

Unveiling the Narratives of Complex Traits and Diseases

*Integrative Bioinformatics
enhanced with KGs*



HEAD OF BIOINFORMATICS @ROTHAMSTED
FOUNDER OF KNETMINER

Dr Keywan Hassani-Pak

Genes are rarely single actors

- **Unmasking Heroes and Villains:** Understanding genes in complex trait genomics is akin to solving a plant biology mystery, where multiple genes play diverse roles.
- **Plotlines of Interaction:** Just as in a thrilling story, their interactions form the plotlines that shape the narrative of the trait.
- **Data Challenges:** Increasing volumes of data, found in various formats and scattered across diverse locations



Our Vision

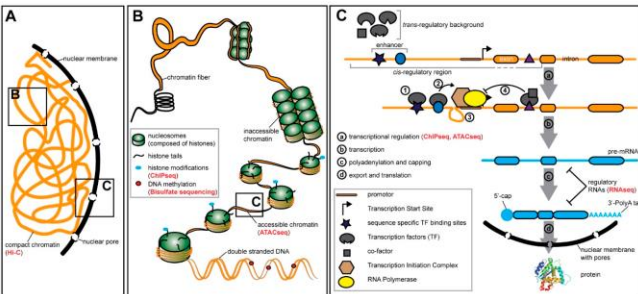
- **Automation & Discovery:** Develop advanced software for automated plotline discovery.
- **Integrating Evidence:** Seamlessly integrate diverse data sources.
- **Compelling Insights:** Present users with compelling, comprehensive explanations.
- **Empowering Science:** Empower researchers to unravel biology's mysteries.

What is KnetMiner?

- Knowledge Graph-based data exploration platform.
- Integrates biological and agricultural datasets.
- Enables interactive exploration and discovery.
- Supports decision-making in research and development.

