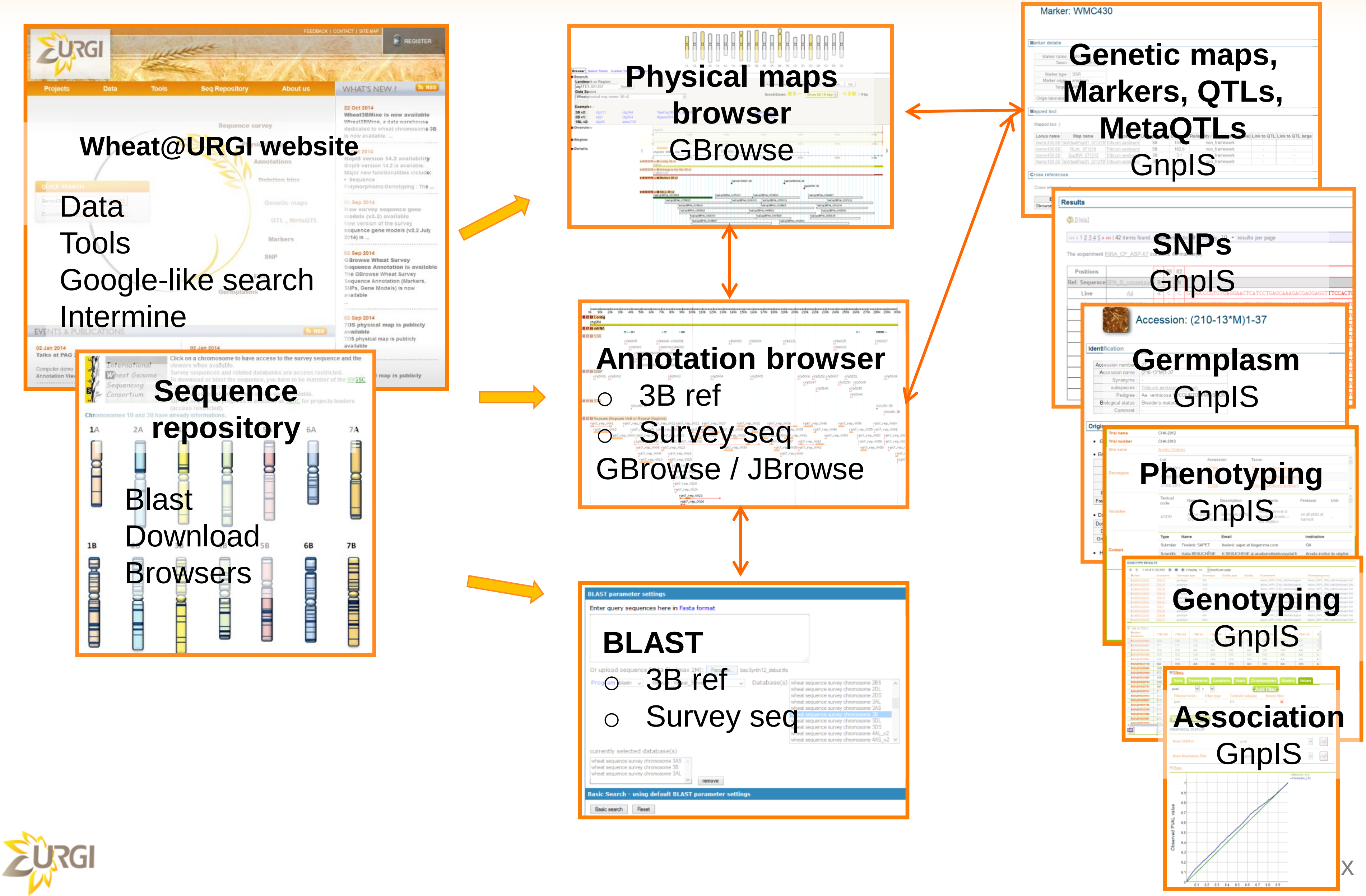
```
traes3bPseudomoleculeV1 GDEC    region  1       774434471       .       .       .       ID=traes3bPseudomoleculeV1;Name=trae
traes3bPseudomoleculeV1 GDEC    scaffold        348386688       349797793       .       +       .       ID=v443_0001;Name=v4
traes3bPseudomoleculeV1 GDEC    scaffold        518597965       519187914       .       +       .       ID=v443_0002;Name=v4
v443_0003       GDEC    scaffold        1       554367  .       +       .       ID=v443_0003;Name=v443_0003;annotation_type=
traes3bPseudomoleculeV1 GDEC    rRNA    712153774       712153848       .       +       .       ID=TRAES3BF186500020CFD_rrna
traes3bPseudomoleculeV1 GDEC    rRNA    83722210        83722342        .       -       .       ID=TRAES3BF274200010CFD_rrna
traes3bPseudomoleculeV1 GDEC    rRNA    84584607        84584739        .       -       .       ID=TRAES3BF274300010CFD_rrna
traes3bPseudomoleculeV1 GDEC    rRNA    305327918       305328048       .       +       .       ID=TRAES3BF275600010CFD_rrna
traes3bPseudomoleculeV1 GDEC    ncRNA   80396151        80396410        .       +       .       ID=TRAES3BF009700010CFD_ncrn
traes3bPseudomoleculeV1 GDEC    ncRNA   610361620       610361880       .       -       .       ID=TRAES3BF069100010CFD_ncrn
traes3bPseudomoleculeV1 GDEC    ncRNA   731231635       731231892       .       +       .       ID=TRAES3BF039300010CFD_ncrn
traes3bPseudomoleculeV1 GDEC    tRNA    579409155       579409223       .       -       .       ID=TRAES3BF001300010CFD_trna
traes3bPseudomoleculeV1 GDEC    tRNA    135826251       135826319       .       +       .       ID=TRAES3BF001600010CFD_trna
traes3bPseudomoleculeV1 GDEC    tRNA    570364498       570364579       .       -       .       ID=TRAES3BF001900010CFD_trna
