Updates on the WheatIS tools

Michael Alaux

HAL Id: hal-04536526
https://hal.inrae.fr/hal-04536526

Submitted on 8 Apr 2024

Update on the WheatIS tools

Michael Alaux
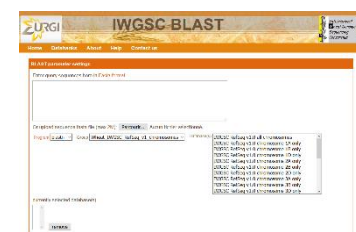
WheatIS data discovery tool

Michael Alaux

SolR search

http://www.wheatis.org/

Michael Alaux

SolR search

Michael Alaux

New data in 2017

transPlant-MIPS (UP):

- CrowsNest: 13324

transPlant-IPK (UP):

- CR-EST: 199220
- GEBIS: 51814
- MetaCrop: 355

transPlant-EBI (UP):

- Ensembl Plants: 218282

transplant-IPGPAS (UP):

- PlantPhenoDB: 2

transplant-IPGPAS (UP):

- PlantPhenoDB: 2

T3 (UP):

- Triticeae Toolbox: 185781

CIMMYT (UP):

- CIMMYT Dspace: 918
- CIMMYT dataverse: 37

UWA (UP):
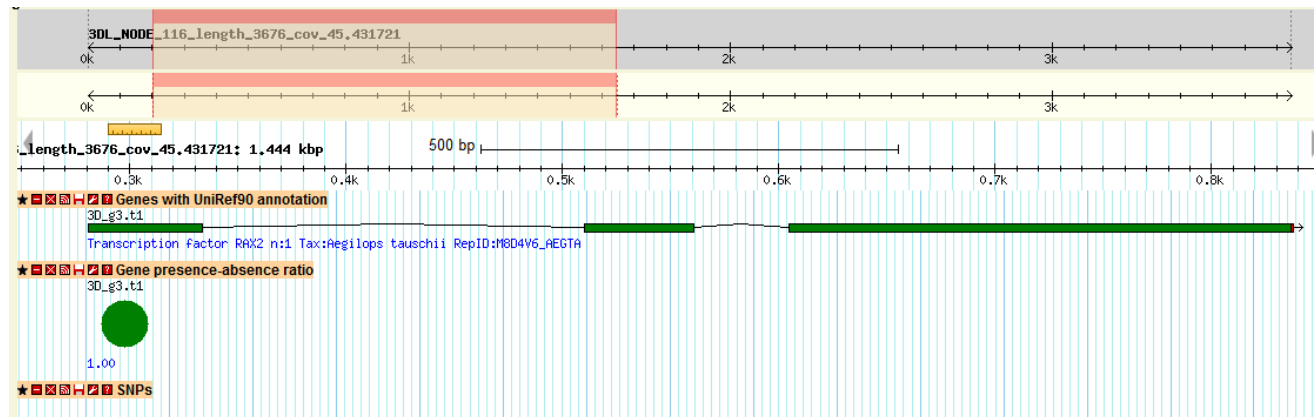
- Wheat Pangenome: 167167   GBrowse

GrainGenes (UP):

- GrainGenes: 548   QTLs

Gramene-WheatIS (UP):

- Gramene: 229851   Genes, GO, pathways, etc.