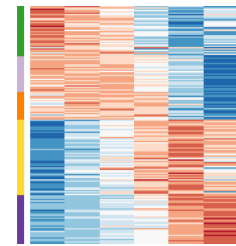


We organise knowledge about genes and traits



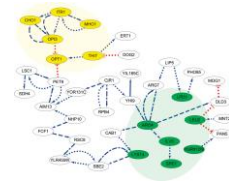
Genetics

- QTL
- GWAS



Omics

- Genomics
- Transcriptomics
- Proteomics
- Metabolomics



Networks

- PPI
- GRN
- Phylogeny



32 million scientific publications

10,000s of published genomes

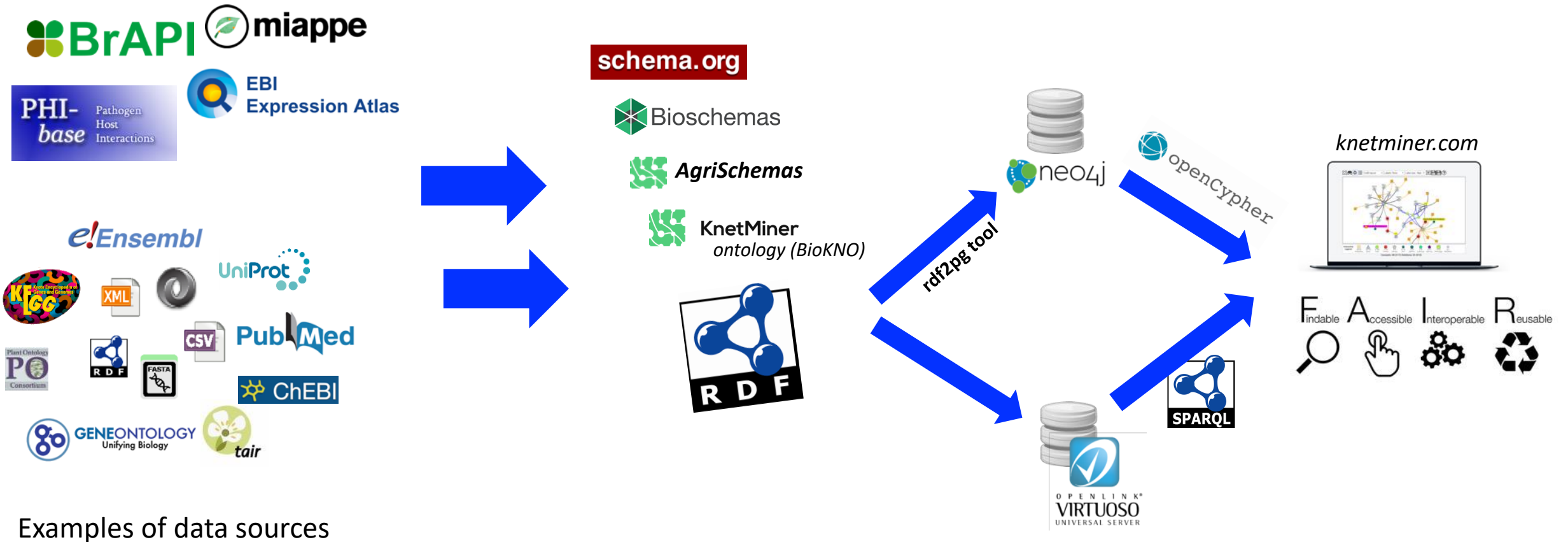
1,000s of databases & ontologies

Knowledge Graph Content

- KG-2023-Cereals
- 7 genomes
- 2M nodes
- 10M edges

	Genes	Pubmed	NER	Gene-Trait	Homology	...
Arabidopsis_thaliana	27628	81501	346692	66897	16811	
Brachypodium_distachyon	34310	1113	3365	664	26199	
Hordeum_vulgare	35825	11771	26204	4267	30485	
Oryza_indica	40745	39476	131154	20767	30180	
Oryza_sativa	35775	39476	145008	28233	27105	
Triticum_aestivum	107545	33320	480210	21193	100822	
Zea_mays	39756	35999	88139	14675	30613	

Lightweight Data Modelling and Integration



Brandizi et al., 2018
<https://knetminer.com/data>

Searching with keywords, genes and genome regions

The screenshot shows the KnetMiner search interface for *Triticum aestivum*. The page includes a navigation bar with the KnetMiner logo, the species name, and links for Tutorial, Cite Us, keywan, and My KnetSpace. The main search area is titled "Search KnetMiner with keywords, gene lists or genomic regions:" and contains three search sections: Keyword Search, Gene List Search, and Genome Region Search. The Keyword Search section has a text input field containing "drought OR heat OR frost" and a button with a plus sign. Below it, a message states "4495 documents and 66135 genes will be found with this query". The Gene List Search section has a text area containing a list of gene IDs: TRAESCS3D02G468400, TRAESCS3D03G0544800, TaSIZ1, CDC5, TPS*, and TPP?. Below the text area, it says "6 Genes" and "Clear all". The Genome Region Search section has four input fields: Chromosome Start (1A), End (580000000), Label (My QTL), and Genes (164). Below these fields, there is a "+ Add region" button and a "Search" button. At the bottom, there is a "Clear Search Fields" link. On the right side, there is a "Example Queries" panel with a list of search examples: Keyword search (AND), Keyword search (OR), Gene names search, Wheat gene symbols, Grain colour + PHS, Grain growth (Region Search), Wheat gene list (20 genes), Drought + Stress resistance (100 genes), and Disease resistance (1000s of genes).

KnetMiner *Triticum aestivum* Tutorial Cite Us keywan My KnetSpace

Search KnetMiner with keywords, gene lists or genomic regions:

Keyword Search

drought OR heat OR frost

4495 documents and 66135 genes will be found with this query

Gene List Search

TRAESCS3D02G468400
TRAESCS3D03G0544800
TaSIZ1
CDC5
TPS*
TPP?

