

# From Farm to FAIR

A Data Sharing Infrastructure for Designing Future Wheat

Chris Rawlings

Computational and Analytical Sciences - Rothamsted Research

DFW WP4 Lead: Rob Davey (EI)

Co-Lead: Chris Rawlings (RRes)

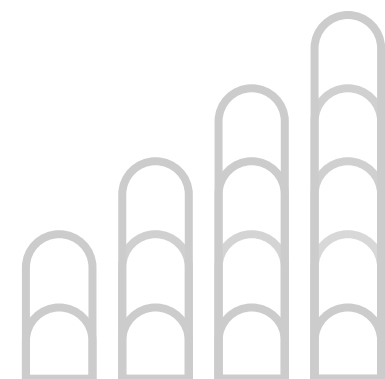


A decorative grey square with rounded corners is positioned in the top-left corner of the slide.

# Wide range of of data to be managed

In DFW combination of “traditional” breeding trials, bespoke research trials and fundamental research datasets creates a challenge

- Field trial datasets – agronomic and developmental traits
- Diversity set genotyping
- Physical phenotyping
- Chemical phenotyping
- High Throughput phenotyping platform data and UAV imaging
- Pangenomic datasets i.e. Wheat 10+
- Epigenetic datasets
- Single cell genomics





# Data Sharing in DFW

- DFW has obligations because of its significance and size as a UK publicly-funded (UKRI-BBSRC) R&D project
- Sharing data within the project is essential to the collaboration
- Important that data management and sharing is not solely the responsibility of the informatics teams, but shared across the project.
- **Data Coordination Task Force** brings different work streams together to coordinate efforts and share best practice.
  - Including breeder representative (Ruth Bryant – RAGT)
  - Ensures development of the data sharing tools and ensures interoperability between the specialist bioinformatic and genomic data resources of individual partners

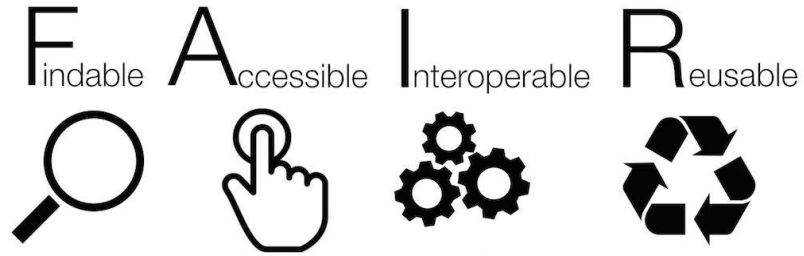


A decorative grey square in the top-left corner.

# Sharing data with other stakeholders

- Breeders – Toolkit data. Bringing data back into project after 2 year embargo  
<https://designingfuturewheat.org.uk/breeders-toolkit/>
- Funders (UKRI-BBSRC)
  - On demand access to catalogue of outputs from the project
    - Publications
    - Conference participation and presentations
    - Knowledge exchange outside research community
    - Digital outputs (datasets and software)
    - Public engagement
    - Collaborations
- Research Community – commitment to data sharing following FAIR principles





- F** The first step in (re)using data is to **find** them. Metadata and data should be easy to find for both humans and computers.
- metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process
- A** Once the user finds the required data, they need to know how they can be **accessed**
- possibly including authentication and authorisation
- I** The data usually need to be **interoperable** with other data.
- E.g. to interoperate with applications or workflows for analysis, storage, and processing
- R** The ultimate goal of FAIR is to optimise the **reuse** of data.
- To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.



# Main DFW-funded Data Resources

Wheat Germplasm Resource	<a href="http://www.seedstor.ac.uk">www.seedstor.ac.uk</a>
CerealsDB	<a href="http://www.cerealsdb.uk.net">www.cerealsdb.uk.net</a>
DFW Field Trials	<a href="http://grassroots.tools/dfw">http://grassroots.tools/dfw</a>
Ensembl Plants - Wheat	<a href="http://plants.ensembl.org/Triticum_aestivum">http://plants.ensembl.org/Triticum_aestivum</a>
KnetMiner	<a href="https://knetminer.com/Triticum_aestivum">https://knetminer.com/Triticum_aestivum</a>
Wheat Expression Browser	<a href="http://www.wheat-expression.com">www.wheat-expression.com</a>
DFW Data Portal	<a href="http://opendata.earlham.ac.uk/wheat">http://opendata.earlham.ac.uk/wheat</a>
DFW Digital Repository	<a href="http://ckan.grassroots.tools">http://ckan.grassroots.tools</a>



# Interoperability – Software and Data



### Grassroots Tools

#### BROWSE FIELD TRIALS

MyLocation



Grassroots Data Repository

Using 128M capture to profile the Watkins bread wheat landrace diversity collection

Grassroots Data Repository

Name	Size	Date	Properties
Watkins_02_watkins_capture_128M		2019-02-29 18:18	
Watkins_02_watkins_capture_128M		2019-02-29 18:18	

#### Designing Future Wheat

Search data

Provider tags: [Germplasm](#) [Genetics](#) [Genomics](#)

DFW CAN statistics

4 experiments 8 plots

### SeedStor

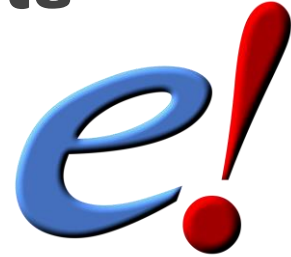
#### SeedStor Homepage

Created by the Germplasm Resources Unit

Quick search Accession Name, Number and StoreCode fields:

Quick Search

Germplasm Resources Unit



### knetminer

dormancy OR germination OR color OR Flavon\* OR proanthocyanidin

Query Suggester

Genome or QTL Search

Gene List

- Tracc2.1288100017988
- Tracc2.1288100017989
- Tracc2.1288100017990
- Tracc2.1288100017991
- Tracc2.1288100017992
- Tracc2.1288100017993
- Tracc2.1288100017994
- Tracc2.1288100017995
- Tracc2.1288100017996

