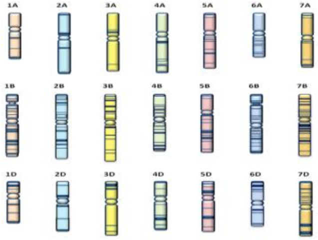
International
Wheat Genome
Sequencing
Consortium

IWGSC Sequence Repository:

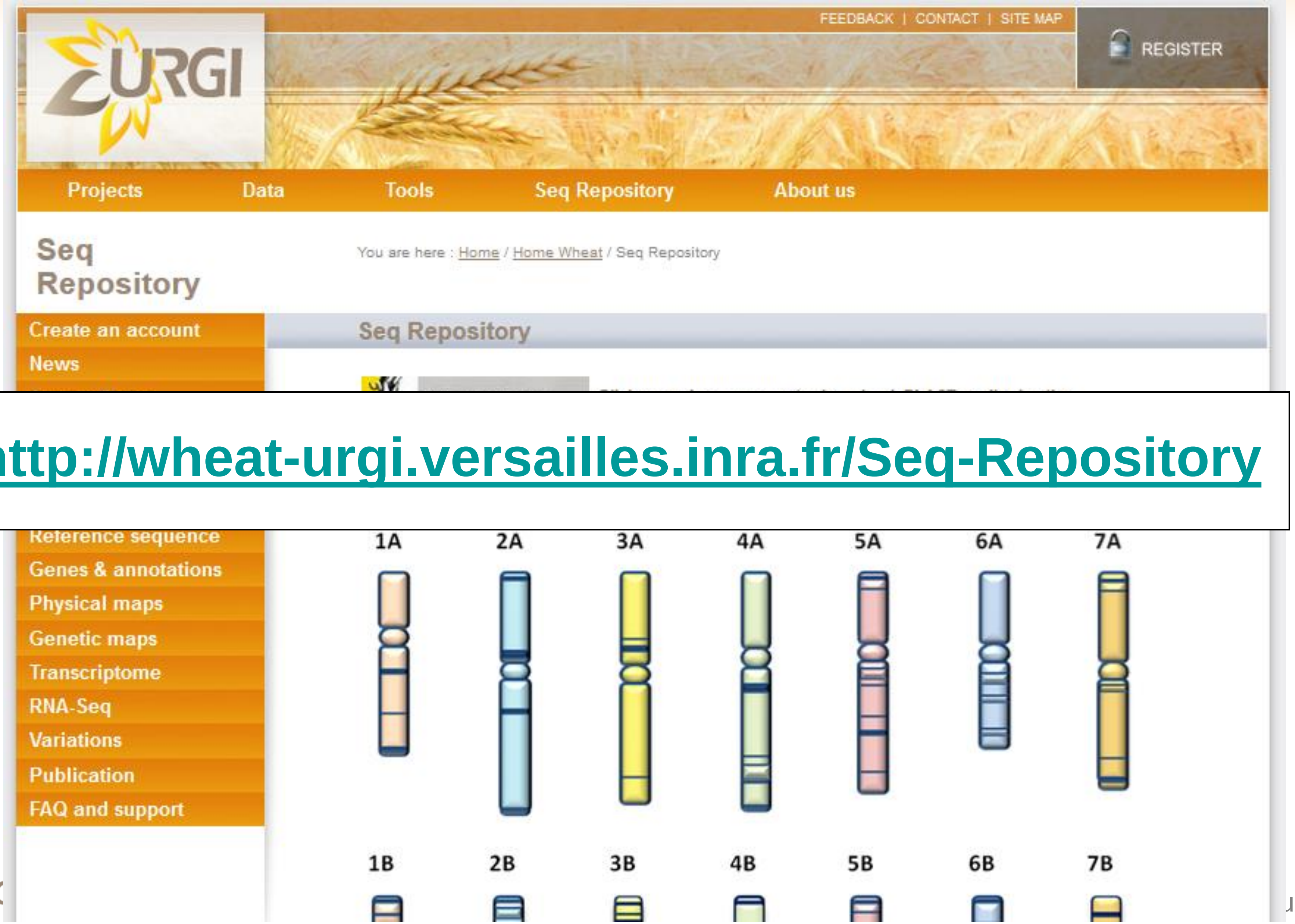
Moving towards tools to facilitate data integration for the reference sequence of wheat

Michael Alaux



IWGSC S&P, San Diego, 12 January 2016

IWGSC Sequence Repository



You are here : Home / Home Wheat / Seq Repository

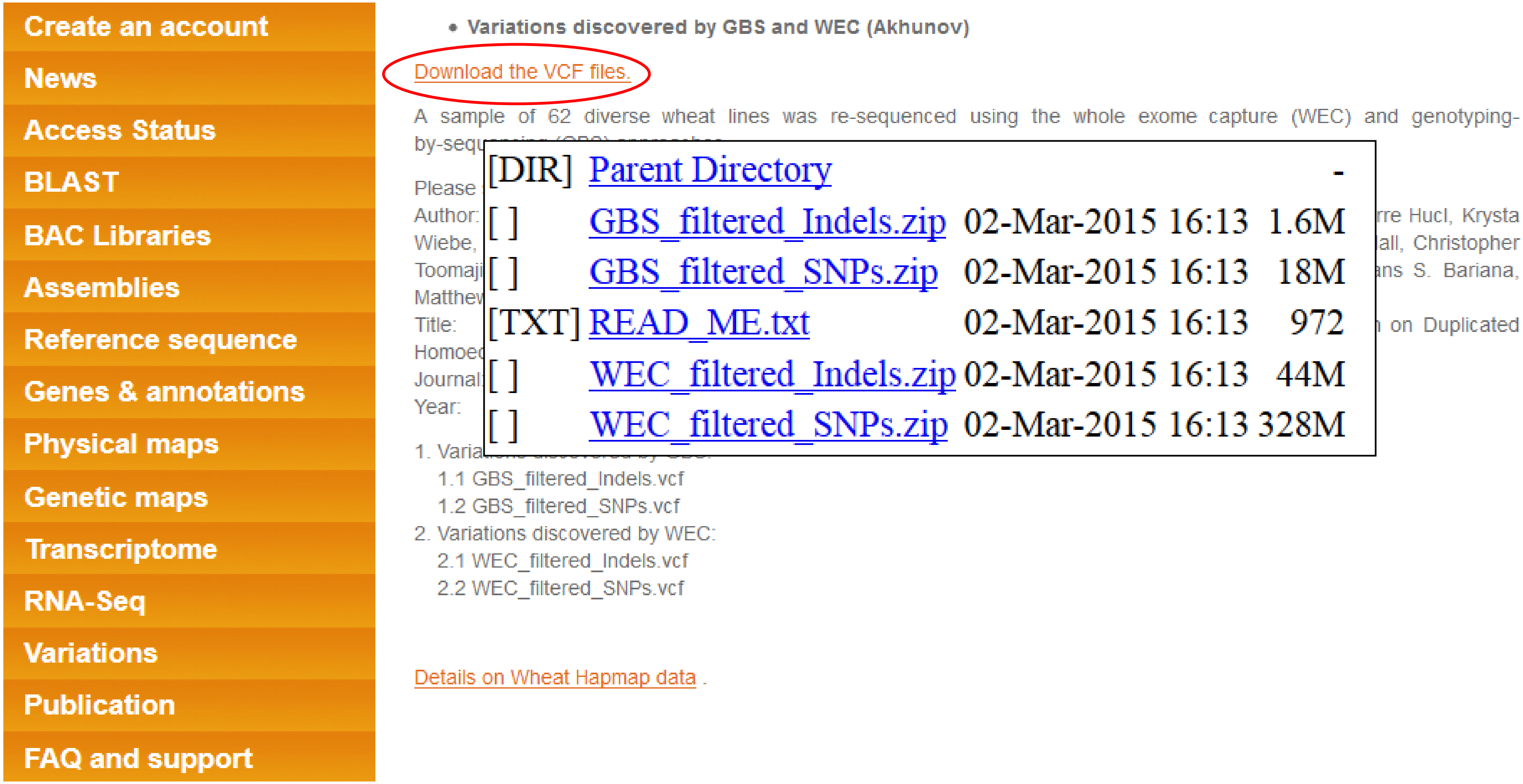
Seq Repository

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

New data available in 2015

Michael Alaux

New Variations data



- **Variations discovered by GBS and WEC (Akhunov)**

Download the VCF files.

