



**HAL**  
open science

# IWGSC data repository news and overview of the plans for the WheatIS

Michael Alaux

► **To cite this version:**

Michael Alaux. IWGSC data repository news and overview of the plans for the WheatIS. PAG XXX  
- Plant and Animal Genome Conference, Jan 2023, San Diego, United States. hal-04537219

**HAL Id: hal-04537219**

**<https://hal.inrae.fr/hal-04537219>**

Submitted on 8 Apr 2024

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

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



Distributed under a Creative Commons Attribution 4.0 International License



# ➤ IWGSC Data Repository news and WheatIS overview

## IWGSC Business Meeting



# ➤ IWGSC Data Repository news

February 2022:

- Data from the reference sequence of the Turkish elite cultivar Sonmez are now available to **download** (IWGSC **news** ).

April 2021:

- **IWGSC RefSeq v2.1** assembly and annotation is available in open access to download, BLAST and display in a browser (IWGSC **announcement** ).

July 2020:

- Chromatin accessibility data are **available** in open access (E. Akhunov).



**INRAE**

IWGSC Business Meeting

Michael Alaux



# ➤ IWGSC Data Repository news

The screenshot displays the IWGSC BLAST web interface. At the top, a genomic track shows a scale from 0 to 350,000,000. Below the track, several tracks are visible, including 'Reference sequence', 'HighConfidenceGenesv2.1', 'LowConfidenceGenesv2.1', 'TaBW280K', and 'ClariTE\_annotation'. A 'Select tracks' button is located on the left. The main content area is titled 'IWGSC BLAST' and includes a navigation menu (Home, Databanks, About, Help, Contact us) and 'BLAST parameter settings'. A text input field for 'Enter query sequences here in Fasta format' is present, along with a 'Browse...' button and a file name 'bacSynth12\_debut.tfa'. The 'Program' is set to 'blastn' and the 'Group' is 'Wheat\_IWGSC\_RefSeq\_v2.1\_chromosomes'. A dropdown menu for 'Database(s)' is open, showing options like 'IWGSC RefSeq v2.1 all chromosomes' and 'IWGSC RefSeq v2.1 chromosome 1A only'. Below the settings, a table displays search results for 'Synth12' queries. The table has columns for Query, Databanks, Subject, Score, Identities (Query length), Percentage, Expect, Start, and End. The results show high identity scores (e.g., 2805) and high percentages (e.g., 99%) for various chromosome regions.

Query	Databanks	Subject	BEST HSP						
			Score	Identities (Query length)	Percentage	Expect	Start	End	
Synth12	IWGSC RefSeq v2.1 chromosome 3B only	<input type="checkbox"/> Chr3B	2805	1559/1560 (1560)	99	0.0	16057592	16059150	
Synth12	IWGSC RefSeq v2.1 chromosome 4A only	<input type="checkbox"/> Chr4A	848	630/731 (1560)	86	0.0	588816565	588817281	
Synth12	IWGSC RefSeq v2.1 chromosome 7D only	<input type="checkbox"/> Chr7D	845	633/732 (1560)	86	0.0	74968086	74968795	
Synth12	IWGSC RefSeq v2.1 chromosome 5A only	<input type="checkbox"/> Chr5A	836	625/731 (1560)	85	0.0	652964188	652964894	
Synth12	IWGSC RefSeq v2.1 chromosome 3D only	<input type="checkbox"/> Chr3D	816	631/752 (1560)	84	0.0	604203107	604203813	
Synth12	IWGSC RefSeq v2.1 chromosome 7A only	<input type="checkbox"/> Chr7A	814	634/744 (1560)	85	0.0	641413686	641412970	



INRAE

IWGSC Business Meeting

Michael Alaux



## ➤ IWGSC Data Repository metrics

<b>IWGSC Data Repository resources</b>	<b>2020</b>	<b>2021</b>	<b>2022</b>
Data Repository website visits	34525	30249	29160
Downloads	34236	36309	25240
Genome browsers visits	82834	31261	34171
BLAST searches	82968	94920	54029