6 Genes Clear all

Genome Region Search

Chromosome Start	End	Label	Genes
1A	580000000	My QTL	164

+ Add region

Search

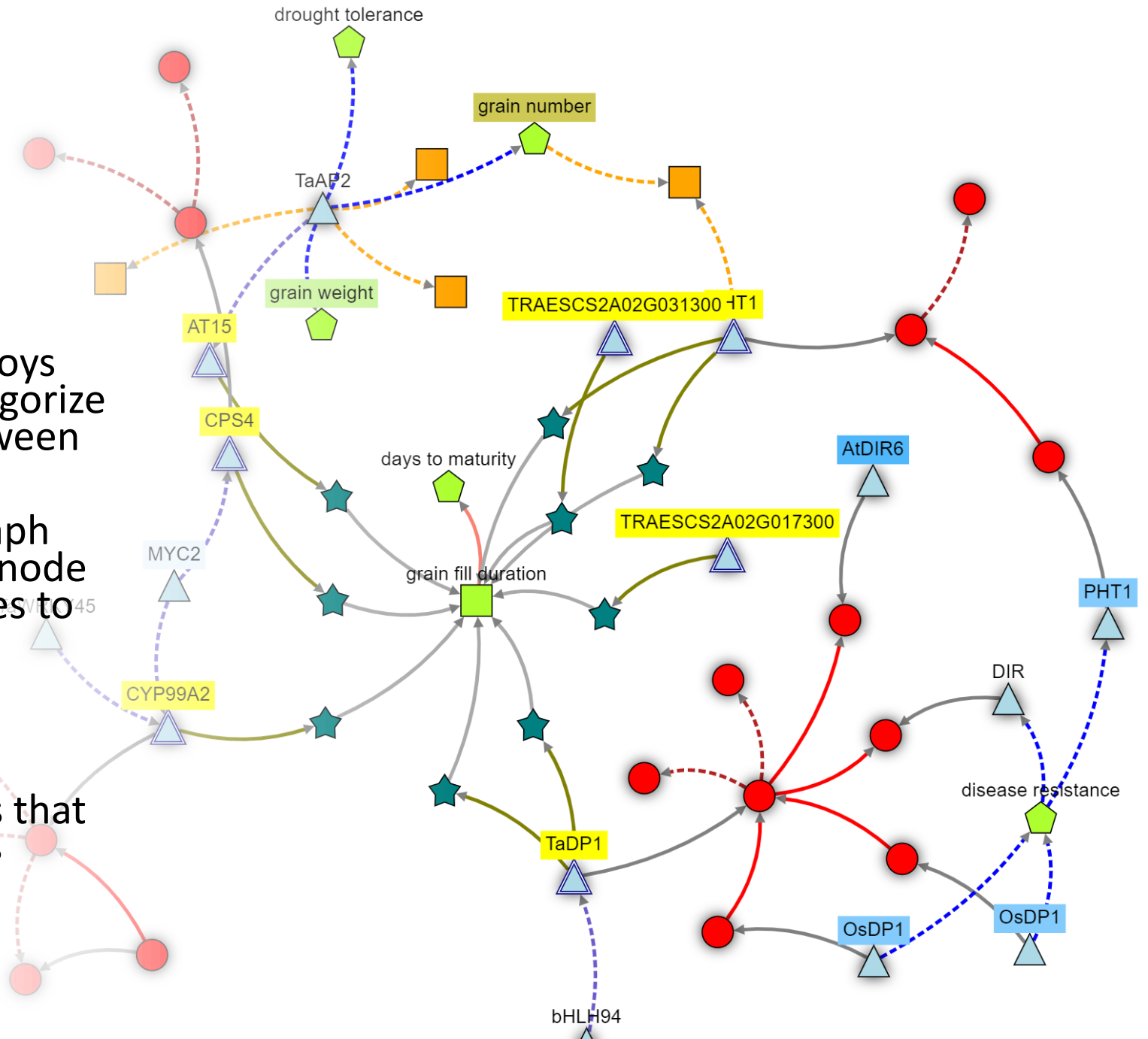
Clear Search Fields

Example Queries

- Keyword search (AND)
- Keyword search (OR)
- Gene names search
- Wheat gene symbols
- Grain colour + PHS
- Grain growth (Region Search)
- Wheat gene list (20 genes)
- Drought + Stress resistance (100 genes)
- Disease resistance (1000s of genes)

