GrainGenes
A Database for Triticeae and Avena

GrainGenes
a global web resource for wheat, barley, rye, and oat data

Taner Sen
USDA-ARS, Crop Improvement and Genetics, Albany, CA
UC Berkeley

PAG 2024
January 13, 2024

GrainGenes

https://wheat.pw.usda.gov – Data, Tools, Community

Search

- Search & Browse GrainGenes
- Genetic Maps at GrainGenes

Submit Your Data to GrainGenes

- Submit Your Data Sets to GrainGenes
- GrainGenes Data Formats

Community Services

- Calendar
- Current Hot Topics
- Data Download
- GrainGenes Mailing List
- Job Listings
- Oatmail Mailing List
- Tutorials

Species Portals on GrainGenes

- Wheat Gene Catalogue
- Annual Wheat Newsletter
- Barley Boulevard
- Barley Genetics Newsletter
- Global Durum Genomic Resources
- Oat Newsletter
- Oat Nomenclature
- PanOat

Upcoming Events

- Genomic selection in breeding. A hands-on short course in R in Madrid Spain
  May 20 2024 to May 24 2024
- 2024 American Oat Workers in Saskatoon, Saskatchewan, Canada
  Jul 21 2024 to Jul 24 2024
- Eucarpia 22nd General Congress in Leipzig, Germany
  Aug 18 2024 to Aug 20 2024

Quick Links

Hot Topics

The Wheat Genome book is out

[Posted on November 20, 2023]

Dear IWGSC Members and Friends,

We are very pleased to announce that the first comprehensive book about the Wheat Genome is now out and available in open access.

You can read or download The Wheat Genome book here:
https://link.springer.com/book/10.1007/978-3-031-38294-9

US Federal Register notice: Movement of Organisms Modified or Produced Through Genetic Engineering

[Note: Five new rules as to the nature, size, and number of mutations in plants are being proposed to reduce regulatory burden. The public is asked to provide comments.]

AGENCY: Animal and Plant Health Inspection Service, USDA.

SUMMARY:

GrainGenes Updates

- December 2023: Added AlphaFold linked proteins to select organisms
- December 2023: Added Triticum urartu genome browser and BLAST
- November 2023: Aegilops umbellutata browser and BLAST are available
- October 2023: 1,000 wheat exome SNP data as a track on the IWGSC Chinese Spring RefSeq v2.1 genome browser
- October-November 2023: KASP primers added to GrainGenes Probes.
- October 2023: Added two Wheat einkorn browsers and BLAST
- September 2023: VMAP 1.1 Track is added to wheat Chinese Spring IWGSC RefSeq v1 genome browser
- September 2023: Genome browsers and BLAST service are available for Einkorn assemblies in a brand-new site
- September 2023: Phosphosite tracks added to IWGSC v2, Morex v3, Sang
- September 2023: A new track is added placing Oat-2018-Consensus (2018 Mrg) map regions on the OT3098 v2 genome assembly
- August 2023: 2022 African Spring Wheat Stem Rust Nursery data is available
- August 2023: GWAS results for pathology, salinity, and agronomic traits added to the IWGSC Chinese Spring v1 browser

more updates....