traes3bPseudomoleculeV1 GDEC    tRNA    137162921       137162991       .       -       .       ID=TRAES3BF002000020CFD_trna
traes3bPseudomoleculeV1 GDEC    tRNA    137042737       137042808       .       -       .       ID=TRAES3BF002000010CFD_trna
v443_0021       GDEC    tRNA    55002   55083   .       +       .       ID=TRAES3BF002100010CFD_trna;Name=TRAES3BF002100010C
v443_2808       GDEC    assembly_gap    181038  182311  .       +       .       ID=gap40458;estimated_length=1274;gap_type=w
v443_2808       GDEC    assembly_gap    210199  211705  .       +       .       ID=gap40459;estimated_length=1507;gap_type=w
traes3bPseudomoleculeV1 GDEC    gene    349172009       349174508       .       +       .       ID=TRAES3BF000100010CFD_g;Na
traes3bPseudomoleculeV1 GDEC    mRNA    349172009       349174508       .       +       .       ID=TRAES3BF000100010CFD_t1;P
traes3bPseudomoleculeV1 GDEC    polypeptide     349172009       349172910       .       +       .       ID=TRAES3BF00010001
traes3bPseudomoleculeV1 GDEC    polypeptide     349173747       349173855       .       +       .       ID=TRAES3BF00010001
traes3bPseudomoleculeV1 GDEC    polypeptide     349174365       349174508       .       +       .       ID=TRAES3BF00010001
traes3bPseudomoleculeV1 GDEC    BAC     364970318       365136153       .       -       .       Name=TaaCsp3BFhA_0007P05
traes3bPseudomoleculeV1 GDEC    BAC     365043167       365176552       .       +       .       Name=TaaCsp3BFhA_0231G07
traes3bPseudomoleculeV1 GDEC    BAC     155025162       155213836       .       +       .       Name=TaaCsp3BFhA_0036N10
traes3bPseudomoleculeV1 GDEC    BAC     155154603       155277530       .       -       .       Name=TaaCsp3BFhA_0105I21
```