UWA Wheat pangenome

ID	Source	Type	Taxon	Description
3D_g3.t1	Wheat Pangenome	SEQUENCE FEATURE	Triticum aestivum	SEQUENCE FEATURE, Wheat Pangenome, 3D_g3.t1, Presence=1.00, Transcription factor RAX2 n:1 Tax:Aegilops tauschii RepID:M8D4V6_AEGTA, Triticum aestivum, mRNA, 3D_g3.t1
3D_g7.t1	Wheat Pangenome	SEQUENCE FEATURE	Triticum aestivum	SEQUENCE FEATURE, Wheat Pangenome, 3D_g7.t1, Presence=1.00, Uncharacterized protein n:8 Tax:Triticeae RepID:A0A077S4R6_WHEAT, Triticum aestivum, mRNA, 3D_g7.t1
3D_g9.t1	Wheat Pangenome	SEQUENCE FEATURE	Triticum aestivum	SEQUENCE FEATURE, Wheat Pangenome, 3D_g9.t1, Presence=1.00, Uncharacterized protein n:1 Tax:Triticum aestivum RepID:A0A080YUN9_WHEAT, Triticum aestivum, mRNA, 3D_g9.t1
3D_g9.t2	Wheat Pangenome	SEQUENCE FEATURE	Triticum aestivum	SEQUENCE FEATURE, Wheat Pangenome, 3D_g9.t2, Presence=1.00, Uncharacterized protein n:1 Tax:Triticum aestivum RepID:A0A080YUN9_WHEAT, Triticum aestivum, mRNA, 3D_g9.t2
3D_g11.t1	Wheat Pangenome	SEQUENCE FEATURE	Triticum aestivum	SEQUENCE FEATURE, Wheat Pangenome, 3D_g11.t1, Presence=1.00, No Hits, Triticum aestivum, mRNA, 3D_g11.t1
3D_g12.t1	Wheat Pangenome	SEQUENCE FEATURE	Triticum aestivum	SEQUENCE FEATURE, Wheat Pangenome, 3D_g12.t1, Presence=1.00, Uncharacterized protein n:5 Tax:Triticeae RepID:W5D2X2_WHEAT, Triticum aestivum, mRNA, 3D_g12.t1
3D_g15.t1	Wheat Pangenome	SEQUENCE FEATURE	Triticum aestivum	SEQUENCE FEATURE, Wheat Pangenome, 3D_g15.t1, Presence=1.00, Uncharacterized protein n:1 Tax:Aegilops tauschii RepID:R7WFP4_AEGTA, Triticum aestivum, mRNA, 3D_g15.t1
3D_g21.t1	Wheat Pangenome	SEQUENCE FEATURE	Triticum aestivum	SEQUENCE FEATURE, Wheat Pangenome, 3D_g21.t1, Presence=1.00, Galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase 2 n:2 Tax:Triticeae RepID:N1QQK4_AEGTA, Triticum aestivum, mRNA, 3D_g21.t1
3D_g23.t1	Wheat Pangenome	SEQUENCE FEATURE	Triticum aestivum	SEQUENCE FEATURE, Wheat Pangenome, 3D_g23.t1, Presence=1.00, GDSL esterase/lipase n:1 Tax:Aegilops tauschii RepID:M8BY45_AEGTA, Triticum aestivum, mRNA, 3D_g23.t1
3D_g27.t1	Wheat Pangenome	SEQUENCE FEATURE	Triticum aestivum	SEQUENCE FEATURE, Wheat Pangenome, 3D_g27.t1, Presence=1.00, Uncharacterized protein n:8 Tax:Triticeae RepID:M0ZDK5_HORVD, Triticum aestivum, mRNA, 3D_g27.t1

Graingenes QTLs

ID	Source	Type	Taxon	Description
QAmc.ocs-4A.1	GrainGenes	QTL	Triticum aestivum	QTL, GrainGenes, QAmc.ocs-4A.1, Amylose content; Identification of genetic loci affecting amylose content and agronomic traits on chromosome 4A of wheat., Triticum aestivum
QCc.mst-2D	GrainGenes	QTL	Triticum aestivum	QTL, GrainGenes, QCc.mst-2D, Wheat stem sawfly attraction; Identification of Novel QTL for Sawfly Resistance in Wheat, Triticum aestivum
QCc.mst-4A	GrainGenes	QTL	Triticum aestivum	QTL, GrainGenes, QCc.mst-4A, Wheat stem sawfly attraction; Identification of Novel QTL for Sawfly Resistance in Wheat, Triticum aestivum
QCkf.ucw-7B	GrainGenes	QTL	Triticum turgidum ssp. durum	QTL, GrainGenes, QCkf.ucw-7B, Cooked firmness; QTL analysis of pasta quality using a composite microsatellite and SNP map of durum wheat., Triticum turgidum ssp. durum
QCol.wak-2B	GrainGenes	QTL	Triticum aestivum	QTL, GrainGenes, QCol.wak-2B, Leaf color/glaucousness; Identifying QTL for high-temperature adult-plant resistance to stripe rust (Puccinia striiformis f. sp. tritici) in the spring wheat (Triticum aestivum L.) cultiva[...]
QDon.crc-3B	GrainGenes	QTL	Triticum aestivum	QTL, GrainGenes, QDon.crc-3B, Deoxynivalenol (DON) accumulation; Molecular mapping of novel genes controlling Fusarium head blight resistance and deoxynivalenol accumulation in spring wheat, Triticum aestivum
QDon.crc-5A	GrainGenes	QTL	Triticum aestivum	QTL, GrainGenes, QDon.crc-5A, Deoxynivalenol (DON) accumulation; Molecular mapping of novel genes controlling Fusarium head blight resistance and deoxynivalenol accumulation in spring wheat, Triticum aestivum