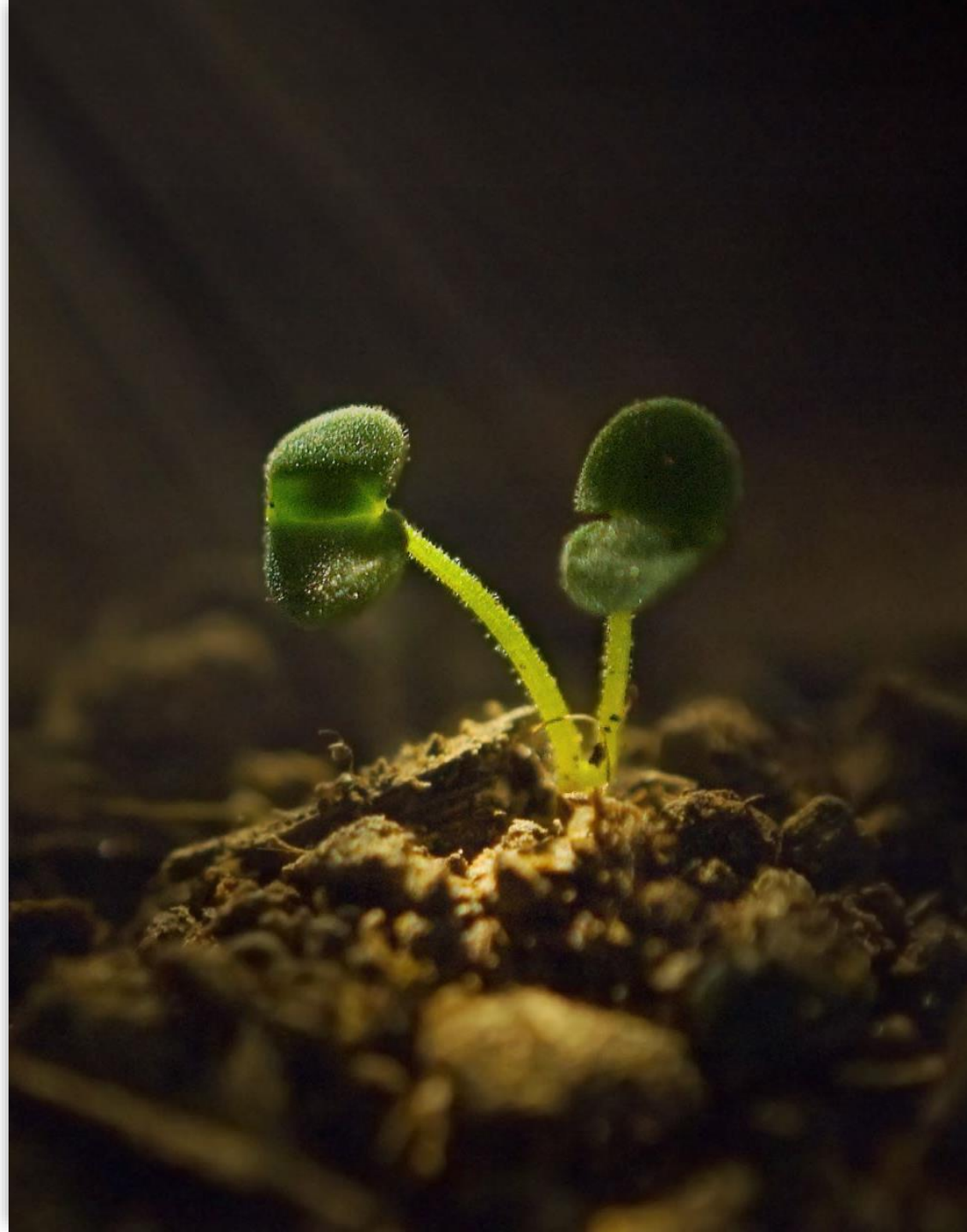
Semantic Search in KnetMiner

- **Structured Ontologies:** Employs structured ontologies to categorize and define relationships between biological concepts.
- **Graph analytics:** Employs graph analytics techniques such as node degree and distance measures to quantify the specificity of information
- **Relationship Discovery:** Uncovering meaningful connections and associations that span multiple linked datasets



Use Cases of KnetMiner in Agriculture

- Crop Improvement: Discovering genetic markers for traits.
- Disease Resistance: Understanding pathways related to plant immunity.
- Functional Annotation: Annotating gene functions and interactions.
- Comparative Analysis: Cross-species and cross-variety comparisons.



Case study

Forward genetics in crops and Arabidopsis

Phenotype: Anthocyanin 22



ID: 75

Study: [Atwell et. al, Nature 2010](#)

DOI: [10.21958/phenotype:75](https://doi.org/10.21958/phenotype:75)

Scoring: Results expressed as binary data, determined by the presence (1) or absence (0) of anthocyanin in all 4 plants / accession after 5wks of growth