A sample of 62 diverse wheat lines was re-sequenced using the whole exome capture (WEC) and genotyping-by-sequ

Please

Author: … rre Hucl, Krysta
Wiebe, … all, Christopher
Toomaji … ans S. Bariana,
Matthew

Title: … on Duplicated
Homoeo

Journal:

Year:

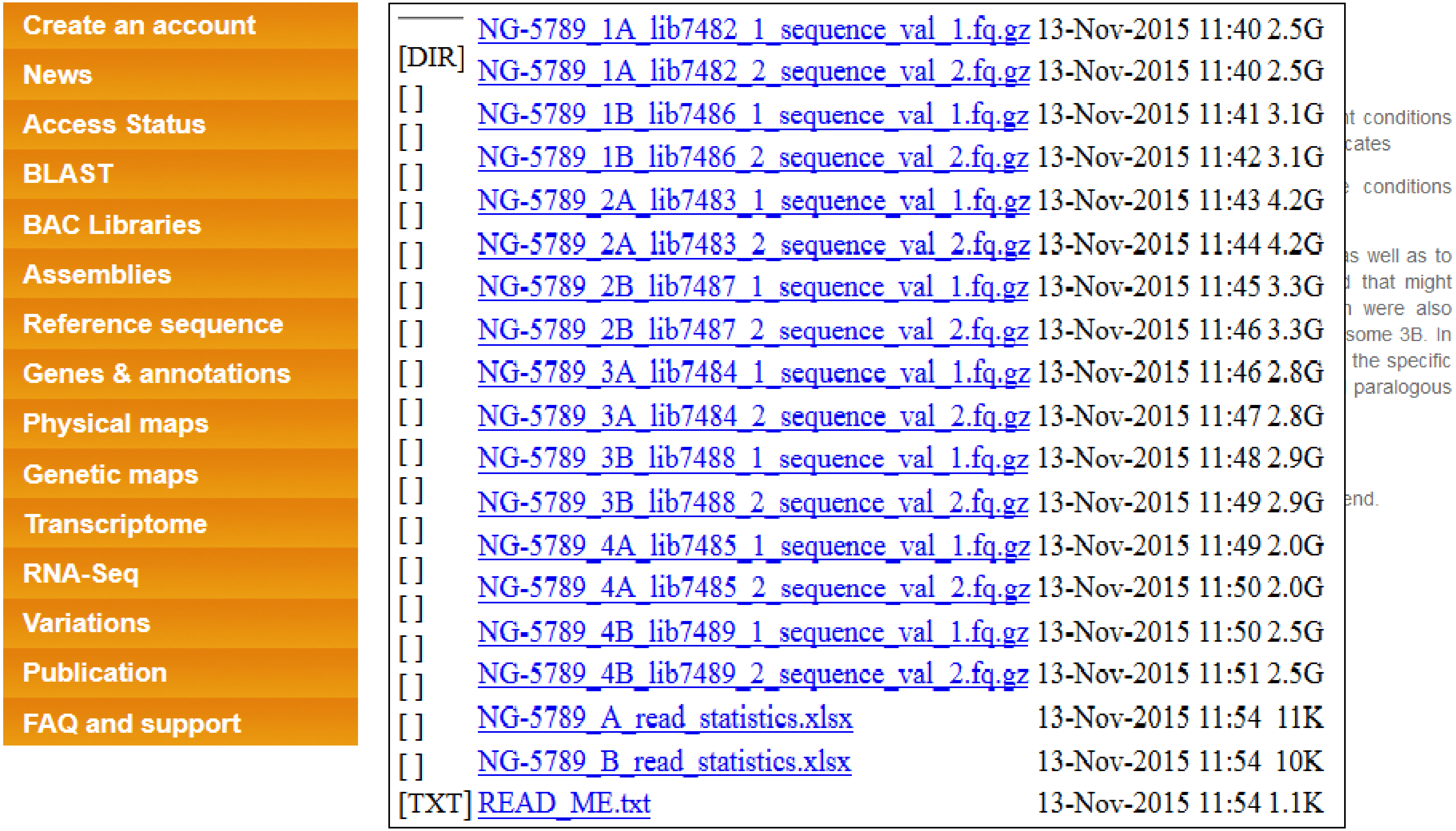
[DIR]	Parent Directory		-
[ ]	GBS_filtered_Indels.zip	02-Mar-2015 16:13	1.6M
[ ]	GBS_filtered_SNPs.zip	02-Mar-2015 16:13	18M
[TXT]	READ_ME.txt	02-Mar-2015 16:13	972
[ ]	WEC_filtered_Indels.zip	02-Mar-2015 16:13	44M
[ ]	WEC_filtered_SNPs.zip	02-Mar-2015 16:13	328M

1. Varia
   1.1 GBS_filtered_Indels.vcf
   1.2 GBS_filtered_SNPs.vcf
2. Variations discovered by WEC:
   2.1 WEC_filtered_Indels.vcf
   2.2 WEC_filtered_SNPs.vcf

Details on Wheat Hapmap data .

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

New RNA-Seq data



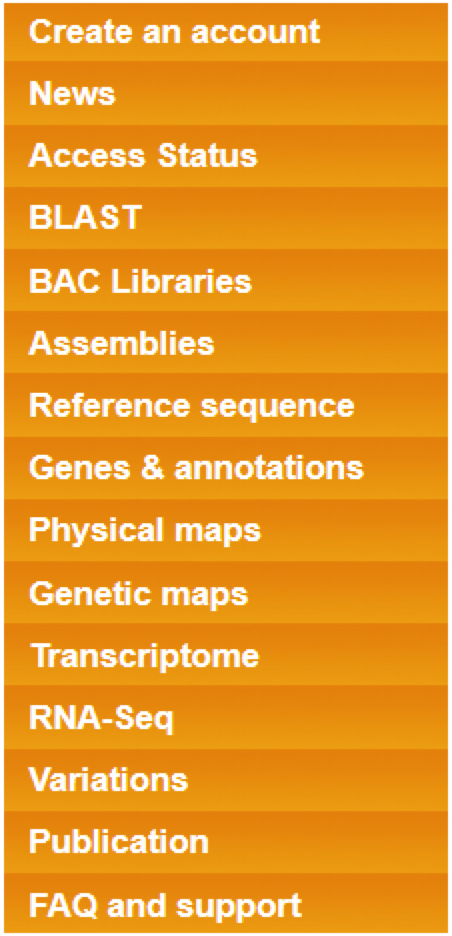
- Create an account
- News
- Access Status
- BLAST
- BAC Libraries
- Assemblies
- Reference sequence
- Genes & annotations
- Physical maps
- Genetic maps
- Transcriptome
- ✓ RNA-Seq
- Variations
- Publication
- FAQ and support

```
——     NG-5789_1A_lib7482_1_sequence_val_1.fq.gz 13-Nov-2015 11:40 2.5G
[DIR]  NG-5789_1A_lib7482_2_sequence_val_2.fq.gz 13-Nov-2015 11:40 2.5G
[ ]    NG-5789_1B_lib7486_1_sequence_val_1.fq.gz 13-Nov-2015 11:41 3.1G
[ ]    NG-5789_1B_lib7486_2_sequence_val_2.fq.gz 13-Nov-2015 11:42 3.1G
[ ]    NG-5789_2A_lib7483_1_sequence_val_1.fq.gz 13-Nov-2015 11:43 4.2G
[ ]    NG-5789_2A_lib7483_2_sequence_val_2.fq.gz 13-Nov-2015 11:44 4.2G
[ ]    NG-5789_2B_lib7487_1_sequence_val_1.fq.gz 13-Nov-2015 11:45 3.3G
[ ]    NG-5789_2B_lib7487_2_sequence_val_2.fq.gz 13-Nov-2015 11:46 3.3G
[ ]    NG-5789_3A_lib7484_1_sequence_val_1.fq.gz 13-Nov-2015 11:46 2.8G
[ ]    NG-5789_3A_lib7484_2_sequence_val_2.fq.gz 13-Nov-2015 11:47 2.8G
[ ]    NG-5789_3B_lib7488_1_sequence_val_1.fq.gz 13-Nov-2015 11:48 2.9G
[ ]    NG-5789_3B_lib7488_2_sequence_val_2.fq.gz 13-Nov-2015 11:49 2.9G
[ ]    NG-5789_4A_lib7485_1_sequence_val_1.fq.gz 13-Nov-2015 11:49 2.0G
[ ]    NG-5789_4A_lib7485_2_sequence_val_2.fq.gz 13-Nov-2015 11:50 2.0G
[ ]    NG-5789_4B_lib7489_1_sequence_val_1.fq.gz 13-Nov-2015 11:50 2.5G
[ ]    NG-5789_4B_lib7489_2_sequence_val_2.fq.gz 13-Nov-2015 11:51 2.5G
[ ]    NG-5789_A_read_statistics.xlsx            13-Nov-2015 11:54  11K
[ ]    NG-5789_B_read_statistics.xlsx            13-Nov-2015 11:54  10K
[ ]
[ ]
[TXT]  READ_ME.txt                               13-Nov-2015 11:54 1.1K
```

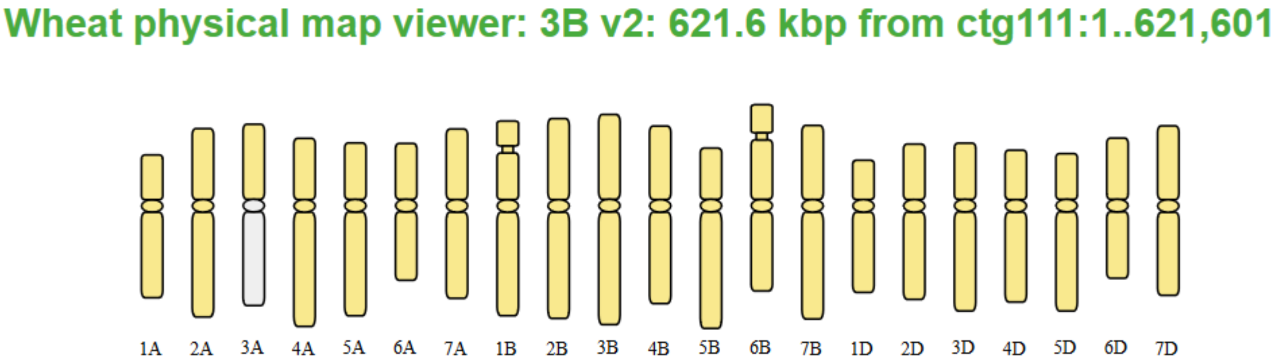
t conditions
cates
conditions
s well as to
that might
were also
some 3B. In
the specific
paralogous
end.