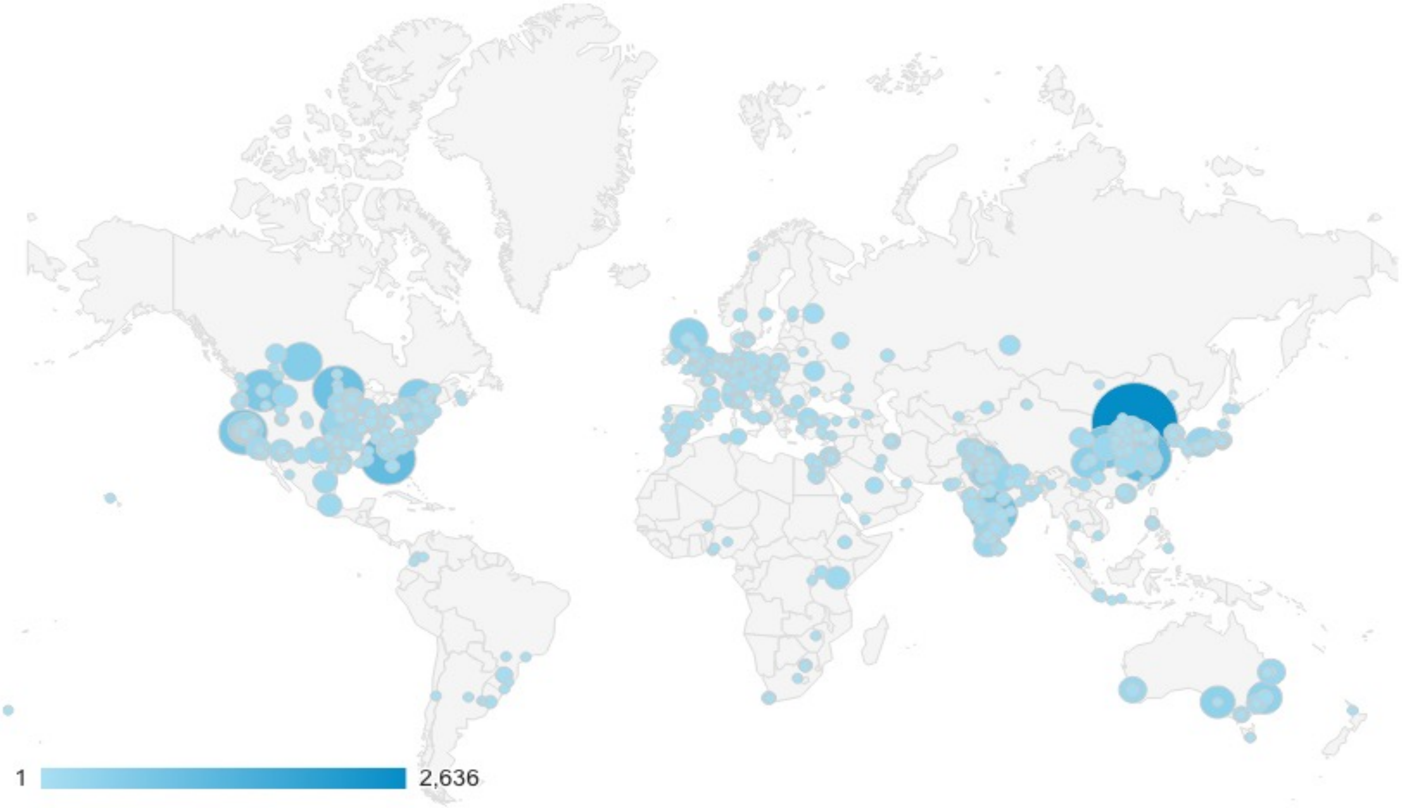
Follow Us

About GrainGenes

GrainGenes

- **Mission**: to serve researchers who work on *Triticeae* and *Avena* with their long-term data storage needs by curating, building and managing integrated views for a wide range of peer-reviewed data types, including genetic, genomic, and pathologic information for wheat, barley, rye, and oat. (since 1992)

- **Stakeholders**: Global small grains researchers



GrainGenes – last year

1. GrainGenes **5-year Project Plan** 2023-2028 was certified

2. **Annual site visitors** to GrainGenes are up around **17%** compared to the previous year – demonstrating the usefulness and impact of GrainGenes

3. We acquired a **new server** through end-of-the-year funds in FY and we are virtualizing the GrainGenes infrastructure