GrainGenes QTL Report: QAmc.ocs-4A.1

[Printable Version] [Submit comment/correction]

QTL	QAmc.ocs-4A.1
Trait Affected	Amylose content
Ontology	TO:0000196 Planteome amylose content
Chromosome	4AS
Positive Significant Marker	Xbcd1738-4A Xcdo1387-4A
Species	Triticum aestivum
Reference	Araki E et al. (1999) Identification of genetic loci affecting amylose co[...] 98:977-984.

ap based on quantitative trait

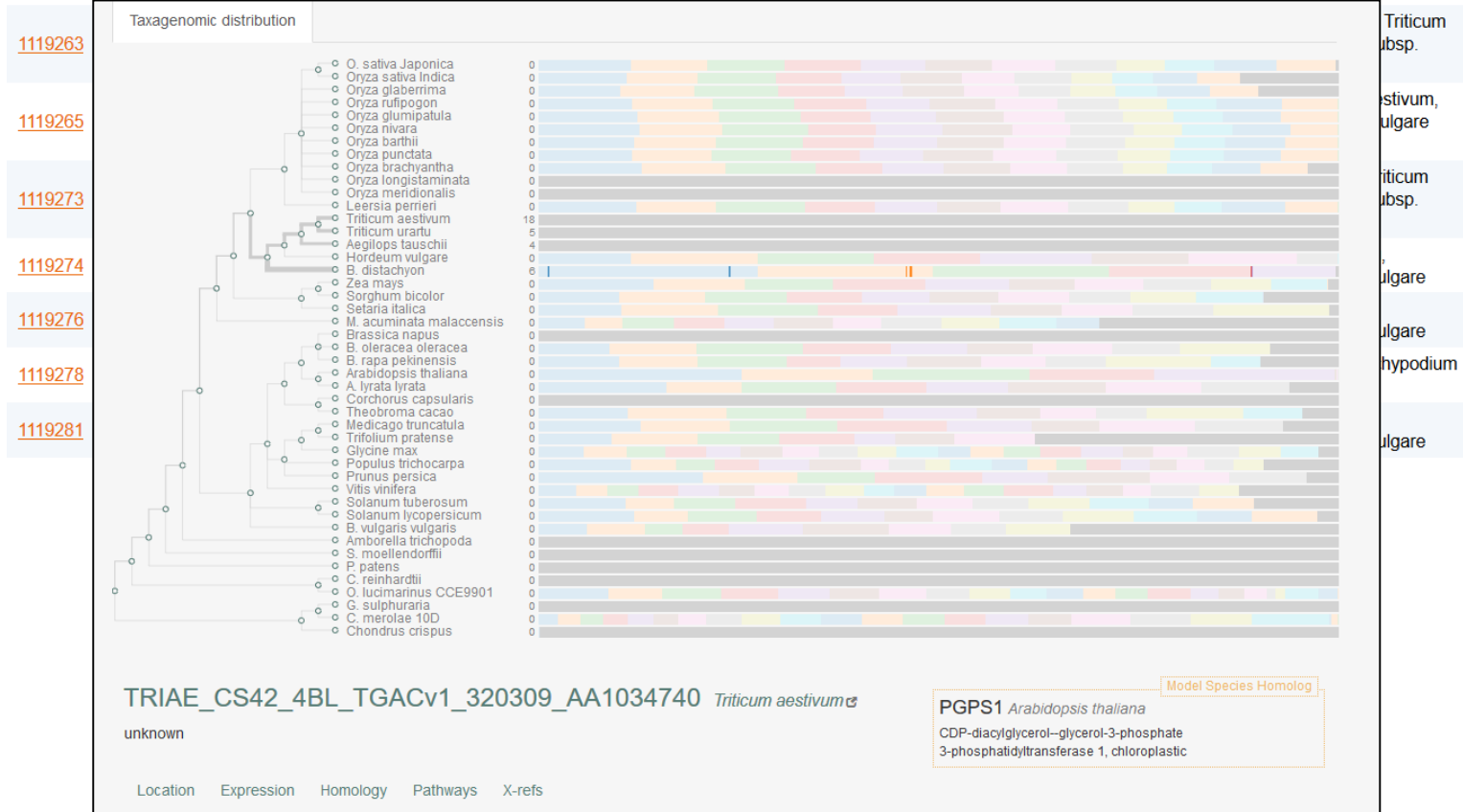## GrainGenes Locus Report: Xbcd1738-4A

[Printable Version] [Submit comment/correction]