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

New Physical maps



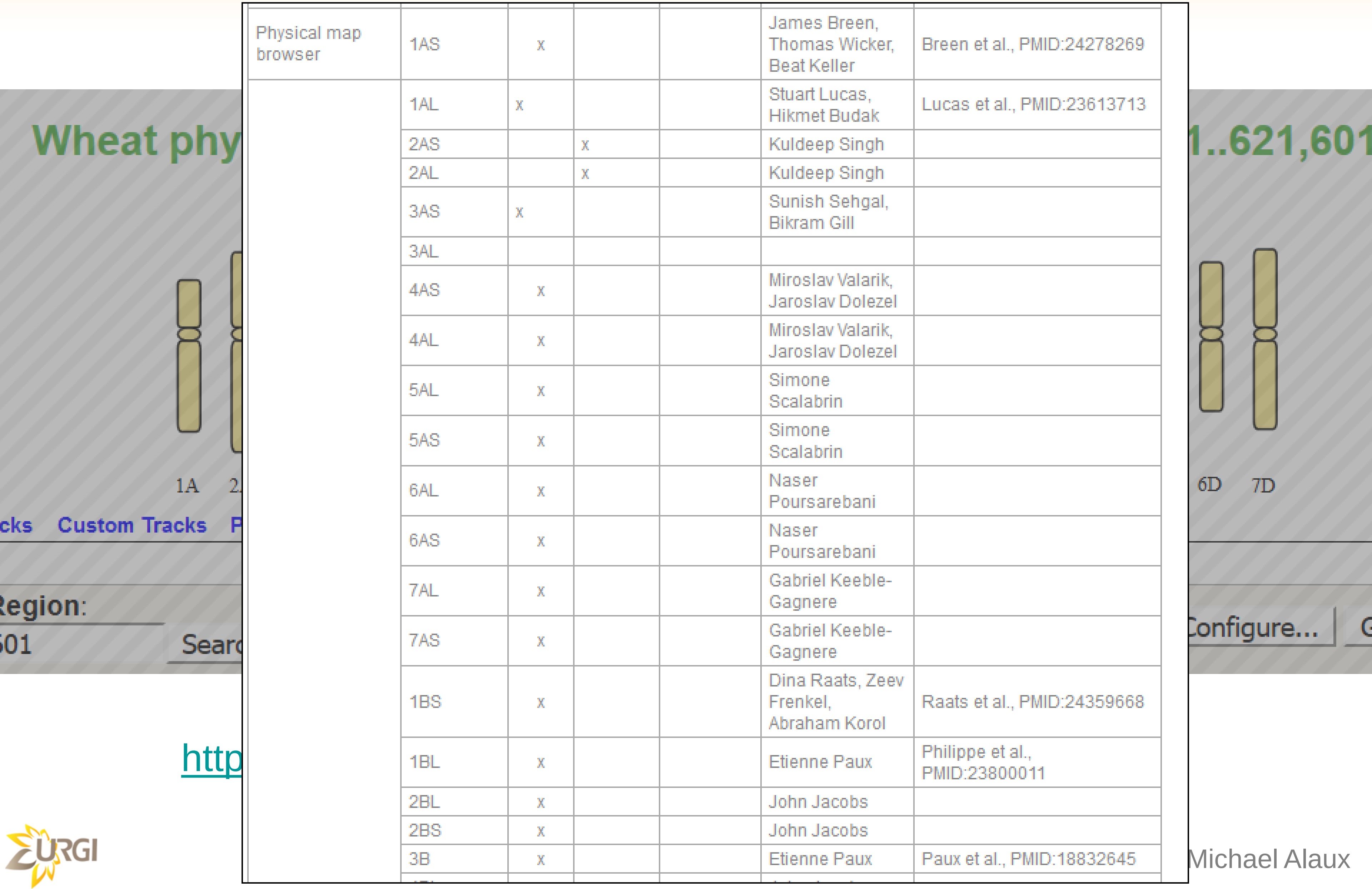
- Create an account
- News
- Access Status
- BLAST
- BAC Libraries
- Assemblies
- Reference sequence
- Genes & annotations
- ✓ Physical maps
- Genetic maps
- Transcriptome
- RNA-Seq
- Variations
- Publication
- FAQ and support



Wheat physical map viewer: 3B v2: 621.6 kbp from ctg111:1..621,601

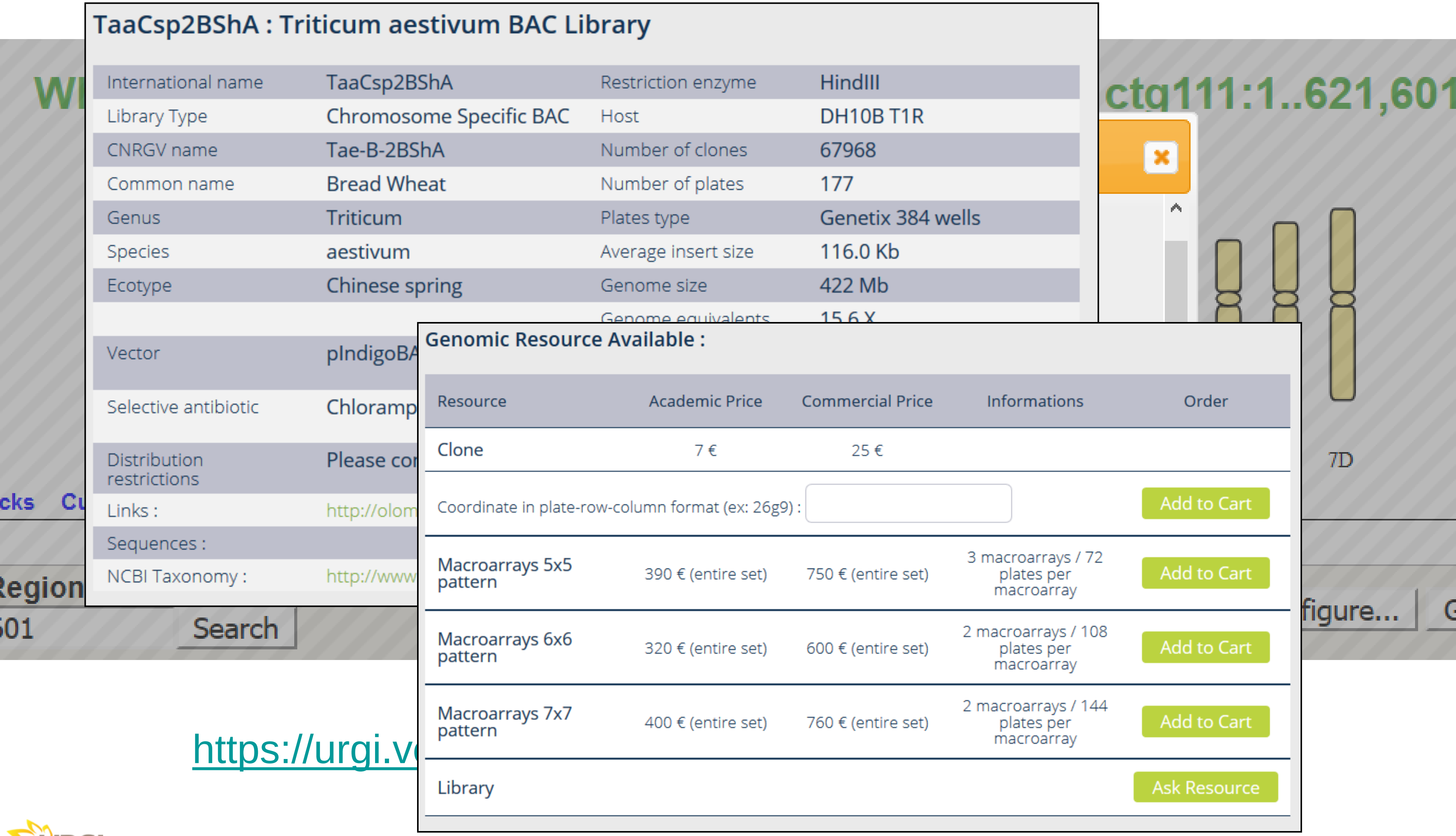
All physical maps (except 3AL) are available !

Exemple of 2BS (Bayer)



Physical map browser	1AS	x			James Breen, Thomas Wicker, Beat Keller	Breen et al., PMID:24278269
	1AL	x			Stuart Lucas, Hikmet Budak	Lucas et al., PMID:23613713
	2AS		x		Kuldeep Singh	
	2AL		x		Kuldeep Singh	
	3AS	x			Sunish Sehgal, Bikram Gill	
	3AL					
	4AS	x			Miroslav Valarik, Jaroslav Dolezel	
	4AL	x			Miroslav Valarik, Jaroslav Dolezel	
	5AL	x			Simone Scalabrin	
	5AS	x			Simone Scalabrin	
	6AL	x			Naser Poursarebani	
	6AS	x			Naser Poursarebani	
	7AL	x			Gabriel Keeble-Gagnere	
	7AS	x			Gabriel Keeble-Gagnere	
	1BS	x			Dina Raats, Zeev Frenkel, Abraham Korol	Raats et al., PMID:24359668
	1BL	x			Etienne Paux	Philippe et al., PMID:23800011
	2BL	x			John Jacobs	
	2BS	x			John Jacobs	
	3B	x			Etienne Paux	Paux et al., PMID:18832645

Exemple of 2BS (Bayer)



TaaCsp2BShA : Triticum aestivum BAC Library

International name	TaaCsp2BShA	Restriction enzyme	HindIII
Library Type	Chromosome Specific BAC	Host	DH10B T1R
CNRGV name	Tae-B-2BShA	Number of clones	67968
Common name	Bread Wheat	Number of plates	177
Genus	Triticum	Plates type	Genetix 384 wells
Species	aestivum	Average insert size	116.0 Kb
Ecotype	Chinese spring	Genome size	422 Mb
		Genome equivalents	15.6 X
Vector	pIndigoBA		
Selective antibiotic	Chloramp		
Distribution restrictions	Please co		
Links :	http://olom		
Sequences :			
NCBI Taxonomy :	http://www		

Genomic Resource Available :