Growth conditions: None

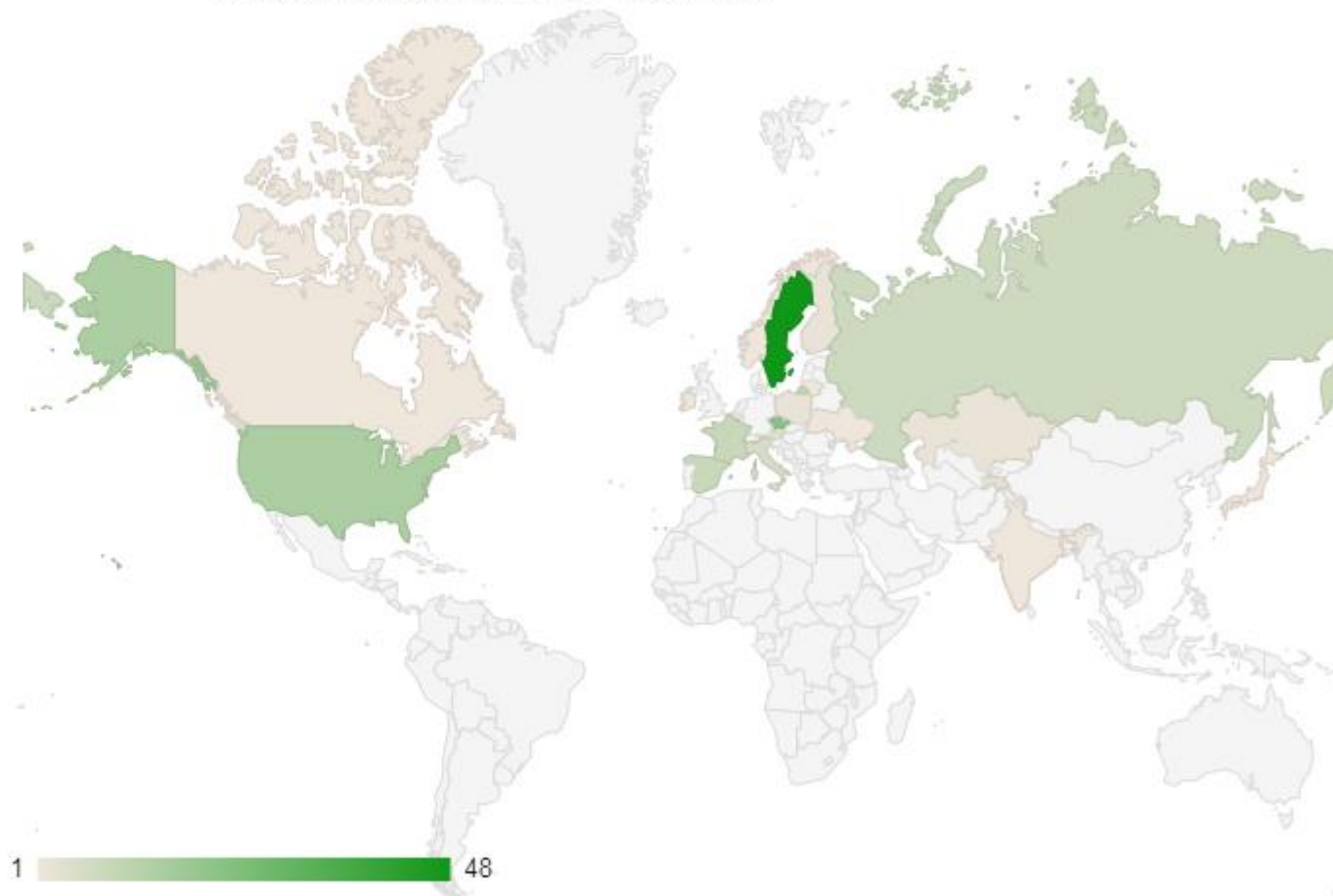
Trait-Ontology: [anthocyanin content \(TO:0000071\)](#)

Environment ontology: [growth chamber study \(EO:0007269\)](#)

Unit: [dimensionless unit \(UO:0000186\)](#)

Shapiro-Wilk*: 6.44e-20 (p-value)

Geographic distribution of 177 accessions



Search KnetMiner with keywords, gene lists or genomic regions:

Keyword Search

anthocyanin

672 documents and 7309 genes will be found with this query

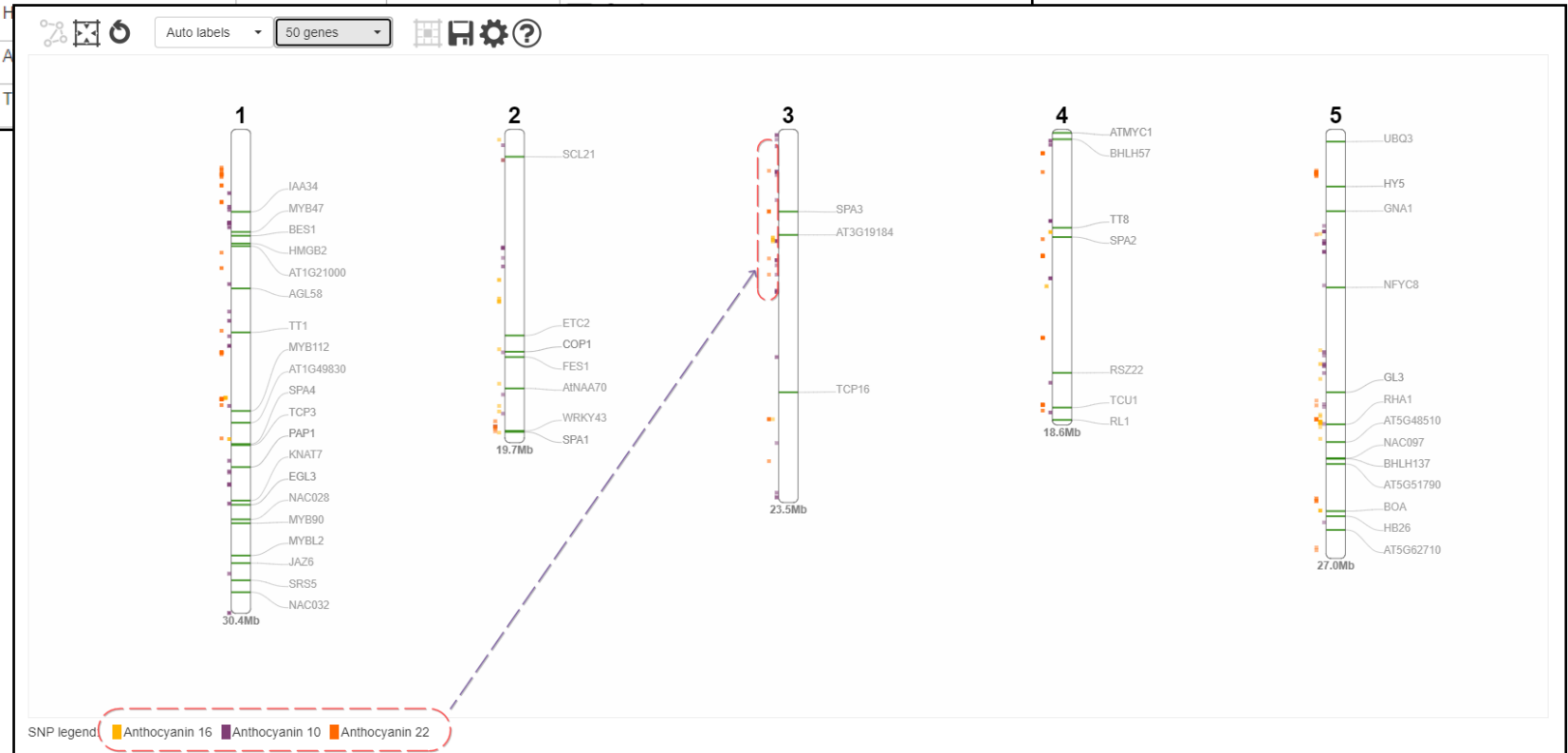
anthocyanin

- Anthocyanin 16
- Anthocyanin 10
- anthocyanin content
- Anthocyanin 22

Visualisation of GWAS data in KnetMiner

ACCESSION	GENE NAME	CHRO	START	EVIDENCE
AT2G32950	COP1	2	13977881	12 3 2 169 1
AT1G56650	PAP1	1	21233555	7 3 158 1 4
AT1G63650	EGL3	1	23599413	6 4 2 139 2
AT2G46340	SPA1	2	19022154	4 4 2 143 1
AT5G41315	GL3	5	16529090	6 4 2 124 1
AT4G09820	TT8	4	6181881	5 4 2 124 1
AT5G11260				
AT4G00480				
AT1G53230				