Locus	Xbcd1738-4A [Marker Report]
Type	RFLP
Chromosome	4A
Chromosome Arm	4AS
Map	Ta-Synthetic/Opata-4A 19.5 Ta-Synthetic/Opata-BARC-4A 18.4 Ta-Synthetic/Opata-GPW-4A 26.5 *[Show Nearby Loci]*
In QTL	QAmc.ocs-4A.1 *[Show all 6]*
Map Data	Wheat, Synthetic x Opata [GBrowser] Wheat, Synthetic x Opata, BARC Wheat, Synthetic x Opata, GPW
Probe	BCD1738
Homology	CAB61839.1 e-value: 9e-37 (AJ242803) putative serine/threonine phosphatase type 2c [Sporobolus stapfianus]
Image	BCD1738 SO autoradiogram

URGI

Michael Alaux

Gramene

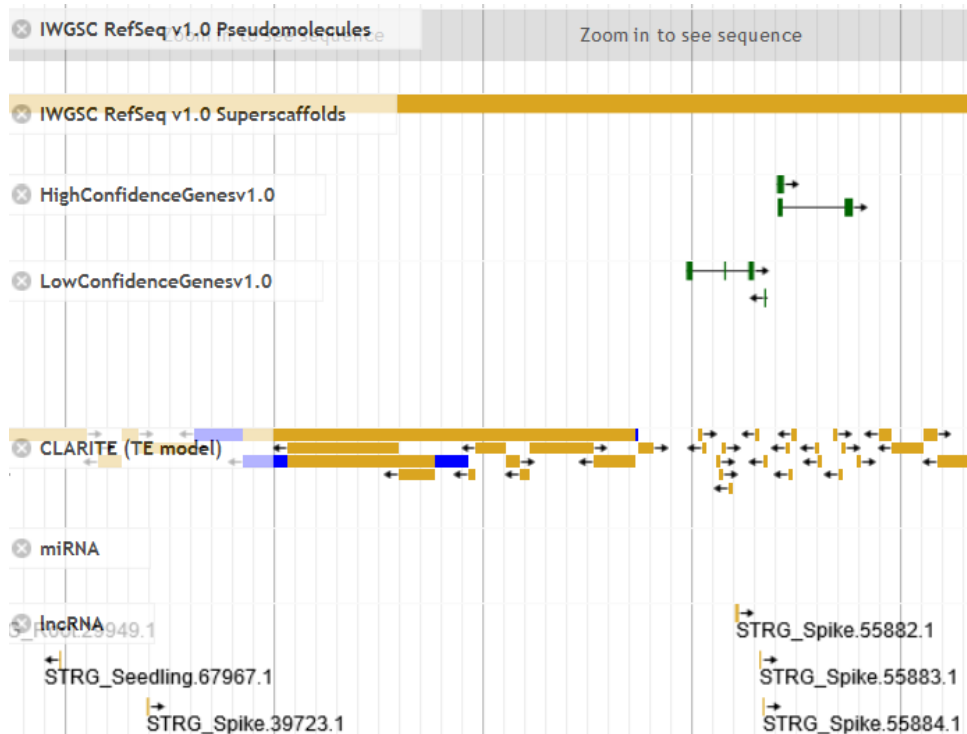| ID	Source	Type	Taxon	Description
1119260	Gramene	Pathway	Triticum aestivum, Triticum urartu, Brachypodium distachyon, Aegilops tauschii, Hordeum vulgare subsp. vulgare	Gramene, Pathway, 1119260, cardiolipin biosynthesis, Triticum aestivum, Triticum urartu, Brachypodium distachyon, Aegilops tauschii, Hordeum vulgare subsp. vulgare
1119261	Gramene	Pathway	Triticum aestivum, Triticum urartu, Brachypodium distachyon, Aegilops tauschii, Hordeum vulgare subsp. vulgare	Gramene, Pathway, 1119261, salicylate biosynthesis, Triticum aestivum, Triticum urartu, Brachypodium distachyon, Aegilops tauschii, Hordeum vulgare subsp. vulgare
1119262	Gramene	Pathway	Triticum aestivum, Triticum urartu, Brachypodium distachyon, Aegilops tauschii, Hordeum vulgare subsp. vulgare	Gramene, Pathway, 1119262, threonine biosynthesis from homoserine, Triticum aestivum, Triticum urartu, Brachypodium distachyon, Aegilops tauschii, Hordeum vulgare subsp. vulgare
1119263				
1119265				
1119273				
1119274				
1119276				
1119278				
1119281				