Resource	Academic Price	Commercial Price	Informations	Order
Clone	7 €	25 €		
Coordinate in plate-row-column format (ex: 26g9) :				Add to Cart
Macroarrays 5x5 pattern	390 € (entire set)	750 € (entire set)	3 macroarrays / 72 plates per macroarray	Add to Cart
Macroarrays 6x6 pattern	320 € (entire set)	600 € (entire set)	2 macroarrays / 108 plates per macroarray	Add to Cart
Macroarrays 7x7 pattern	400 € (entire set)	760 € (entire set)	2 macroarrays / 144 plates per macroarray	Add to Cart
Library				Ask Resource

https://urgi.ve

URGI

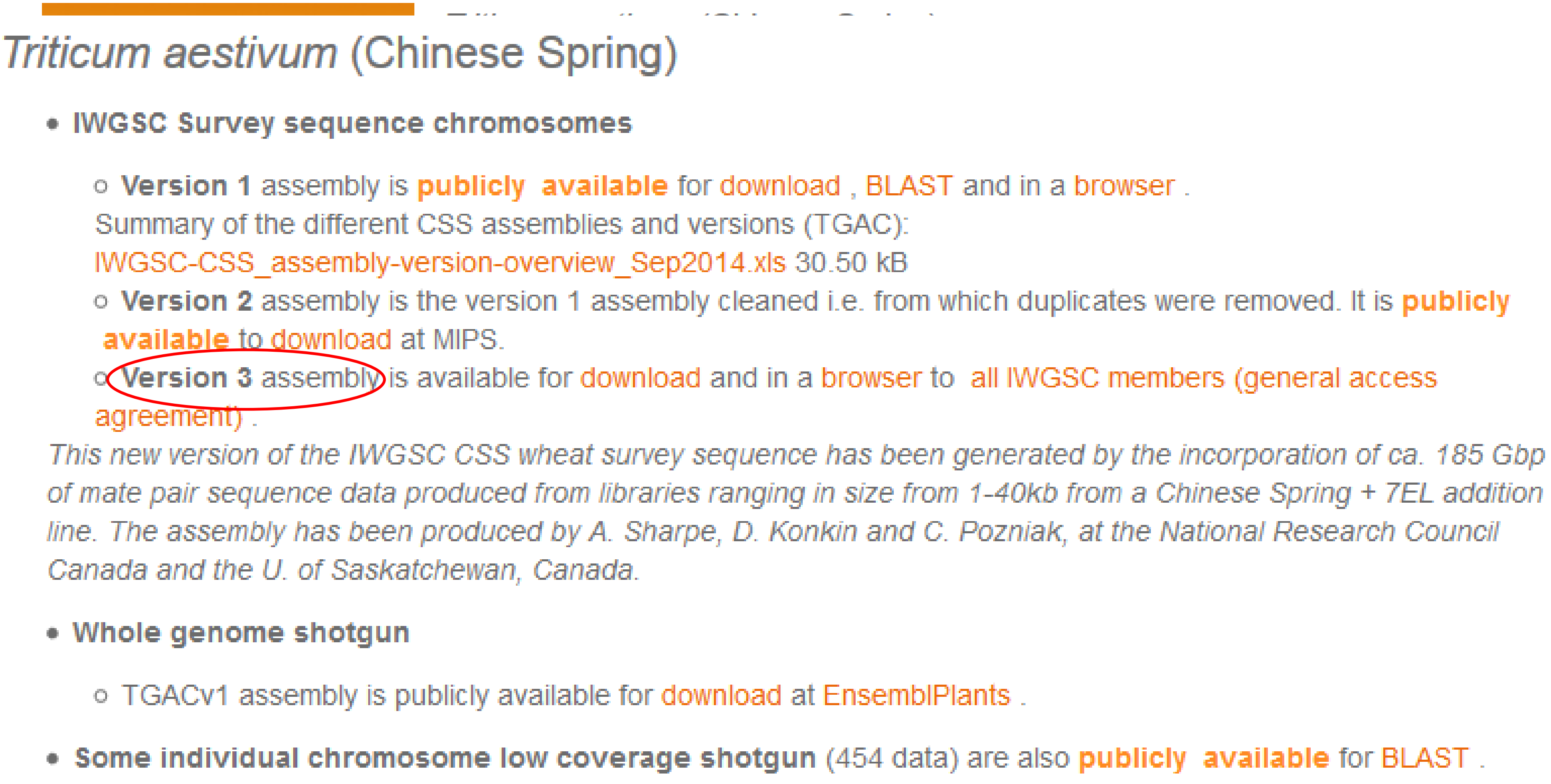
Exemple of 2BS (Bayer)

2BS-077.C12

```
AAGCTTAATGGGAAAGGAAGGGGTAAAGTGATGAGGTTGCAGCAGCGACT
AAGCTTTCGGGACGAAGCGTCGCCGTTTCGAGGCGGAACCTGGACGGAAC
AAGCTTTGGTGGCATTTGGTCTCTCCGTCGCTGGTCTTCACTTTTTGGTC
AAGCTTTCGATATTCATCATATTTCTTTCCTCGTTTCTAATGCTTTCTCA
AAGCTTCAATAGCCCTCAACTAAGCATACGTCTAGAATCTATGGGTATCC
AAGCTTGACGCGCCCTTCTGTAAGTGCTGCTGGTGGTGAAAGTGATTTGT
AAGCTTGTAGGCATTAGATCAATGATGGATAATTATGTCGGATGCGATTC
AAGCTTAGGCTTTTGGCCTAACTTGGTAATCACTCAGTAGCATGGATGGT
AAGCTTCTTGATCAGTTATTATTTCTATTAGTATTTCTTCAAAATGCCAA
AAGCTTCTCGGAATGAGTGAAGTTA
AAGCTTAGCCTTTTGGCCTAAGTTGGTCTATGGCCATGGGTAGCCTGGCT
AAGCTTGGAGTAGAAAGAGGCATGGGTGGCACCTCCAGCGCCCATCAAGT
AAGCTTGCGACATCTAAACCTGCTTCTGCTATTA
AAGCTTGATTTTGGGGCTTGTAAAGATGGGTTCCAGAGTCCAGGATCTTC
AAGCTTGAACAAGAAAATATCAATCTGAAAGCTGCTCTTGGGAAGAAAAA
AAGCTTGTGGGAGACTCTCTTCTTCAATTTGCATAAAATTGTATATTTCC
AAGCTTAGTATTTTATACAGCATTGAAAAGCATATGAATTTGCCTGAAAT
AAGCTTGTGGGAGACTCTCTTCTTCAATTTGCACGAAGTTA
AAGCTTTCATGCACGCGACAAAAAGTGTCAAATGATGATTGTTTCTCTCA
AAGCTTTCATCTAAGGAAATGGGGTCGTGGAGTTGCTGTAAGTAGGCTAG
AAGCTTTGCTAGTCTTGATACCCATGGTAATGGGATTGCTCAGTCCTTGT
AAGCTTCCCAAATTGGAACGAAGCTATCGTTTGAAAACCCTAAATGGGCA
AAGCTTGATGCACTCACTACACAACTCACAAAAATACATGGATGGCACCA
AAGCTTGTGTGGTTTCTTCTGATCCACTCGGAGGACTAAATCTCCTTCTT
```

https:/ ub/

New Assembly



Triticum aestivum (Chinese Spring)

- **IWGSC Survey sequence chromosomes**
  - **Version 1** assembly is **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
  - **Version 2** assembly is the version 1 assembly cleaned i.e. from which duplicates were removed. It is **publicly available** to download at MIPS.
  - **Version 3** assembly is available for download and in a browser to all IWGSC members (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.

- **Whole genome shotgun**
  - TGACv1 assembly is publicly available for download at EnsemblPlants .
- **Some individual chromosome low coverage shotgun** (454 data) are also **publicly available** for BLAST .

Survey sequence v3

Triticum aestivum (Chinese Spring)

- **IWGSC Survey sequence chromosomes**
  - **Version 1** assembly is **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
  - **Version 2** assembly is the version 1 assembly cleaned i.e. from which duplicates were removed. It is **publicly available** to download at MIPS.
  - **Version 3** assembly is available for download and in a browser to all IWGSC members (general access agreement) .

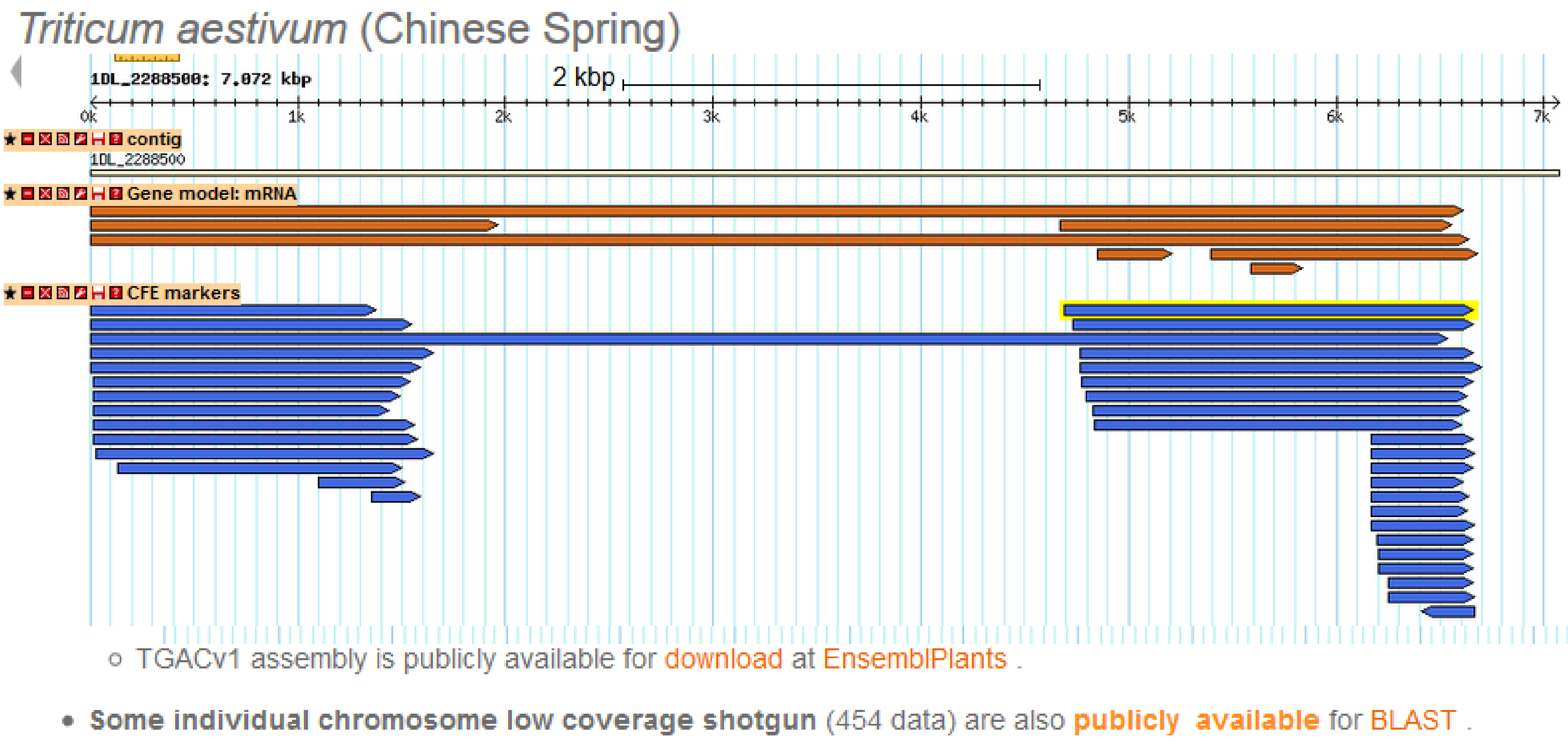
  This new version of the IWGSC CSS wheat survey sequence has been generated by the incorporation of ca. 185 Gbp of mate pai… b from a Chinese Spring + 7EL addition line. The as… k, at the National Research Council Canada and the U. of Saskatchewan, Canada.