**INRAE**

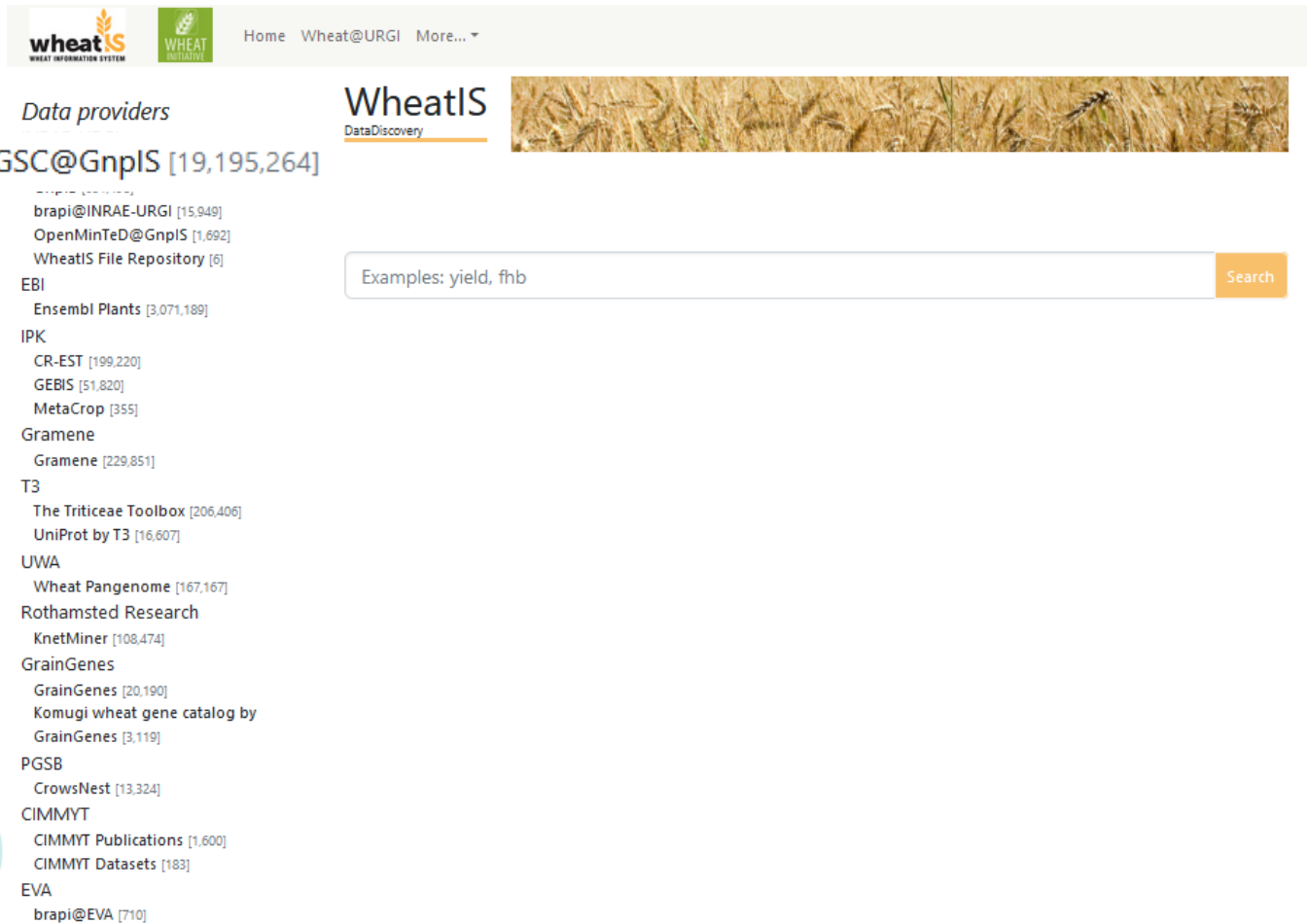
IWGSC Business Meeting

Michael Alaux



# ➤ WheatIS news

- New version of the WheatIS Data Discovery tool
  - Ontologies
  - BrAPI sources



The screenshot shows the WheatIS Data Discovery tool interface. At the top, there is a navigation bar with the WheatIS logo (Wheat Information System) and the Wheat Initiative logo, along with links for Home, Wheat@URGI, and More... A search bar is located in the top right corner with the text "Examples: yield, fhb" and a "Search" button. Below the navigation bar, the main content area is titled "WheatIS DataDiscovery" and features a banner image of a wheat field. On the left side, there is a list of "Data providers" including IWGSC@GnpIS [19,195,264], brapi@INRAE-URGI [15,949], OpenMinTeD@GnpIS [1,692], WheatIS File Repository [6], EBI, Ensembl Plants [3,071,189], IPK, CR-EST [199,220], GEBIS [51,820], MetaCrop [355], Gramene, Gramene [229,851], T3, The Triticeae Toolbox [206,406], UniProt by T3 [16,607], UWA, Wheat Pangenome [167,167], Rothamsted Research, KnetMiner [108,474], GrainGenes, GrainGenes [20,190], Komugi wheat gene catalog by GrainGenes [3,119], PGSB, CrowsNest [13,324], CIMMYT, CIMMYT Publications [1,600], CIMMYT Datasets [183], and EVA, brapi@EVA [710].



## ➤ WheatIS news

- Short demo

<https://urgi.versailles.inrae.fr/wheatis/>



**INRAE**

IWGSC Business Meeting

Michael Alaux



## ➤ WheatIS plan

- Data updates of existing sources
  - Ensembl, GrainGenes, Gramene
- Addition of new IWGSC sources
  - Apollo from Australia
- Promotion of the WheatIS Data Discovery tool
  - Talks
  - Videos
- Workshop to support the use of the wheat gene nomenclature and guidelines



INRAE

IWGSC Business Meeting

Michael Alaux