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

Wheat dataflow overview

Michael Alaux

Wheat dataflow overview



Wheat@URGI website

- Data
- Tools
- Google-like search
- Intermine

Sequence repository

- Blast
- Download
- Browsers

Physical maps browser

GBrowse

Annotation browser

- ○ 3B ref
- ○ Survey seq

GBrowse / JBrowse

BLAST

- ○ 3B ref
- ○ Survey seq

Genetic maps, Markers, QTLs, MetaQTLs

GnpIS

SNPs

GnpIS

Germplasm

GnpIS

Phenotyping

GnpIS

Genotyping

GnpIS

Association

GnpIS

Acknowledgments

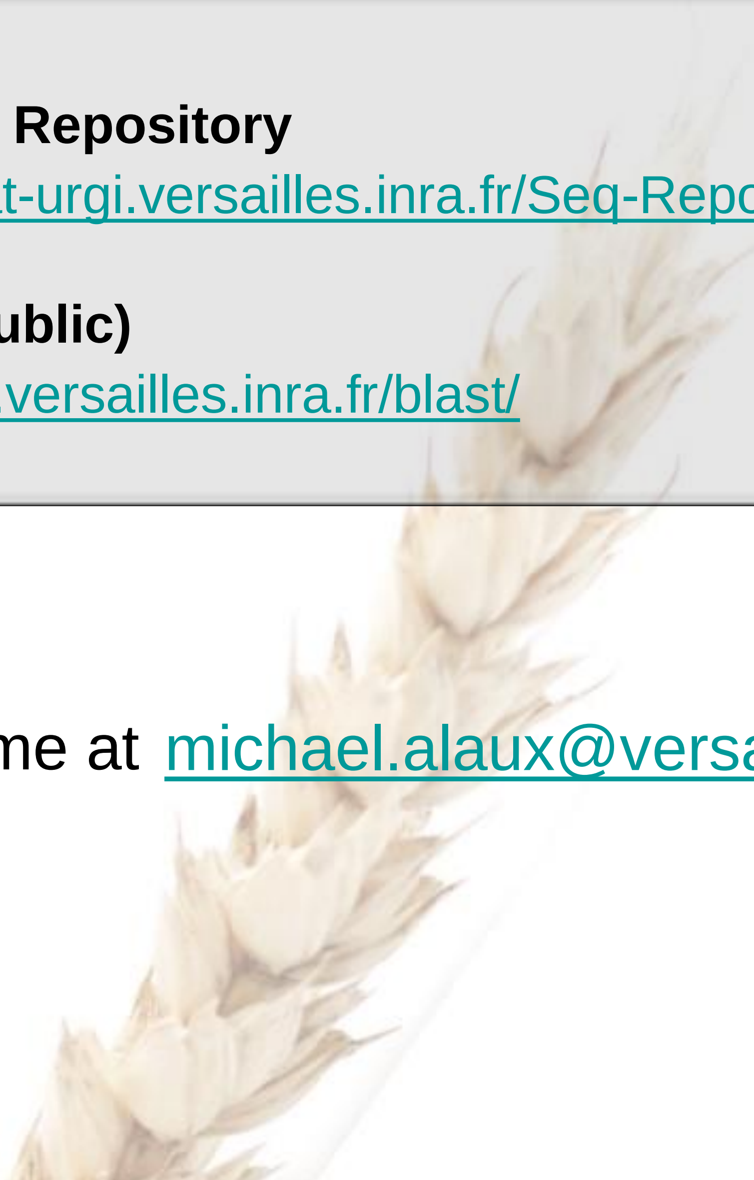
Alaux M.
Letellier T.
Couderc L.
Alfama F.
Jamilloux V.
Flores R.
Guerche C.
Loaec M.
Steinbach D.
Quesneville H.

Choulet F.
Paux E.

Rogers J.
Eversole K.

All data providers

Questions



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

BLAST (public)
https://urgi.versailles.inra.fr/blast/

Contact me at michael.alaux@versailles.inra.fr