Taxagenomic distribution

Species	Count
O. sativa Japonica	0
Oryza sativa Indica	0
Oryza glaberrima	0
Oryza rufipogon	0
Oryza glumipatula	0
Oryza nivara	0
Oryza barthii	0
Oryza punctata	0
Oryza brachyantha	0
Oryza longistaminata	0
Oryza meridionalis	0
Leersia perrieri	0
Triticum aestivum	18
Triticum urartu	5
Aegilops tauschii	4
Hordeum vulgare	0
B. distachyon	6
Zea mays	0
Sorghum bicolor	0
Setaria italica	0
M. acuminata malaccensis	0
Brassica napus	0
B. oleracea oleracea	0
B. rapa pekinensis	0
Arabidopsis thaliana	0
A. lyrata lyrata	0
Corchorus capsularis	0
Theobroma cacao	0
Medicago truncatula	0
Trifolium pratense	0
Glycine max	0
Populus trichocarpa	0
Prunus persica	0
Vitis vinifera	0
Solanum tuberosum	0
Solanum lycopersicum	0
B. vulgaris vulgaris	0
Amborella trichopoda	0
S. moellendorffii	0
P. patens	0
C. reinhardtii	0
O. lucimarinus CCE9901	0
G. sulphuraria	0
C. merolae 10D	0
Chondrus crispus	0

TRIAE_CS42_4BL_TGACv1_320309_AA1034740 *Triticum aestivum*

unknown

Model Species Homolog

PGPS1 *Arabidopsis thaliana*
CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase 1, chloroplastic

Location   Expression   Homology   Pathways   X-refs

URGI

Michael Alaux

In progress at URGI

URGI (UP):

- IWGSC@GnpIS: 23295514   New data from the IWGSC RefSeq v1.0
- GnpIS: 58000
- OpenMinTeD: 3398   New data from publications
- WheatIS File Repository: 6

Michael Alaux

In progress at URGI

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 [...]

Michael Alaux

In progress at URGI

Bug fixed

- It is now possible to find metadata containing special caracters such as « . », « - », « _ »

SolR package

You'd like to join the WheatIS federation in order to make your data discoverable through the WheatIS search tool ?

You should contact us to have support and to choose the best option that fits your needs.

- The simpler option is to extract your data in the dedicated CSV format and to provide us this file (either stored on your web server, or by any other way). We then will be able to fetch and index this file at URGI and make your data visible within the WheatIS search tool.
- If you want to keep full control over your data, you can also install and configure an on-premise Solr instance. Once installation completed, some data indexed and network opened to URGI's servers, we'll be able to add your node to our federation and you'll be visible in the WheatIS search tool.

To ease the WheatIS nodes to expose their data into the WheatIS search tool, we developed a package to help the installation, the configuration and the management of a WheatIS node, or to get primary insights about how to format your data. Download the package:

WheatIS-search-tool-
sha1_solr_7fec39d8006b2c2163ddc72f294c65b17358f46f_etl_script_c4cb7cfb6d139b75640133ddd78b67dbdcf04866.tar.gz
(415.17 kB)

Please contact us to get support and to add your data to the WheatIS search.