[DIR]	Parent Directory		-
[ ]	wheat_survey_sequence_v3.tar.gz	30-Dec-2015 10:53	92M

- **Whole genome shotgun**
  - TGACv1 assembly is publicly available for download at EnsemblPlants .
- **Some individual chromosome low coverage shotgun** (454 data) are also **publicly available** for BLAST .

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

Survey sequence v3



Triticum aestivum (Chinese Spring)

- ○ TGACv1 assembly is publicly available for download at EnsemblPlants .

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

https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_iwgsc_survey_sequence_v3/

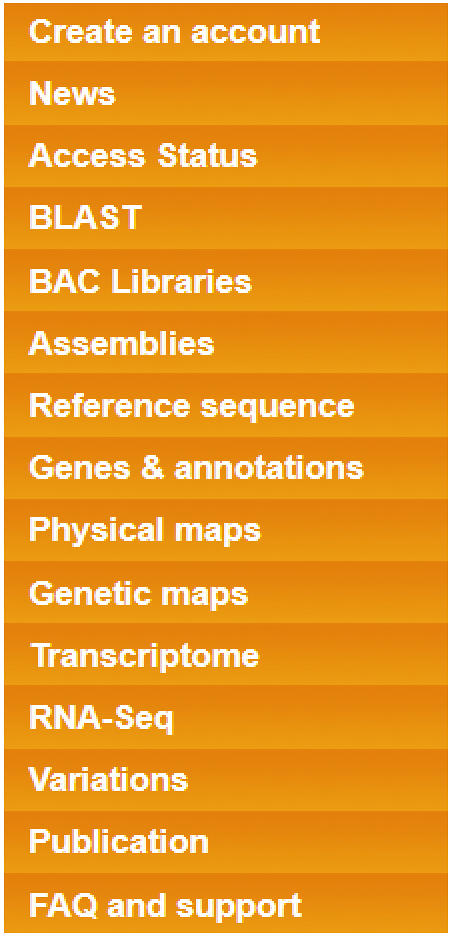
URGI

Michael Alaux

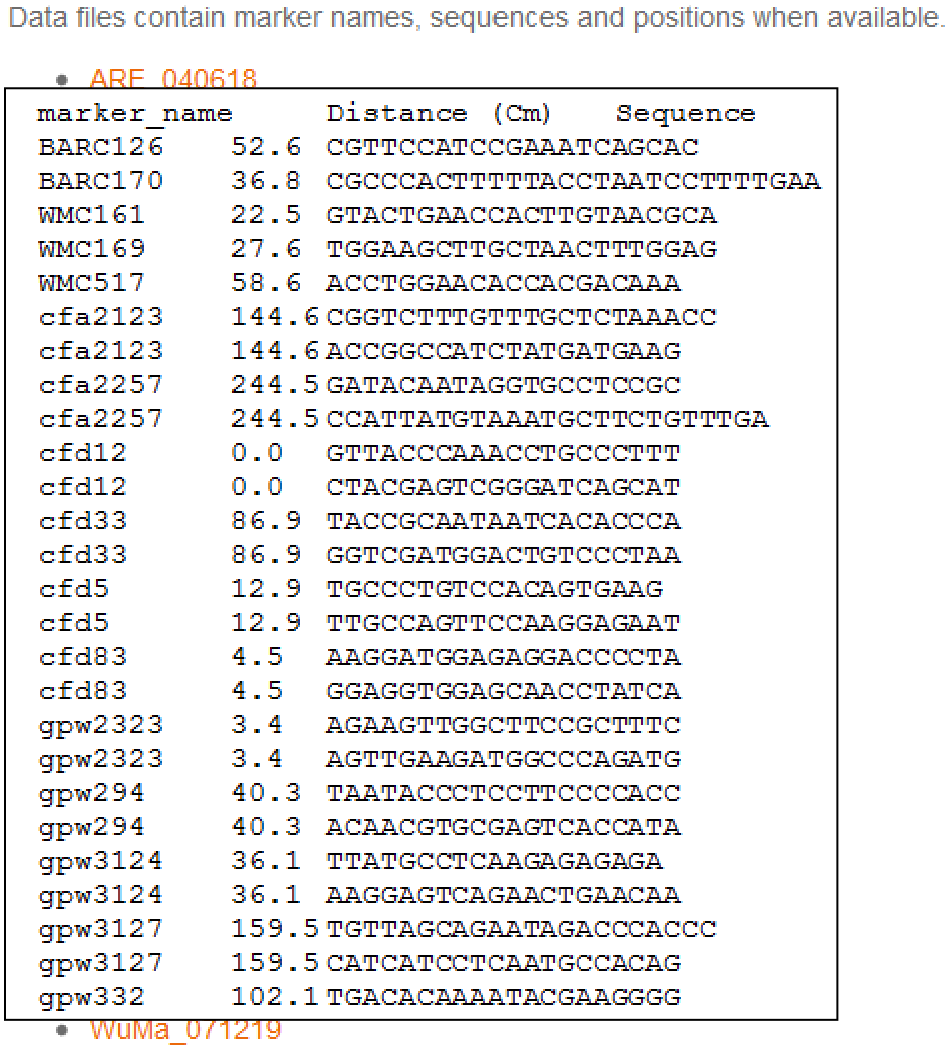
New tools functionalities in 2015

Michael Alaux

Genetic maps (feedback from PAG 2015)



✓



Data files contain marker names, sequences and positions when available.

- ARE_040618

```
marker_name      Distance (Cm)    Sequence
BARC126     52.6  CGTTCCATCCGAAATCAGCAC
BARC170     36.8  CGCCCACTTTTTACCTAATCCTTTTGAA
WMC161      22.5  GTACTGAACCACTTGTAACGCA
WMC169      27.6  TGGAAGCTTGCTAACTTTGGAG
WMC517      58.6  ACCTGGAACACCACGACAAA
cfa2123     144.6 CGGTCTTTGTTTGCTCTAAACC
cfa2123     144.6 ACCGGCCATCTATGATGAAG
cfa2257     244.5 GATACAATAGGTGCCTCCGC
cfa2257     244.5 CCATTATGTAAATGCTTCTGTTTGA
cfd12       0.0   GTTACCCAAACCTGCCCTTT
cfd12       0.0   CTACGAGTCGGGATCAGCAT
cfd33       86.9  TACCGCAATAATCACACCCA
cfd33       86.9  GGTCGATGGACTGTCCCTAA
cfd5        12.9  TGCCCTGTCCACAGTGAAG
cfd5        12.9  TTGCCAGTTCCAAGGAGAAT
cfd83       4.5   AAGGATGGAGAGGACCCCTA
cfd83       4.5   GGAGGTGGAGCAACCTATCA
gpw2323     3.4   AGAAGTTGGCTTCCGCTTTC
gpw2323     3.4   AGTTGAAGATGGCCCAGATG
gpw294      40.3  TAATACCCTCCTTCCCCACC
gpw294      40.3  ACAACGTGCGAGTCACCATA
gpw3124     36.1  TTATGCCTCAAGAGAGAGA
gpw3124     36.1  AAGGAGTCAGAACTGAACAA
gpw3127     159.5 TGTTAGCAGAATAGACCCACCC
gpw3127     159.5 CATCATCCTCAAGTGCCACAG
gpw332      102.1 TGACACAAAATACGAAGGGG
```