4. **100s of new QTLs** were curated into GrainGenes

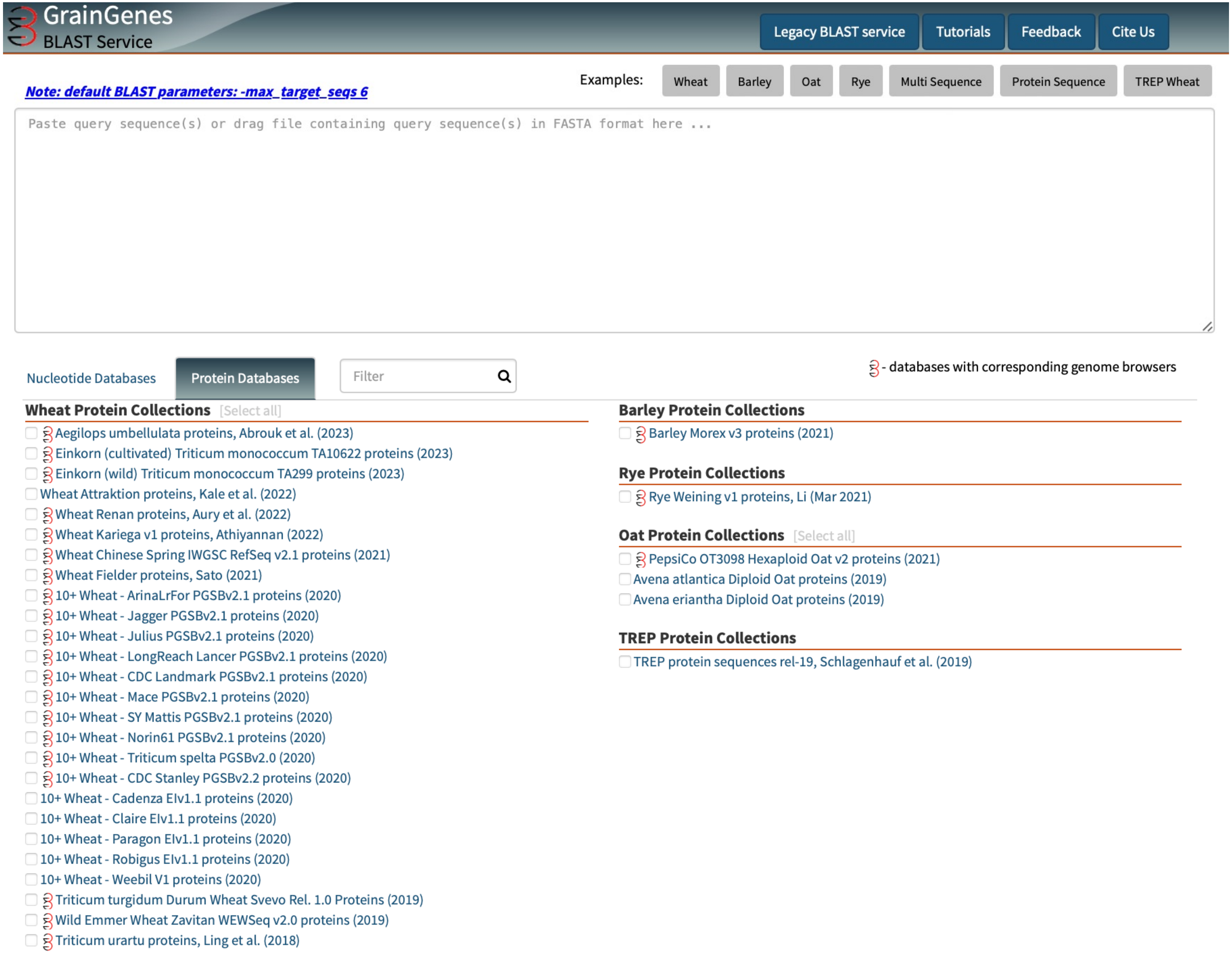
5. GrainGenes now has **176 BLAST databases, including 81 JBrowse-linked databases**