ROTHAMSTED RESEARCH







# Field Trials Grassroots Service

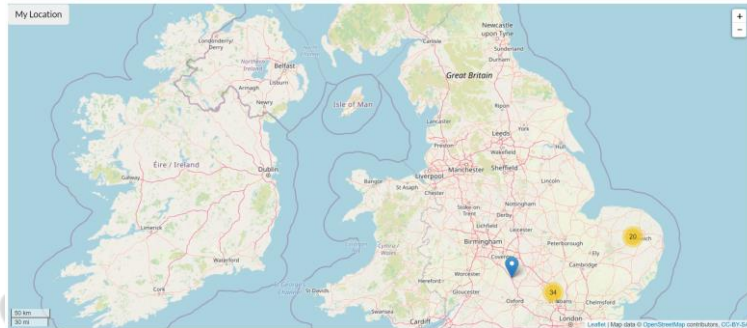
- Provide a template for submitting metadata about DFW Field Experiments
  - To allow investigators to review ongoing and planned experiments and identify opportunities for adding value to experiments
  - Submitted using web-based forms and uploading spreadsheets
- Standardised template for submitting the genotype and phenotype data from DFW field trials
  - Working with field managers and biologists to standardize these spreadsheets and collection mechanisms
  - To facilitate publishing of data compliant with FAIR sharing principles
- Working with individual PIs to facilitate publication of datasets and integration with resources such as Ensembl Plants



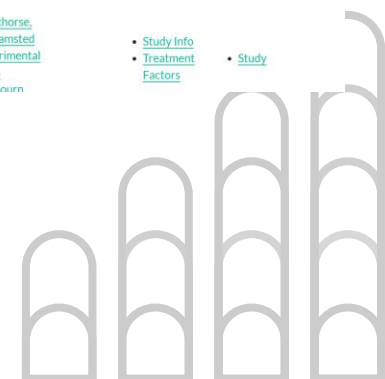
# Field Trials - Findable

- 55 Trials currently submitted
- The experimental data can be accessed using a map-based view and a searchable table of the data...

## BROWSE FIELD TRIALS



Programme	Field Trial	Study	Team	Description	Sowing Date	Harvest Date	Plots	Address	Shape Data	Popup Info	Links
Designing Future Wheat	DFW - Designing Future Wheat - Work package 2 (WP2) - Added value and resilience	1st vs 3rd wheat take-all resistance trial	Rothamsted Research					<a href="#">Stackyard RES St Albans UK AL5 2BQ</a>		<ul style="list-style-type: none"> <li>• <a href="#">Study Info</a></li> <li>• <a href="#">Treatment Factors</a></li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Study</a></li> </ul>
Designing Future Wheat	6th Year Paragon x Watkins Mapping populations 2017	2017 DFW Paragon x Watkins Mapping Populations 6th Year	Andrew Riche	7 PxW Mapping populations grown at 2 N levels plus 2 Robigus x Watkins mapping populations	2016-10-19	2017-08-15		<a href="#">Meadow, Rothamsted Experimental Farm Redbourn GB</a>		<ul style="list-style-type: none"> <li>• <a href="#">Study Info</a></li> <li>• <a href="#">Treatment Factors</a></li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Study</a></li> </ul>
Designing Future Wheat	DFW - Designing Future Wheat - Work package 2 (WP2) - Added value and resilience	AxC NILs (2)	Rothamsted Research					<a href="#">Black Horse St Albans United Kingdom AL3 7PX</a>		<ul style="list-style-type: none"> <li>• <a href="#">Study Info</a></li> <li>• <a href="#">Treatment Factors</a></li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Study</a></li> </ul>
Designing Future Wheat	DFW Academic Toolkit RRes	DFW Academic Toolkit RRes, Harvest 2016	Andrew Riche					<a href="#">Blackhorse, Rothamsted Experimental Farm Redbourn</a>		<ul style="list-style-type: none"> <li>• <a href="#">Study Info</a></li> <li>• <a href="#">Treatment Factors</a></li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Study</a></li> </ul>



# Field Trials - Findable

- ... or via a text-based search web page

## SEARCH FIELD TRIALS

A service to search field trial data

For more information and help, go to the [user documentation](#)

Simple options  Advanced options

Search

toolkit

Type

Any

Page

0

Page size

10

Submit

Show 10 entries

Search:

Rank	Type	Title	Info	Link
1	Study	DFW Toolkit lines	Broad Mead UK MK43 0XF	<a href="#">View Study</a>
2	Field Trial	DFW WP3 - DFW Academic Toolkit Trials	DFW WP3	<a href="#">View Field Trial</a>
3	Field Trial	DFW WP3 - DFW Breeders Toolkit Trials	DFW WP3	<a href="#">View Field Trial</a>
4	Field Trial	Andrew Riche - DFW Academic Toolkit RRes	Andrew Riche	<a href="#">View Field Trial</a>
5	Study	DFW Academic Toolkit Trial H2019	Black Horse St Albans United Kingdom AL3 7PX	<a href="#">View Study</a>
6	Study	DFW Toolkit lines 2nd year	Black Horse St Albans United Kingdom AL3 7PX	<a href="#">View Study</a>
7	Study	DFW Academic Toolkit RRes Harvest 2020	Meadow, Rothamsted Experimental Farm Redbourn	<a href="#">View Study</a>