- WuMa_071219

URGI

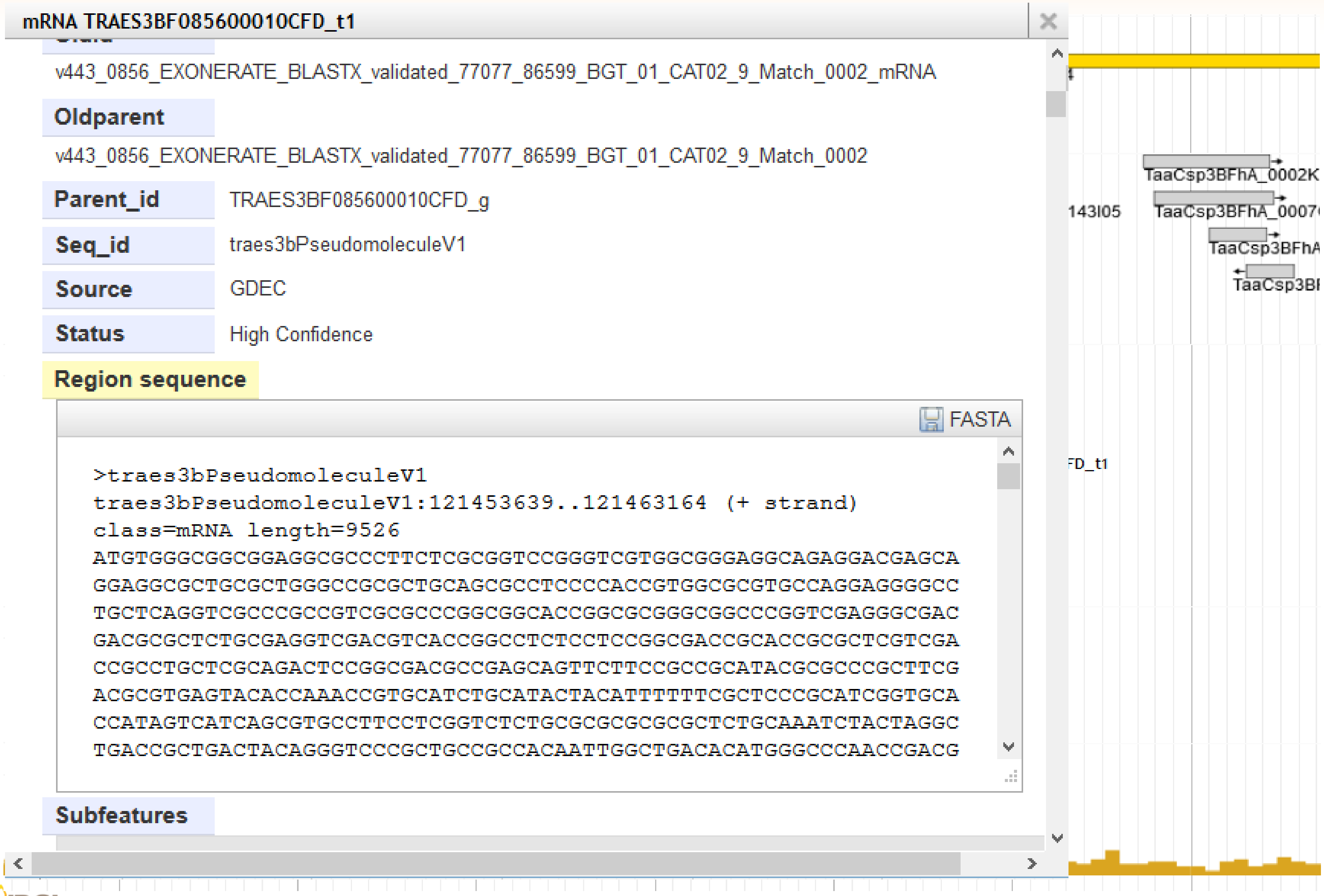
http://wheat-urgi.versailles.inra.fr/Seq-Repository

Michael Alaux

3B reference JBrowse

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

3B reference JBrowse



mRNA TRAES3BF085600010CFD_t1

v443_0856_EXONERATE_BLASTX_validated_77077_86599_BGT_01_CAT02_9_Match_0002_mRNA

Oldparent

v443_0856_EXONERATE_BLASTX_validated_77077_86599_BGT_01_CAT02_9_Match_0002

Parent_id	TRAES3BF085600010CFD_g
Seq_id	traes3bPseudomoleculeV1
Source	GDEC
Status	High Confidence

Region sequence

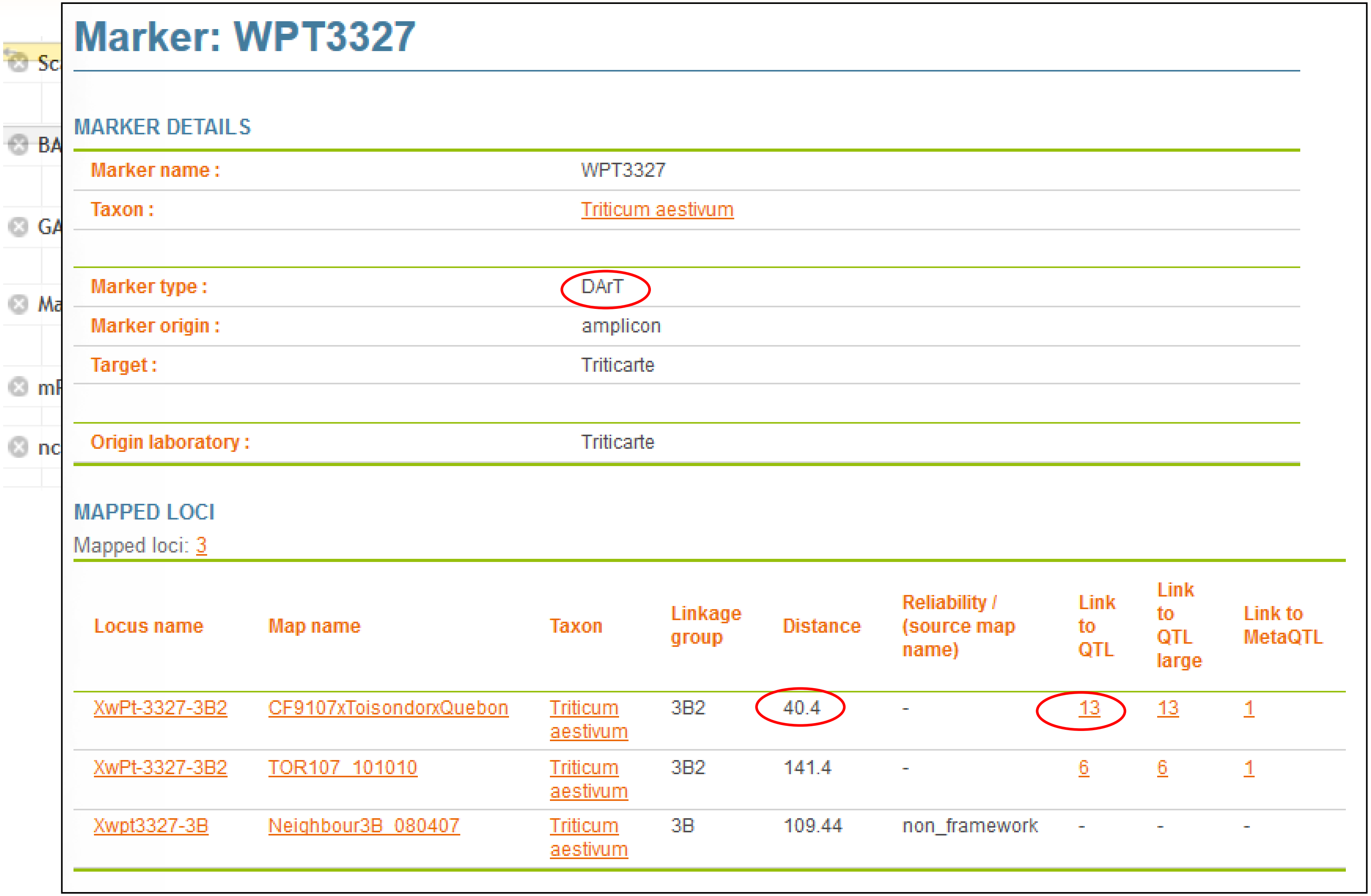
FASTA

```
>traes3bPseudomoleculeV1
traes3bPseudomoleculeV1:121453639..121463164 (+ strand)
class=mRNA length=9526
ATGTGGGCGGCGGAGGCGCCCTTCTCGCGGTCCGGGTCGTGGCGGGAGGCAGAGGACGAGCA
GGAGGCGCTGCGCTGGGCCGCGCTGCAGCGCCTCCCCACCGTGGCGCGTGCCAGGAGGGGCC
TGCTCAGGTCGCCCGCCGTCGCGCCCGGCGGCACCGGCGCGGGCGGCCCGGTCGAGGGCGAC
GACGCGCTCTGCGAGGTCGACGTCACCGGCCTCTCCTCCGGCGACCGCACCGCGCTCGTCGA
CCGCCTGCTCGCAGACTCCGGCGACGCCGAGCAGTTCTTCCGCCGCATACGCGCCCGCTTCG
ACGCGTGAGTACACCAAACCGTGCATCTGCATACTACATTTTTTCGCTCCCGCATCGGTGCA
CCATAGTCATCAGCGTGCCTTCCTCGGTCTCTGCGCGCGCGCGCTCTGCAAATCTACTAGGC
TGACCGCTGACTACAGGGTCCCGCTGCCGCCACAATTGGCTGACACATGGGCCCAACCGACG
```

Subfeatures

https://urgi.versailles.inra.fr/jbrowse/gmod_jbrowse/?data=myData/Wheat/annotation/3B

3B reference JBrowse



Marker: WPT3327

MARKER DETAILS

Marker name :	WPT3327
Taxon :	Triticum aestivum
Marker type :	DArT
Marker origin :	amplicon
Target :	Triticarte
Origin laboratory :	Triticarte