Trait
anthocyanin content
Anthocyanin 10



AraGWAS Catalog

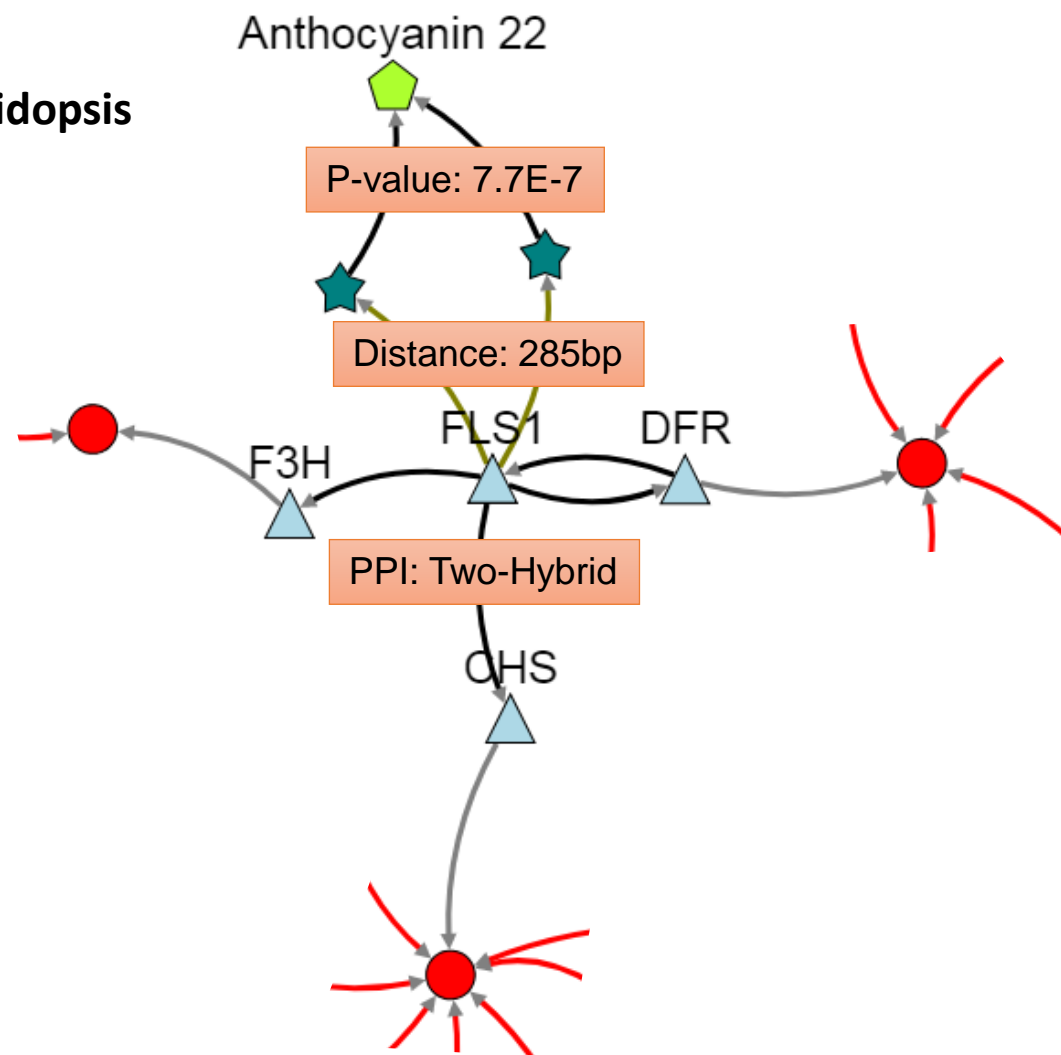
[Home](#) > [Studies](#) > Anthocyanin 22

[STUDY DETAILS](#)

Description

Name: Anthocyanin 22
DOI: [10.21958/gwas:75](https://doi.org/10.21958/gwas:75)
Phenotype ontology: anthocyanin content
Phenotype description: Results expressed as binary data, determined by the absence (0) of anthocyanin in all 4 plants / accessions in growth
Genotype: Full imputed genotype v1.0.0
Transformation: raw
Method: amm
AraPheno link: [Anthocyanin 22](#)
Original publication: [Atwell et. al, Nature 2010](#)
Number of samples: 177 (from 27 different countries)
Total associations: 4746020
Bonferroni threshold: 1.054e-8
Permutation threshold: 4.774e-8
N hits (Bonferroni): 0
N hits (permutation): 0

Arabidopsis



Case study

Reverse genetics in wheat

Trait-based analysis of differentially expressed genes

Search KnetMiner with keywords, gene lists or genomic regions:

Keyword Search

dormancy OR germination OR color OR flavon* OR proanthocyanidin

4441 documents and 78898 genes will be found with this query

Gene List Search

TraesCS3B02G007400
 TraesCS2A02G025700
 TraesCS2D02G530600
 TraesCS2A02G527700
 TraesCS2B02G558400
 TraesCS2B02G038700
 TraesCS3B02G257900

12 Genes Clear all

Genome Region Search

Search

Clear Search Fields

Example Queries

- Keyword search (AND)
- Keyword search (OR)
- Gene names search
- Wheat gene symbols
- Grain colour + PHS
- Grain growth (Region Search)
- Wheat gene list (20 genes)
- Drought + Stress resistance (100 genes)
- Disease resistance (1000s of genes)