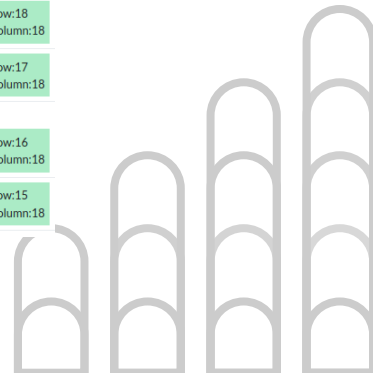
# Field Trials

Each Study has a plot-based tabular view

## FIELD TRIAL DETAILS

Study: DFW Academic Toolkit RRes, Harvest 2020

20	Row:20 Column:1	Row:20 Column:2	Row:20 Column:3	Row:20 Column:4	Row:20 Column:5	Row:20 Column:6	Row:20 Column:7	Row:20 Column:8	Row:20 Column:9	Row:20 Column:10	Row:20 Column:11	Row:20 Column:12	Row:20 Column:13	Row:20 Column:14	Row:20 Column:15	Row:20 Column:16	Row:20 Column:17	Row:20 Column:18
19	Row:19 Column:1	Row:19 Column:2	Row:19 Column:3	Row:19 Column:4	Row:19 Column:5	Row:19 Column:6	Row:19 Column:7	Row:19 Column:8	Row:19 Column:9	Row:19 Column:10	Row:19 Column:11	Row:19 Column:12	Row:19 Column:13	Row:19 Column:14	Row:19 Column:15	Row:19 Column:16	Row:19 Column:17	Row:19 Column:18
18	Row:18 Column:1	Row:18 Column:2	Row:18 Column:3	Row:18 Column:4	Row:18 Column:5	Row:18 Column:6	Row:18 Column:7	Row:18 Column:8	Row:18 Column:9	Row:18 Column:10	Row:18 Column:11	Row:18 Column:12	Row:18 Column:13	Row:18 Column:14	Row:18 Column:15	Row:18 Column:16	Row:18 Column:17	Row:18 Column:18
17	Row:17 Column:1	Row:17 Column:2	Row:17 Column:3	Row:17 Column:4	Row:17 Column:5	Row:17 Column:6	Row:17 Column:7	Row:17 Column:8	Row:17 Column:9	Row:17 Column:10	Row:17 Column:11	Row:17 Column:12	Row:17 Column:13	Row:17 Column:14	Row:17 Column:15	Row:17 Column:16	Row:17 Column:17	Row:17 Column:18
16	Row:16 Column:1	Row:16 Column:2	Row:16 Column:3	Row:16 Column:4	Row:16 Column:5	Row:16 Column:6	Row:16 Column:7	Row:16 Column:8	Row:16 Column:9	Row:16 Column:10	Row:16 Column:11	Row:16 Column:12	Row:16 Column:13	Row:16 Column:14	Row:16 Column:15	Row:16 Column:16	Row:16 Column:17	Row:16 Column:18
15	Row:15 Column:1	Row:15 Column:2	Row:15 Column:3	Row:15 Column:4	Row:15 Column:5	Row:15 Column:6	Row:15 Column:7	Row:15 Column:8	Row:15 Column:9	Row:15 Column:10	Row:15 Column:11	Row:15 Column:12	Row:15 Column:13	Row:15 Column:14	Row:15 Column:15	Row:15 Column:16	Row:15 Column:17	Row:15 Column:18





# Field Trial – Findable

## PLOT DETAILS

x

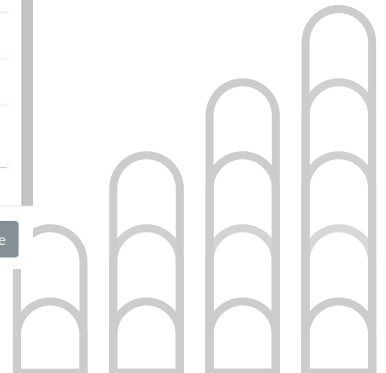
Row: 20  
 Column: 1  
 Length: 3.594m  
 Width: 1.8m  
 Study Design:  
 Sowing Date: 2019-10-30  
 Harvest Date: 2020-08-10  
 Treatment:  
 Comment: Slight height segregation



Replicate	Rack	Accession	Pedigree	Gene Bank	Links
1 (Current Plot)	1	DFW SEL 0208		<a href="#">Germplasm Resources Unit</a>	
3 (Plot Row:3 - Col:23)	1	DFW SEL 0208		<a href="#">Germplasm Resources Unit</a>	
2 (Plot Row:14 - Col:15)	1	DFW SEL 0208		<a href="#">Germplasm Resources Unit</a>	

## PHENOTYPES

Close

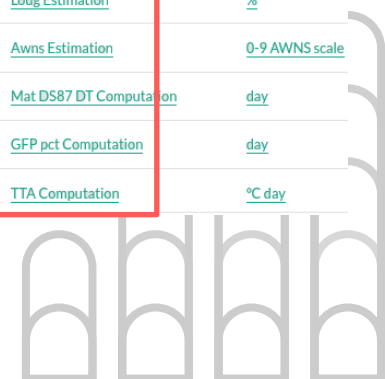


# Field Trial – Reusable Plot Phenotypes

Phenotypic traits, measurements & units are all ontological terms

## PHENOTYPES

Replicate	Rack	Date	Raw Value	Corrected Value	Trait	Measurement	Unit
1 (Current Plot)	1		2020-07-31		<a href="#">Maturity time</a>	<a href="#">Mat DS87 DT Computation</a>	<a href="#">day</a>
1 (Current Plot)	1		57		<a href="#">Grain filling period</a>	<a href="#">GFP pct Computation</a>	<a href="#">day</a>
1 (Current Plot)	1		1743.2		<a href="#">Anthesis thermal time</a>	<a href="#">TTA Computation</a>	<a href="#">°C day</a>
1 (Current Plot)	1		2672.05		<a href="#">Physiological maturity thermal time</a>	<a href="#">TTM Computation</a>	<a href="#">°C day</a>
1 (Current Plot)	1	2020-07-22	78.25		<a href="#">Plant height</a>	<a href="#">PH Measurement</a>	<a href="#">cm</a>
1 (Current Plot)	1		6.858225978		<a href="#">Grain yield</a>	<a href="#">GY Computation</a>	<a href="#">t/ha</a>
1 (Current Plot)	1	2020-08-05	0		<a href="#">Lodging incidence</a>	<a href="#">Lodg Estimation</a>	<a href="#">%</a>
1 (Current Plot)	1	2020-07-06	0		<a href="#">Spike awniness</a>	<a href="#">Awns Estimation</a>	<a href="#">0-9 AWNS scale</a>
3 (Plot Row:3 - Col:23)	1		2020-08-04		<a href="#">Maturity time</a>	<a href="#">Mat DS87 DT Computation</a>	<a href="#">day</a>
3 (Plot Row:3 - Col:23)	1		59		<a href="#">Grain filling period</a>	<a href="#">GFP pct Computation</a>	<a href="#">day</a>
3 (Plot Row:3 - Col:23)	1		1765.65		<a href="#">Anthesis thermal time</a>	<a href="#">TTA Computation</a>	<a href="#">°C day</a>