MAPPED LOCI

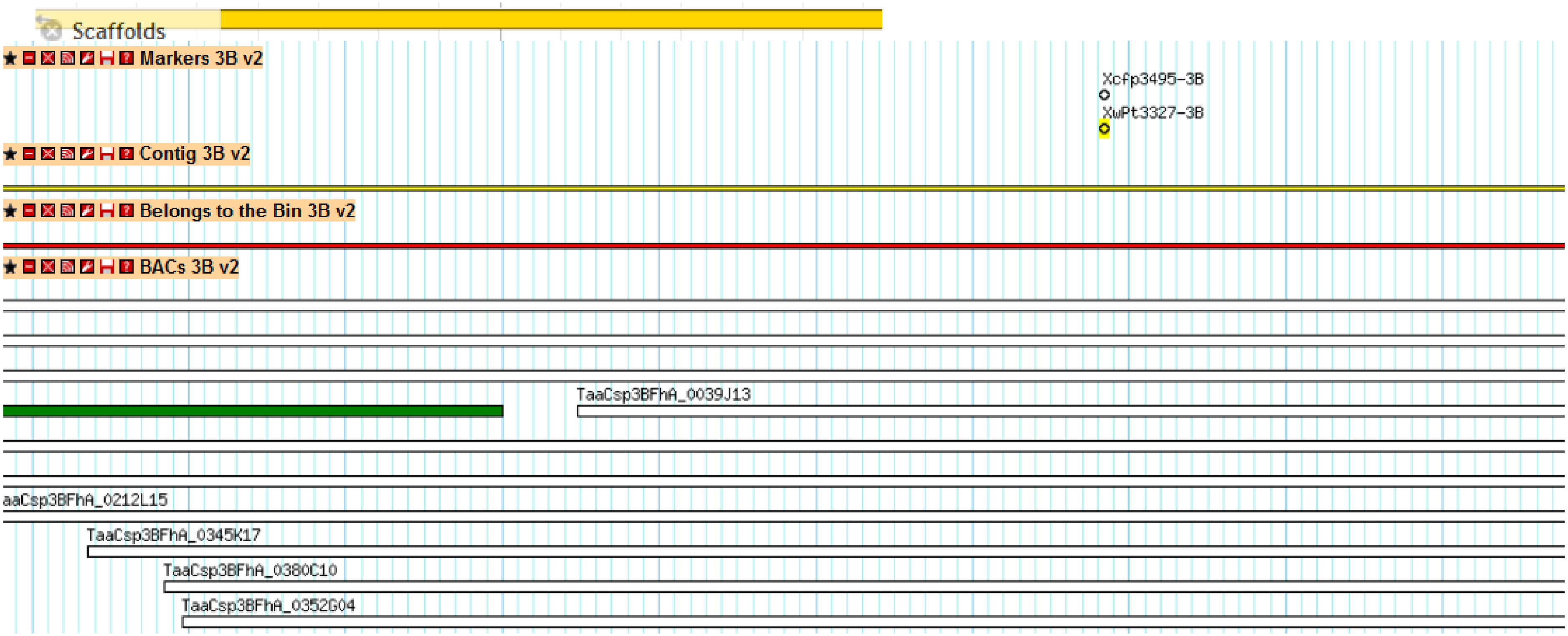
Mapped loci: 3

Locus name	Map name	Taxon	Linkage group	Distance	Reliability / (source map name)	Link to QTL	Link to QTL large	Link to MetaQTL
XwPt-3327-3B2	CF9107xToisondorxQuebon	Triticum aestivum	3B2	40.4	-	13	13	1
XwPt-3327-3B2	TOR107_101010	Triticum aestivum	3B2	141.4	-	6	6	1
Xwpt3327-3B	Neighbour3B_080407	Triticum aestivum	3B	109.44	non_framework	-	-	-

https://urgi.versailles.inra.fr/jbrowse/gmod_jbrowse/?data=myData/Wheat/annotation/3B

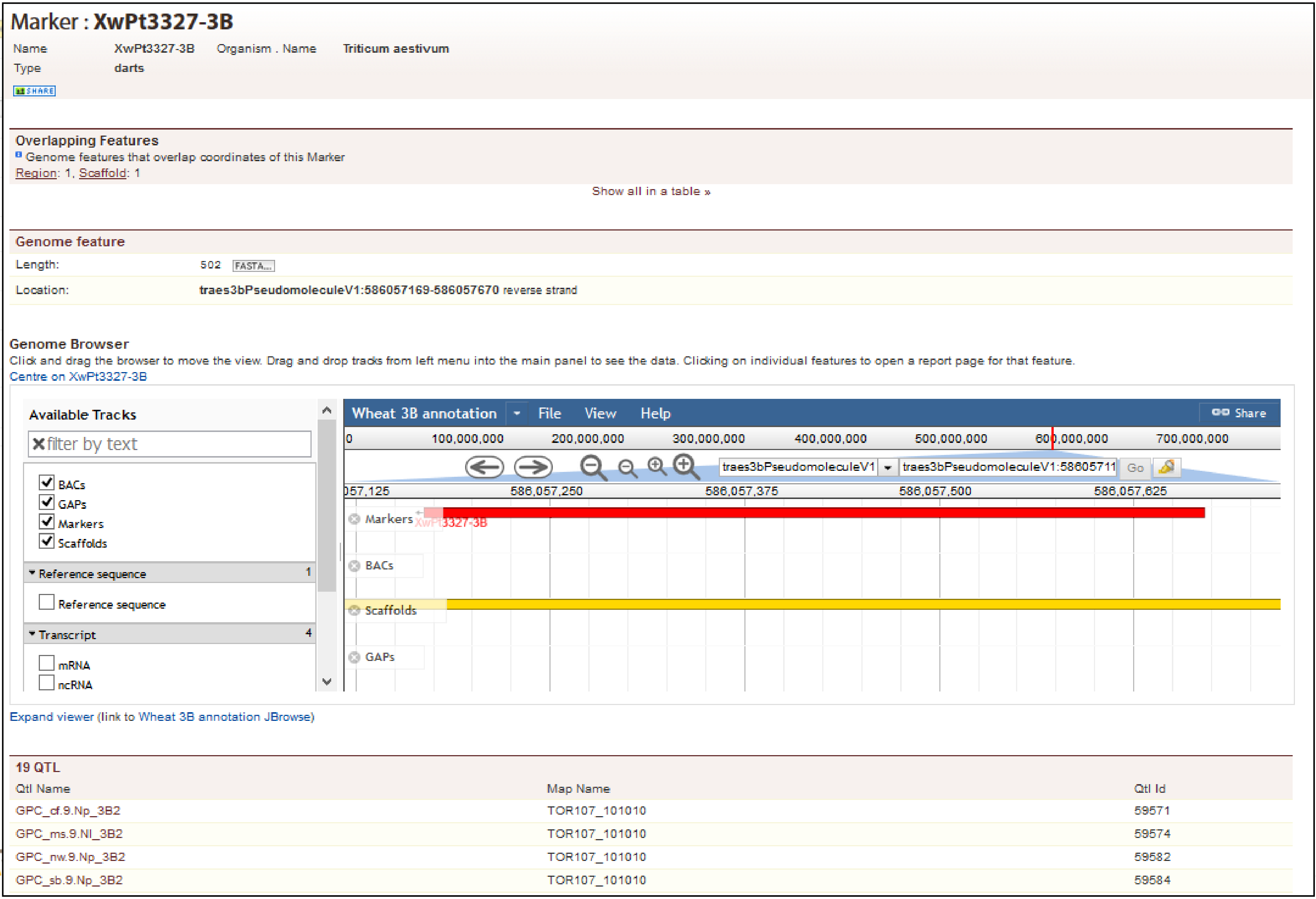
ael Alaux

3B reference JBrowse



https://urgi.versailles.inra.fr/jbrowse/gmod_jbrowse/?data=myData/Wheat/annotation/3B

3B reference JBrowse



Marker : XwPt3327-3B

Name	XwPt3327-3B	Organism . Name	**Triticum aestivum**
Type	darts		

SHARE

Overlapping Features

Genome features that overlap coordinates of this Marker

Region: 1, Scaffold: 1

Show all in a table »

Genome feature

Length:	502 FASTA...
Location:	traes3bPseudomoleculeV1:586057169-586057670 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.

Centre on XwPt3327-3B

Available Tracks

filter by text

☑ BACs
☑ GAPs
☑ Markers
☑ Scaffolds

Reference sequence 1

☐ Reference sequence

Transcript 4

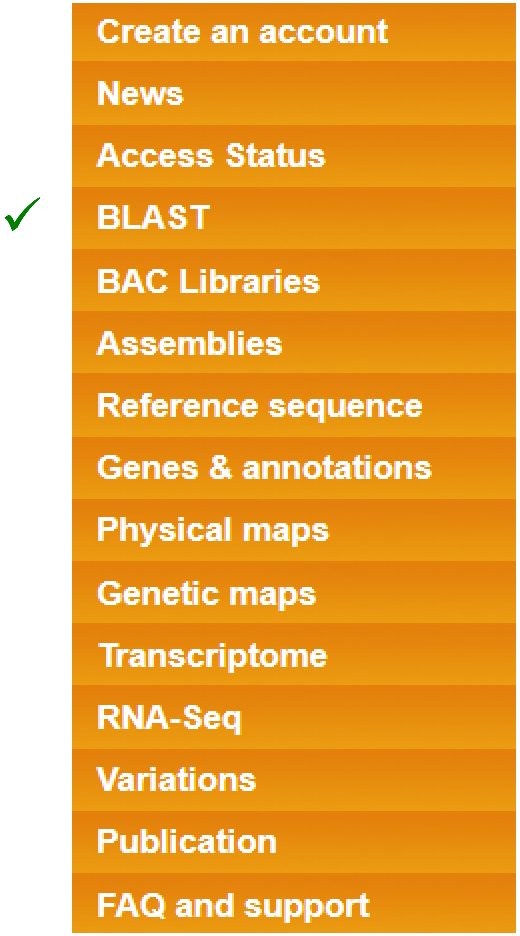
☐ mRNA
☐ ncRNA

Expand viewer (link to Wheat 3B annotation JBrowse)