Powered by

Apache
Solr

Michael Alaux
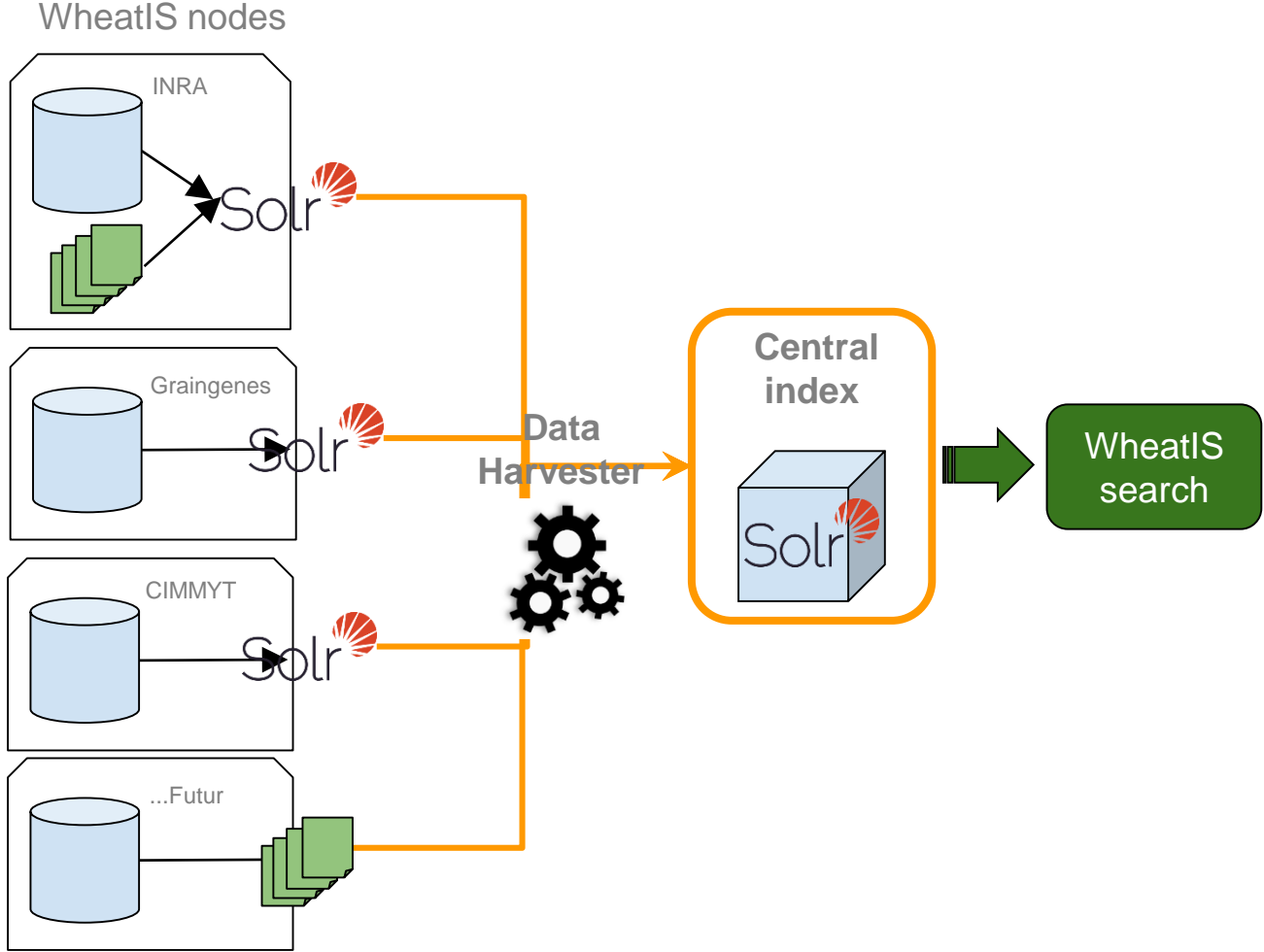
Perspectives

Perspectives

- Central index approach
  - FTP/Web repository on each node (like Trackhub)
  - Metadata harvested / indexed
    - CSV Files
    - Export from existing solr with sligth configuration adaptation

- Pros
  - Lighter implementation for new nodes
  - Easy update of the Solr engine
  - Easy addition of new functionalities

- Cons / things to check
  - Centralisation of queried metadata
    - Description metadata only
    - All content of CSV files
  - Delay in index updates (day or week)

Data standards

Data standards

http://www.wheatis.org/

Data standards

- News will be presented by Hadi Quesneville.

Michael Alaux

WheatIS data submission

Michael Alaux

DSpace - iRODS

http://www.wheatis.org/

Michael Alaux

DSpace - iRODS

Michael Alaux

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

DSpace - iRODS

- DSpace metadata are indexed in the SoIR search

- We will add IWGSC public data / metadata soon

Michael Alaux

Acknowledgements

Michael Alaux

Questions

WheatIS:

http://www.wheatis.org/

Download the SolR package to add your node in the search:

http://wheat-urgi.versailles.inra.fr/Projects/Wheat-Information-System/SolR-tool-package

Contacts: michael.alaux@inra.fr

raphael.flores@inra.fr

thomas.letellier@inra.fr

wheatis-contact@wheatis.org