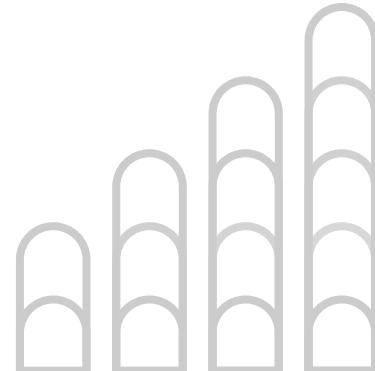




# Field Trials Grassroots Service

The service also integrates with other services and data repositories

- Any materials with publicly-available accessions automatically generates links to any data on the DFW data portal relevant to that accession
- The materials also automatically generate links to the Grassroots GRU service to view the seed details



# DFW Data Portal

Grassroots Data Repository

SEARCH LOGIN











## Using 12Mb capture to profile the Watkins bread wheat landrace diversity collection

*Anthony Hall & Laura-Jayne Gardiner*

Wheat has been domesticated into a large number of agricultural environments and has a remarkable ability to adapt to diverse environments. To understand this process, we survey genotype and DNA methylation across the Watkins bread wheat landrace collection, representing global genetic diversity. For each accession, we use gene based sequence capture (12Mb) to focus on the functionally relevant portion of the genome, followed by paired end sequencing on the Hiseq4000. All Watkins accessions therefore have genotypic data available plus sequence data generated after bisulfite treatment to allow methylation calls.

This data is made available under the [Toronto Agreement](#)

Location: [Home](#) > [under\\_license](#) > [toronto](#) > [Gardiner\\_2018-01-29\\_Watkins-diversity-12Mb](#)

Name	Size	Date	Properties
 <a href="#">Watkins BS exome capture data/</a>		2018-02-20 18:18	  
 <a href="#">Watkins exome capture data/</a>		2018-02-20 18:18	  

Brought to you by mod\_eirods-dav Copyright (c) 2016-18, Earlham Institute and Utrecht University.  
Images taken from the [Amiga Image Storage System](#) by Martin Merz.

[https://opendata.earlham.ac.uk/wheat/under\\_license/toronto/](https://opendata.earlham.ac.uk/wheat/under_license/toronto/)



# DFW Data Portal datasets

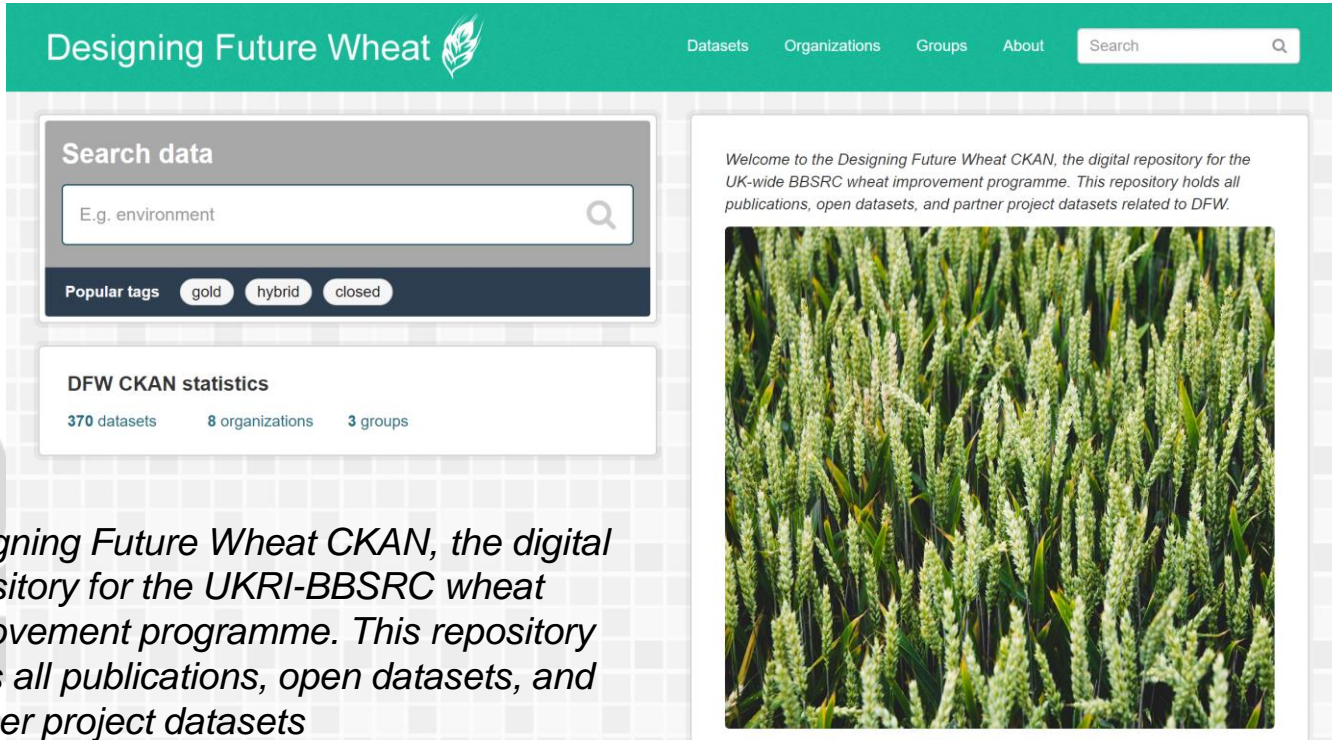
## 24 TB data hosted

- 116M ./Ball\_2018-06-13\_Breeder\_toolkit
- 1.6G ./Gardiner\_2018-07-04\_Wheat-gene-promoter-capture
- 4.7M ./Griffiths\_2018-10-22\_DFW-TK-NIL-Set2-2018-phenotypes
- 13T ./King\_2018-05-09\_Cadenza\_RefSeqV1.0\_Alignments
- 22G ./Clavijo\_2017-04-25\_Wheat-LoLa
- 279G ./Zhou\_2019\_RobxCla\_UAV\_Image\_Data
- 14G ./Martin\_etal\_2018\_Alabdullah\_etal\_2019\_wheat\_meiosis\_transcriptome\_and\_co-expression\_network
- 5.7G ./White\_2020-02-06\_consensus\_evolved\_fielder\_wheat
- 4.0G ./Riche\_2020-07-13\_DFW\_Academic\_Toolkit\_Harvest\_2020
- 6.4T ./Wulff\_2018-01-31\_OWWC
- 1.5T ./Gardiner\_2018-01-29\_Watkins-diversity-12Mb
- 262G ./Brinton\_etal\_2020-05-20-Haplotypes-for-wheat-breeding
- 3.4T ./Wulff\_2019-03-13\_OWWC\_lineage1
- 44G ./Gardiner\_et\_al\_2017\_Watkins\_exome\_capture\_SNP\_calls\_VCF
- 13G ./Ramirez-Gonzalez\_etal\_2018-06025-Transcriptome-Landscape