19 QTL

Qtl Name	Map Name	Qtl Id
GPC_cf.9.Np_3B2	TOR107_101010	59571
GPC_ms.9.NI_3B2	TOR107_101010	59574
GPC_nw.9.Np_3B2	TOR107_101010	59582
GPC_sb.9.Np_3B2	TOR107_101010	59584

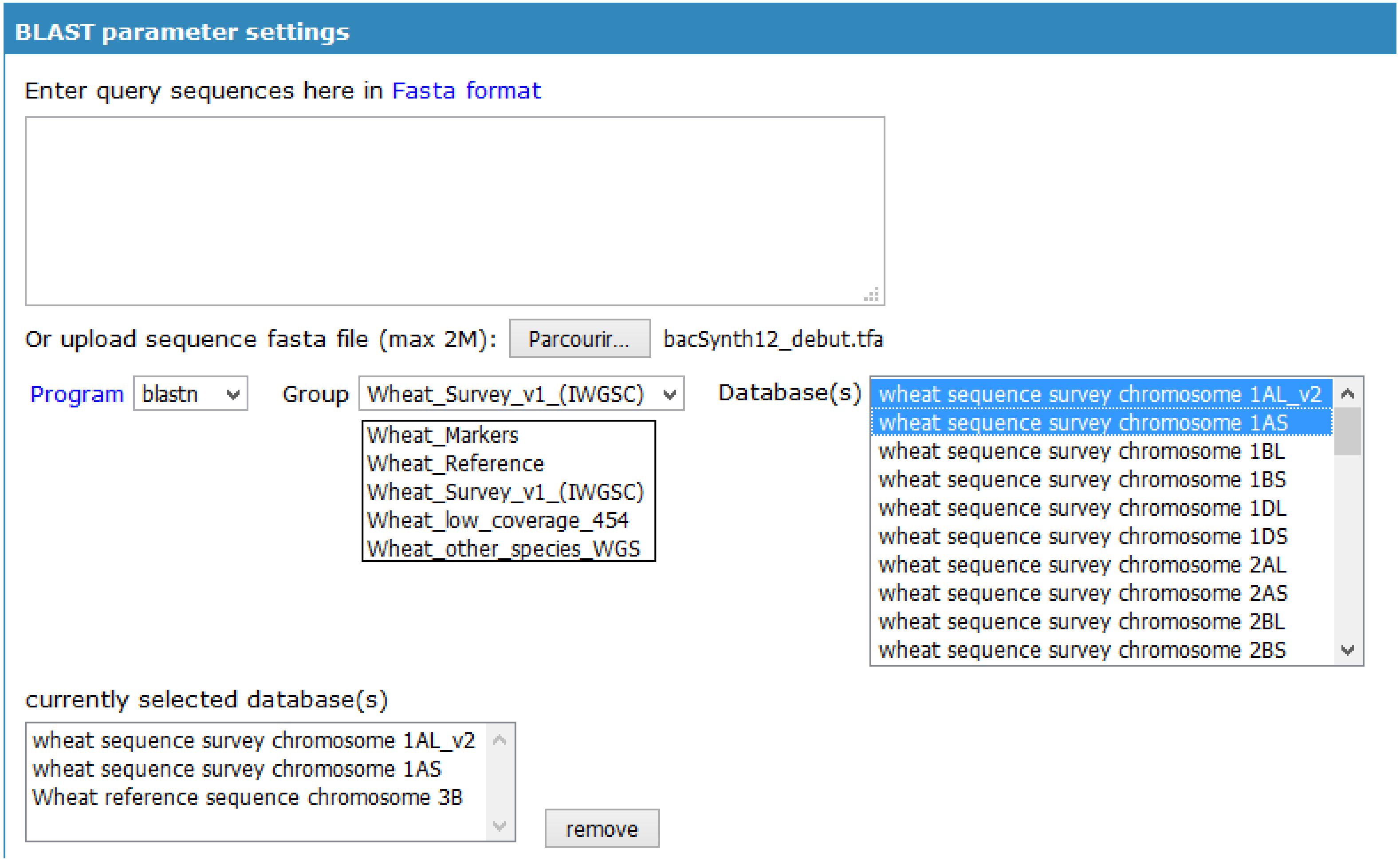
BLAST



News functionalities (feedbacks from PAG 2015)

- addition of a "download all" button
- link from the matching contigs to the browsers (zoomed in the region)
- link to EnsemblPlants chromosomes.

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_Survey_v1_(IWGSC)

Database(s):

- wheat sequence survey chromosome 1AL_v2
- wheat sequence survey chromosome 1AS
- wheat sequence survey chromosome 1BL
- wheat sequence survey chromosome 1BS
- wheat sequence survey chromosome 1DL
- wheat sequence survey chromosome 1DS
- wheat sequence survey chromosome 2AL
- wheat sequence survey chromosome 2AS
- wheat sequence survey chromosome 2BL
- wheat sequence survey chromosome 2BS

currently selected database(s)

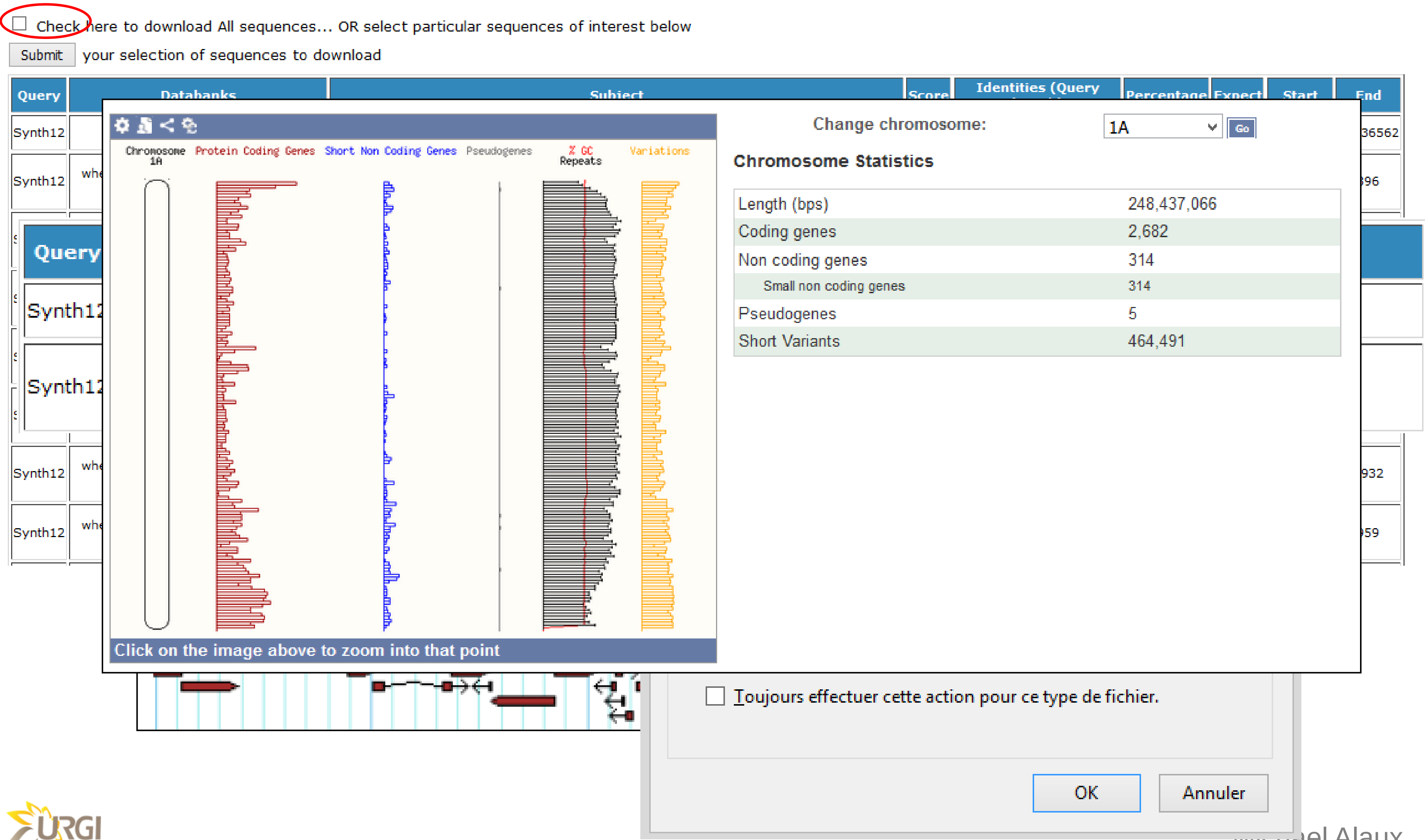
- wheat sequence survey chromosome 1AL_v2
- wheat sequence survey chromosome 1AS
- Wheat reference sequence chromosome 3B

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

URGI

Michael Alaux

BLAST



☐ Check here to download All sequences... OR select particular sequences of interest below

Submit your selection of sequences to download

Click on the image above to zoom into that point

Change chromosome: 1A Go

Chromosome Statistics

Length (bps)	248,437,066
Coding genes	2,682
Non coding genes	314
Small non coding genes	314
Pseudogenes	5
Short Variants	464,491

☐ Toujours effectuer cette action pour ce type de fichier.

What's next?

For 2016

- Physical map browser:
  - Add the 3AL
  - New versions of the 1D, 4D, 6D using WGP

- Survey sequence v3 browser:
  - Add the annotations

- **NRGene IWGSC-WGA** when available:
  - Download
  - BLAST
  - Browser
    - Scaffolds
    - Annotations
    - Your sequences mapped on the scaffolds

For 2016

- Add all the assemblies in the BLAST:
  - Survey sequence v3
  - TGACv1
  - NRGene IWGSC-WGA

Acknowledgements

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

Choulet F.
Paux E.

Rogers J.
Eversole K.
IWGSC Coordinating Committee

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