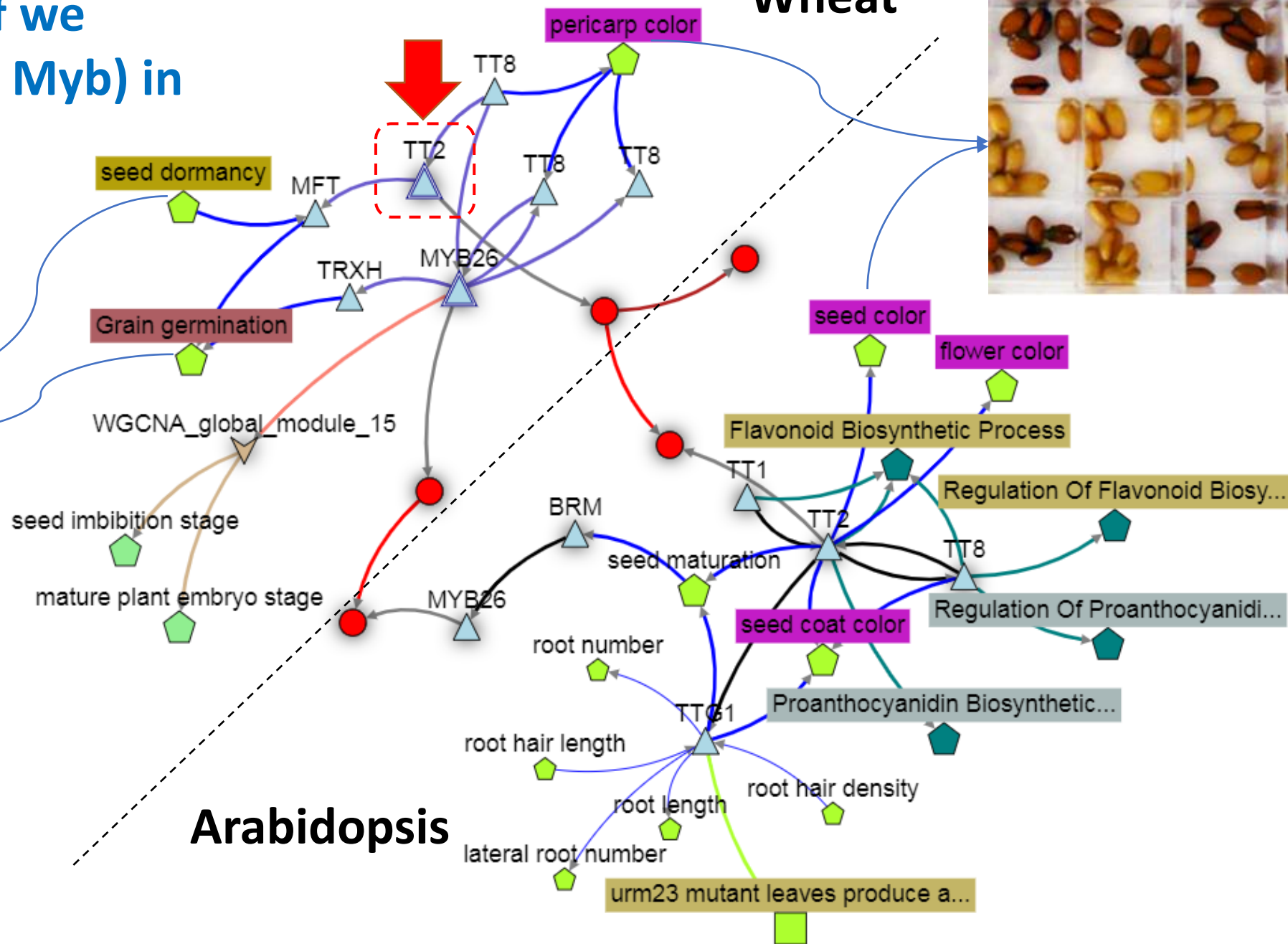
BioProc Publication Trait Path Phenotype PlantOntologyTerm Reaction MolFunc Protein

Accession	Symbol	Chr	Nt start	Evidence	KnetScore	Select
TRAESCS3D02G468400	TT2	3D	570801163	9 100 7	437.62	<input type="checkbox"/>
TRAESCS2B02G038700	CHS	2B	17881640	2 3 5 3 100 5 1	257.79	<input type="checkbox"/>
TRAESCS2A02G025700	CHS	2A	12101768	2 3 5 100 5 1	255.04	<input type="checkbox"/>
TRAESCS2A02G527700	CHS	2A	747086930	2 3 5 2 100 5 1	254.12	<input type="checkbox"/>
TRAESCS2D02G530600	CHS	2D	617051125	2 3 5 2 100 5 1	254.12	<input type="checkbox"/>
TRAESCS2B02G558400	CHS	2B	752784988	2 3 5 2 100 5 1	253.96	<input type="checkbox"/>
TRAESCS3D02G224600	DFR	3D	305876136	1 1 1 4 3 71 4	233.27	<input type="checkbox"/>
TRAESCS3B02G257900	DFR	3B	415923305	1 1 1 3 2 68 3	210.53	<input type="checkbox"/>
TRAESCS3A02G226600	DFR	3A	423967211	1 1 1 3 3 65 2	199.31	<input type="checkbox"/>

What happens if we knockout *TT2* (R Myb) in wheat?



Wheat



Arabidopsis

Thank you!



ROTHAMSTED
RESEARCH



<https://knetminer.com>