Since July 2020 (new server) 418 unique visitors and over 23 TB of transferred data  
4673 users in total from Apr 2018



# DFW Research Outputs - CKAN



The screenshot shows the CKAN website for 'Designing Future Wheat'. The header is green with the site name and a wheat leaf icon. Navigation links for 'Datasets', 'Organizations', 'Groups', and 'About' are present, along with a search bar. A search data box contains the text 'E.g. environment' and a search icon. Below it, 'Popular tags' include 'gold', 'hybrid', and 'closed'. A statistics box shows '370 datasets', '8 organizations', and '3 groups'. A main content area features a welcome message and a photograph of a wheat field.

Designing Future Wheat 

Datasets Organizations Groups About Search


**Search data**

Popular tags gold hybrid closed

**DFW CKAN statistics**

370 datasets 8 organizations 3 groups

Welcome to the Designing Future Wheat CKAN, the digital repository for the UK-wide BBSRC wheat improvement programme. This repository holds all publications, open datasets, and partner project datasets related to DFW.



*Designing Future Wheat CKAN, the digital repository for the UKRI-BBSRC wheat improvement programme. This repository holds all publications, open datasets, and partner project datasets*

# DFW Research Outputs in CKAN

**Organizations**

- Rothamsted Research 216
- John Innes Centre 87
- Earlham Institute 23
- National Institute ... 18
- University of Bristol 13
- University of Notti... 11
- European Bioinforma... 2
- Quadram Institute B... 1

**Groups**

- DFW Publications 360
- Open Data 6
- Open Wild Wheat Con... 2

**Tags**

- gold 121
- hybrid 111
- closed 54
- green 44
- bronze 30
- book chapter 14
- grassroots 6
- toronto 4
- Toronto 4
- diversity 2

Q

**370 datasets found** Order by: Relevance ▾

---

**Hidden variation in polyploid wheat drives local adaptation**

Wheat has been domesticated into a large number of agricultural environments and has the ability to adapt to diverse environments. To understand this process, we survey...

[PDF](#)

---

**Uncovering hidden variation in polyploid wheat**

Comprehensive reverse genetic resources, which have been key to understanding gene function in diploid model organisms, are missing in many polyploid crops. Young polyploid...

[PDF](#) [GitHub](#) [HTML](#)

---

**LYS3 encodes a prolamin-box-binding transcription factor that controls embryo...**

Mutations at the LYS3 locus in barley have multiple effects on grain development, including an increase in embryo size and a decrease in endosperm starch content. The gene...

[PDF](#) [XML](#) [API](#)

---

**Transfer of a starch phenotype from wild wheat to bread wheat by deletion of ...**

Our previous genetic analysis of a tetraploid wild wheat species, *Aegilops peregrina*, predicted that a single gene per haploid genome, *Bgc-1*, controls B-type starch granule...

[PDF](#)

---

**A carbohydrate-binding protein, FLOURY ENDOSPERM 6 influences the initiation ...**

Previously, we identified a quantitative trait locus on the group 4 chromosomes of *Aegilops* and bread wheat that controls B-type starch-granule content. Here, we identify a...

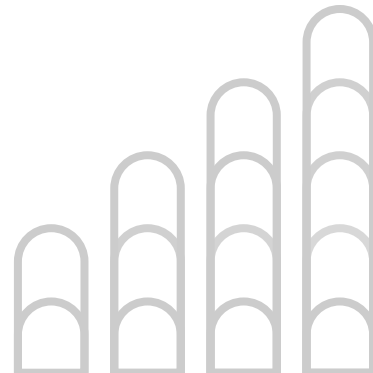
[PDF](#) [DOCX](#)

---

**Acrylamide in food: Progress in and prospects for genetic and agronomic solut...**

Acrylamide is a processing contaminant and Group 2a carcinogen that was discovered in foodstuffs in 2002. Its presence in a range of popular foods has become one of the most...

[PDF](#) [XML](#)




# CKAN Example entry

**A carbohydrate-binding protein, FLOURY ENDOSPERM 6 influences the initiation of A- and B-type starch granules in wheat**

Followers  
**0**

**Organization**



**National Institute of Agricultural Botany**  
NIAB datasets read more

**Social**

Twitter

Facebook

**License**

Creative Commons Attribution  
[View Data](#)

Dataset
Groups
Activity Stream

## A carbohydrate-binding protein, FLOURY ENDOSPERM 6 influences the initiation of A- and B-type starch granules in wheat

Previously, we identified a quantitative trait locus on the group 4 chromosomes of *Aegilops* and bread wheat that controls B-type starch-granule content. Here, we identify a candidate gene by fine-mapping in *Aegilops* and confirm its function using wheat TILLING mutants. This gene is orthologous to the FLOURY ENDOSPERM 6 (FLO6) gene of rice and barley and the PTST2 gene of *Arabidopsis*. In *Triticaceae* endosperm, reduction in the gene dose of functional FLO6 alleles results in reduction, or loss, of B-granules. This is due to repression of granule initiation in late-grain development, but has no deleterious impact on the synthesis of A-granules. The complete absence of functional FLO6, however, results in reduced numbers of normal A-type and B-type granules and the production of highly-abnormal granules that vary in size and shape. This polymorphous starch seen in a wheat flo6 triple mutant is similar to that observed in the barley mutant *Franubot*. Analysis of *Franubot* (fractured Nubel) starch suggests that the mutant A-granules are not fractured but compound, due to stimulation of granule initiation in plastids during early-grain development. Thus, in different situations in *Triticaceae*, FLO6 either stimulates or represses granule initiation.

**Data and Resources**

**Open Access PDF**

**Supplementary Material**  
Supplementary Figures and Tables

**Sequence Read Data**  
BioSample record deposited at the EMBL-EBI European Nucleotide Archive (ENA)

**MK848198 FASTA Sequence**  
MK848198.1 *Aegilops sharonensis* x *Aegilops umbellulata* chromosome 4U floury...

**MK848199 FASTA Sequence**  
MK848199.1 *Aegilops sharonensis* x *Aegilops umbellulata* chromosome 4S floury...

**MK848200 FASTA Sequence**  
MK848200.1 *Aegilops peregrina* chromosome 4U floury endosperm 6 (FLO6) gene...

**MK848201 FASTA Sequence**  
MK848201.1 *Aegilops peregrina* chromosome 4S floury endosperm 6 (FLO6) gene...

[Explore](#)

[Explore](#)

[Explore](#)

[Explore](#)

[Explore](#)

[Explore](#)

[Explore](#)

## Additional Info

Field	Value
Source	
Version	
Authors	<ul style="list-style-type: none"> <li>• <b>Name:</b> Chia, Tansy, <b>Type:</b> Corresponding Author,</li> <li>• <b>Name:</b> Chirico, Marcella, <b>Type:</b> Author,</li> <li>• <b>Name:</b> King, Rob, <b>Type:</b> Author,</li> <li>• <b>Name:</b> Ramirez-Gonzalez, Ricardo, <b>Type:</b> Author,</li> <li>• <b>Name:</b> Saccomanno, Benedetta, <b>Type:</b> Author,</li> <li>• <b>Name:</b> Seung, David, <b>Type:</b> Author,</li> <li>• <b>Name:</b> Simmonds, James, <b>Type:</b> Author,</li> <li>• <b>Name:</b> Trick, Martin, <b>Type:</b> Author,</li> <li>• <b>Name:</b> Uauy, Cristobal, <b>Type:</b> Author,</li> <li>• <b>Name:</b> Verhoeven, Tamara, <b>Type:</b> Author,</li> <li>• <b>Name:</b> Trafford, Kay, <b>Type:</b> Author,</li> </ul>
Maintainer	
Maintainer Email	
Article Host Type	publisher
Article Is Open Access	true
Article License Type	cc-by-nc-nd
Article Version Type	publishedVersion
Citation Report	<a href="https://scite.ai/reports/10.1101/643759">https://scite.ai/reports/10.1101/643759</a>
DFW Organisation	NIAB
DFW Work Package	2
DOI	10.1101/643759
Date Last Updated	2019-07-09T05:10:07.396155
Evidence	open (via page says license)
Journal Is Open Access	false
Open Access Status	hybrid
PDF URL	<a href="https://www.biorxiv.org/content/biorxiv/early/2019/05/23/643759.full.pdf">https://www.biorxiv.org/content/biorxiv/early/2019/05/23/643759.full.pdf</a>
Publisher URL	<a href="https://doi.org/10.1101/643759">https://doi.org/10.1101/643759</a>

# Interoperability – Software and Data



### Grassroots Tools

#### BROWSE FIELD TRIALS

MyLocation



Grassroots Data Repository

Using 128M capture to profile the Watkins bread wheat diversity collection

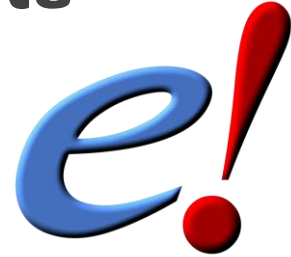
Grassroots Data Repository

Name	Size	Date	Properties
Watkins_02_watkins_capture_128M		2019-02-29 18:18	
Watkins_02_watkins_capture_128M		2019-02-29 18:18	

Designing Future Wheat

Search data

DFW CAN statistics



### SeedStor

#### SeedStor Homepage

Created by the Germplasm Resources Unit

Search Interface

Help Pages

Quick search Accession Name, Number and StoreCode fields:

Quick Search

Germplasm Resources Unit

### knetminer

dormancy OR germination OR color OR Flavon\* OR proanthocyanidin

Query Suggester

Genome or QTL Search


Gene List

- Tracc2.1288100017988
- Tracc2.1288100017989
- Tracc2.1288100017990
- Tracc2.1288100017991
- Tracc2.1288100017992
- Tracc2.1288100017993
- Tracc2.1288100017994
- Tracc2.1288100017995


ROTHAMSTED RESEARCH


# Germplasm Resources Unit

[www.seedstor.ac.uk](http://www.seedstor.ac.uk)

SeedStor (Version 1.08)   SeedStor Home   Search Interface   Help Pages   GRU Information -   

# SeedStor



A photograph of a display case containing several potted wheat plants of different varieties, with informational cards and posters on the wall behind them.

A display of the history of wheat cultivation.

## Germplasm Resources Unit

## SeedStor Homepage

*Created by the Germplasm Resources Unit*



SeedStor is an information management system created by the team at the Germplasm Resources Unit at the John Innes Centre, assisted by funding from the BBSRC. Written largely in PHP it provides a web-based modular interface to an underlying MySQL database. The system is being used to track, curate and monitor the use of the various accessions held in seed collections at the GRU Seed Bank.

[Search Interface](#)

[Help Pages](#)

*Quick search Accession Name, Number and StoreCode fields:*

Quick Search

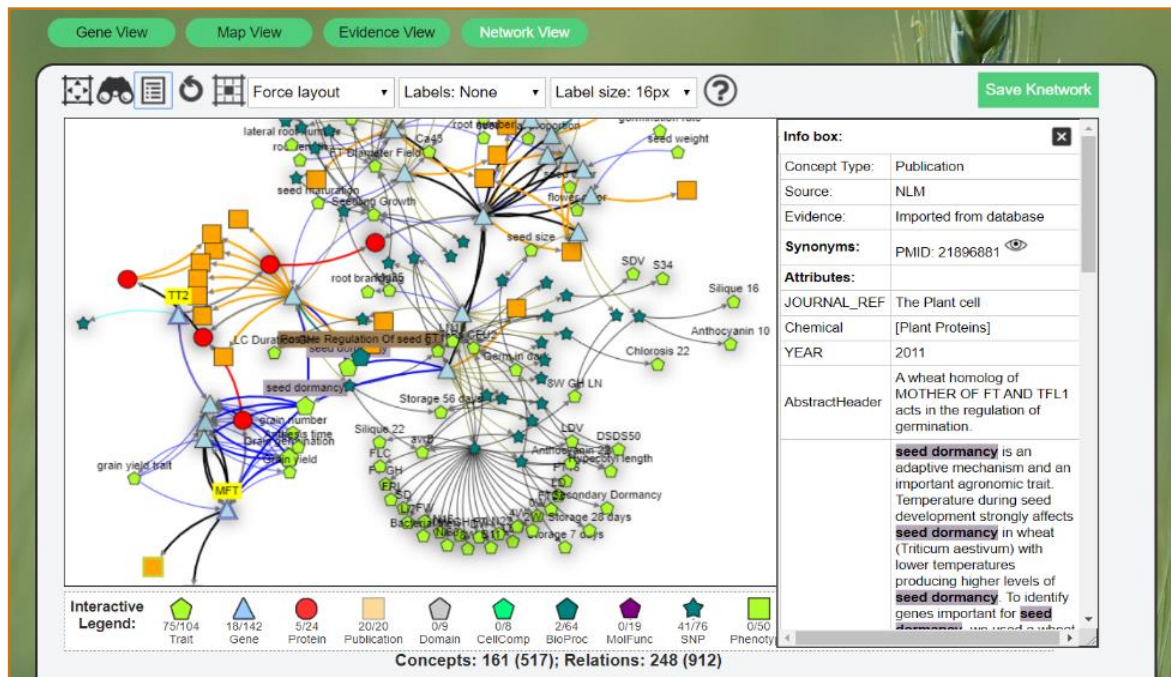
[Go](#)

# KnetMiner 4.0 released with user workspaces for gene networks

by Keywan Hassani-Pak

NEWS

<https://knetminer.com>



## A roadmap for gene functional characterisation in crops with large genomes: Lessons from polyploid wheat

Featured in eLife

<https://elifesciences.org/articles/55646>





**We need your help!** If you've ever found our data helpful, please take our impact survey (15 min). Your replies will help keep the data flowing to the scientific community.
 Take the survey

**Search**

Search ... Go

e.g. [TraesCS3D02G273600](#) or [3D:2585940-2634711](#) or [Carboxy\\*](#)

For information about the assembly and annotation please view the [IWGSC announcement](#).

The previous wheat assembly ([TGACv1](#)) and every other plant from release 37 is available in the [Ensembl Plants archive](#) site.


**About *Triticum aestivum***

*Triticum aestivum* (bread wheat) is a major global cereal grain essential to human nutrition. Wheat was one of the first cereals to be domesticated, originating in the [fertile crescent](#) around 7000 years ago. Bread wheat is hexaploid, with a genome size estimated at ~17 Gb, composed of three closely-related and independently maintained genomes that are the result of a series of naturally occurring hybridisation events. The ancestral progenitor genomes are considered to be [Triticum urartu](#) (the A-genome donor) and an unknown grass thought to be related to *Aegilops speltoides* (the B-genome donor). This first hybridisation event produced tetraploid [emmer wheat](#) (*AABB*, *T. dicoccoides*) which hybridized again with [Aegilops tauschii](#) (the D-genome donor) to produce modern bread wheat.

Guidelines for gene nomenclature in wheat can be found in the 2013 edition of the Wheat Gene Catalogue available in [GrainGenes](#). The Wheat Gene Catalogue is the internationally agreed rules of nomenclature for wheat genes.

Taxonomy ID [4565](#)

Data source [International Wheat Genome Sequencing Consortium](#)

 [More information and statistics](#)

**Genome assembly:** [IWGSC](#)

-  [More information and statistics](#)
-  [Download DNA sequence \(FASTA\)](#)
-  [Convert your data to IWGSC coordinates](#)
-  [Display your data in Ensembl Plants](#)

Other cultivars






[View karyotype](#)




[Example region](#)


**Gene annotation**

**What can I find?** Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

-  [More about this genebuild](#)
-  [Download genes, cDNAs, ncRNA, proteins - FASTA - GFF3](#)
-  [Update your old Ensembl IDs](#)

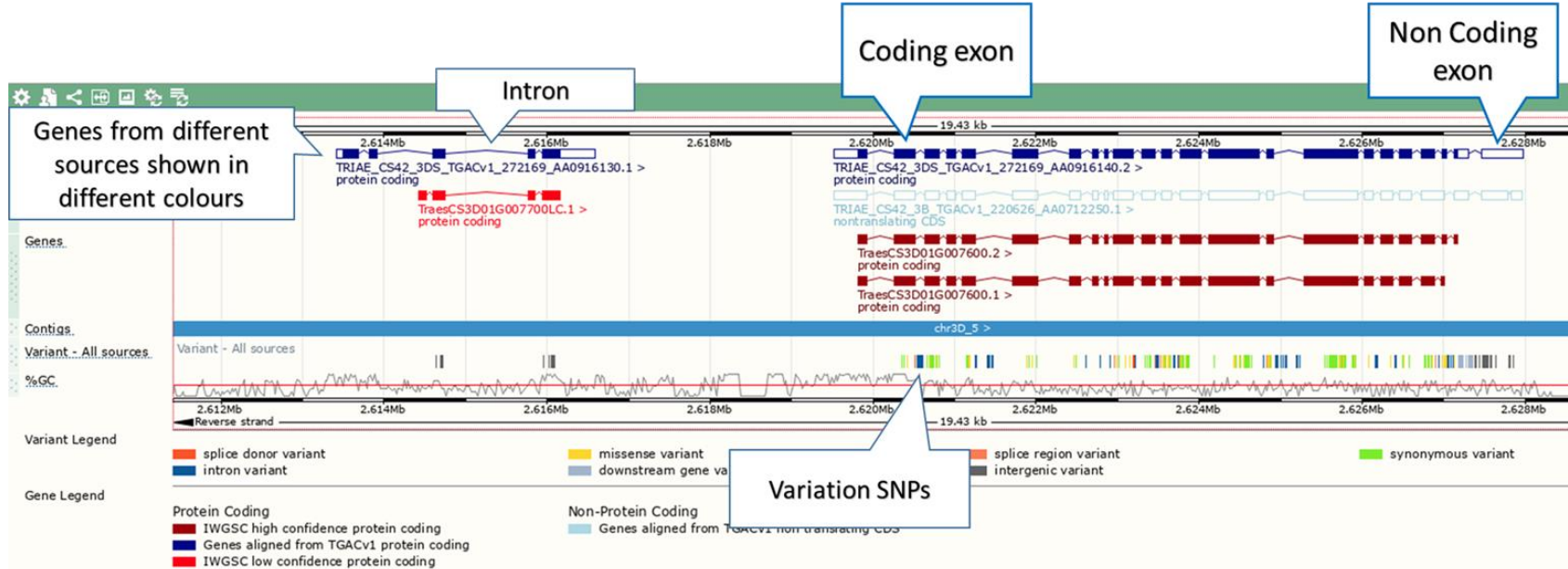


**Example gene**



**Example transcript**

# The IWGSC RefSeq v1.0 in Ensembl Plants



# The TILLING population in Ensembl Plants

## Cadenza0225.chr3D.379537010 SNP

Most severe consequence

**missense variant** | [See all predicted consequences](#)

Alleles

**G/A** | Highest population MAF: < 0.01

Location

[Chromosome 3D:379537010](#) (forward strand) | VCF: 3D 379537010 Cadenza0225.chr3D.379537010 G A

HGVS names

This variant has 5 HGVS names - [Show](#)

Original source

[EMS-induced mutations](#) from sequenced TILLING populations. Seeds can be ordered from [UK SeedStor](#) or [US Dubcovsky lab](#) .

About this variant

This variant overlaps [2 transcripts](#).

### Gene and Transcript consequences

Show/hide columns										Filter	
Gene	Transcript (strand)	Allele (transcript allele)	Consequence Type	Position in transcript	Position in CDS	Position in protein	Amino acid	Codons	SIFT	Detail	
<a href="#">TraesCS3D01G273600</a>	<a href="#">TraesCS3D01G273600.1</a> (-) biotype: protein_coding	A (T)	<b>missense variant</b>	2171 (out of 2742)	2171 (out of 2742)	724 (out of 913)	A/V	GCG/GTG	<b>0.01</b>	<a href="#">Show</a>	
<a href="#">TraesCS3D01G273600</a>	<a href="#">TraesCS3D01G273600.2</a> (-) biotype: protein_coding	A (T)	<b>missense variant</b>	2261 (out of 2688)	2261 (out of 2688)	754 (out of 895)	A/V	GCG/GTG	<b>deleterious</b>	<a href="#">how</a>	

No overlap with Ensembl Regulatory features

# Interoperability – Software and Data



**Grassroots Tools**  
BROWSE FIELD TRIALS

My Location



Grassroots Data Repository

Using 128M capture to profile the Watkins bread wheat landrace diversity collection

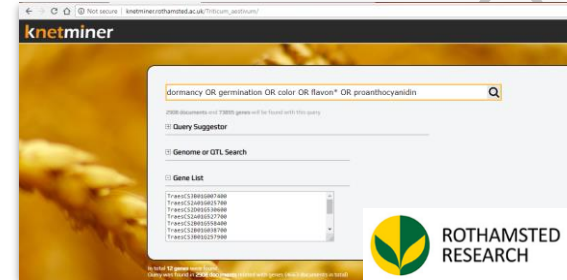
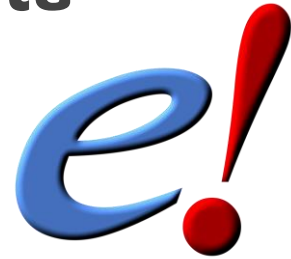
Grassroots Data Repository

Name	Size	Date	Properties
Watkins_22_2016_0428_03M		2016-04-29 18:18	
Watkins_22_2016_0428_03M		2016-04-29 18:18	

Designing Future Wheat

Search data

DFW CAN statistics



knetminer

dormancy OR germination OR color OR flavone\* OR proanthocyanidin

Query Suggester

Genome or GTL Search

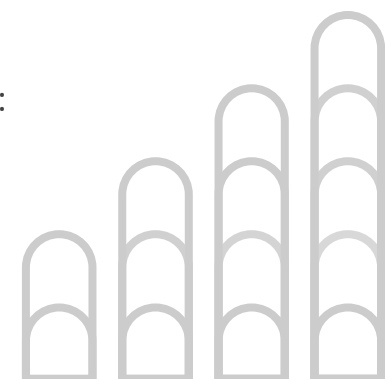
Gene List

- Trasac1.2389100017300
- Trasac1.2389100017300
- Trasac1.2389100017300
- Trasac1.2389100017300
- Trasac1.2389100017300
- Trasac1.2389100017300



# What we have learned

- That the technical challenges of developing a data sharing infrastructure are significant but not overwhelming
  - Would not have been practical if we had started from scratch
- Ensuring that the tools we develop are useable by the project members takes longer and more resource than predicted
  - Major contributions needed by the trial managers who are already overcommitted
- Reaching other data submitters can almost require 1:1 support from DCTF or others
- Developing a common culture of data sharing takes time and effort
- Being open with our progress and research outputs benefits relationships:
  - With funders – transparency
  - With community – new collaborations



# Acknowledgements

- University of Bristol
  - **Paul Wilkinson**
  - Mark Winfield
  - Keith Edwards
  - Gary Barker
- John Innes Centre
  - **Luzie Wingen**
  - Simon Griffiths
  - Ricardo Ramirez Gonzalez
  - Cristobal Uauy
  - Graham Moore
- Rothamsted Research
  - **Chris Rawlings**
  - Keywan Hassani-Pak
  - Marco Brandizi
- Earlham Institute
  - **Simon Tyrrell**
  - **Xingdong Bian**
  - **Alice Minotto**
  - Toni Etuk
  - Felix Shaw
  - Jon Wright
  - Paul Bailey
  - Bernardo Clavijo
  - Luis Yanes
  - Catherine Hunter
- EMBL-EBI
  - Guy Naamati
  - Dan Bolser
  - Paul Kersey (now at Kew)