Genome Browsers

Wheat

- Aegilops umbellulata genome assembly; Abrouk et al. (2023)
- Wheat einkorn 2 assemblies; Ahmed et al. (2023)
- Wheat cultivar Sonmez genome assembly; Akpinar et al. (2022)
- Wheat cultivar Renan genome assembly; Aury et al. (2022)
- Wheat cultivar Attraktion genome assembly; Kale et al. (2022)
- Aegilops longisimma genome assembly; Avni et al. (2022)
- Aegilops speltoides genome assembly; Avni et al. (2022)
- Aegilops sharonensis genome assembly; Avni et al. (2022)
- Wheat Kariega genome assembly; Athiyannan et al. (2022)
- Wheat Fielder genome assembly; Sato et al. (2021)
- Aegilops tauschii Aet v5.0 genome assembly; Wang et al. (2021)
- Aegilops tauschii (wheat) 4 assemblies; Zhou et al. (2021)
- Wheat Chinese Spring IWGSC RefSeq v2.1 genome assembly (2021)
- 10+ Wheat Genome Project (2020)
- Triticum turgidum Durum Wheat Svevo (RefSeq Rel. 1.0) (2019)
- Wild Emmer Wheat Zavitan WEWSeq v2.0 genome assembly (2019)
- Wheat Chinese Spring IWGSC RefSeq v1.0 genome assembly (2018)
- Triticum urartu genome assembly (2018)
- Aegilops tauschii Aet v4.0 genome assembly (2017)
- Wild Emmer Wheat Zavitan WEWSeq v1.0 genome assembly (2017)
- Hexaploid Wheat Pangenome, Montenegro et al. (2017)

BLAST - Protein



Note: default BLAST parameters: -max_target_seqs 6

Examples: Wheat | Barley | Oat | Rye | Multi Sequence | Protein Sequence | TREP Wheat

`Paste query sequence(s) or drag file containing query sequence(s) in FASTA format here ...`

Nucleotide Databases | **Protein Databases** | Filter

§ - databases with corresponding genome browsers

Wheat Protein Collections [Select all]

- ☐ § Aegilops umbellulata proteins, Abrouk et al. (2023)
- ☐ § Einkorn (cultivated) Triticum monococcum TA10622 proteins (2023)
- ☐ § Einkorn (wild) Triticum monococcum TA299 proteins (2023)
- ☐ Wheat Attraktion proteins, Kale et al. (2022)
- ☐ § Wheat Renan proteins, Aury et al. (2022)
- ☐ § Wheat Kariega v1 proteins, Athiyannan (2022)
- ☐ § Wheat Chinese Spring IWGSC RefSeq v2.1 proteins (2021)
- ☐ § Wheat Fielder proteins, Sato (2021)
- ☐ § 10+ Wheat - ArinaLrFor PGSBv2.1 proteins (2020)
- ☐ § 10+ Wheat - Jagger PGSBv2.1 proteins (2020)
- ☐ § 10+ Wheat - Julius PGSBv2.1 proteins (2020)
- ☐ § 10+ Wheat - LongReach Lancer PGSBv2.1 proteins (2020)
- ☐ § 10+ Wheat - CDC Landmark PGSBv2.1 proteins (2020)
- ☐ § 10+ Wheat - Mace PGSBv2.1 proteins (2020)
- ☐ § 10+ Wheat - SY Mattis PGSBv2.1 proteins (2020)
- ☐ § 10+ Wheat - Norin61 PGSBv2.1 proteins (2020)
- ☐ § 10+ Wheat - Triticum spelta PGSBv2.0 (2020)
- ☐ § 10+ Wheat - CDC Stanley PGSBv2.2 proteins (2020)
- ☐ 10+ Wheat - Cadenza EIv1.1 proteins (2020)
- ☐ 10+ Wheat - Claire EIv1.1 proteins (2020)
- ☐ 10+ Wheat - Paragon EIv1.1 proteins (2020)
- ☐ 10+ Wheat - Robigus EIv1.1 proteins (2020)
- ☐ 10+ Wheat - Weebil V1 proteins (2020)
- ☐ § Triticum turgidum Durum Wheat Svevo Rel. 1.0 Proteins (2019)
- ☐ § Wild Emmer Wheat Zavitan WEWSeq v2.0 proteins (2019)
- ☐ § Triticum urartu proteins, Ling et al. (2018)

Barley Protein Collections

- ☐ § Barley Morex v3 proteins (2021)

Rye Protein Collections

- ☐ § Rye Weining v1 proteins, Li (Mar 2021)

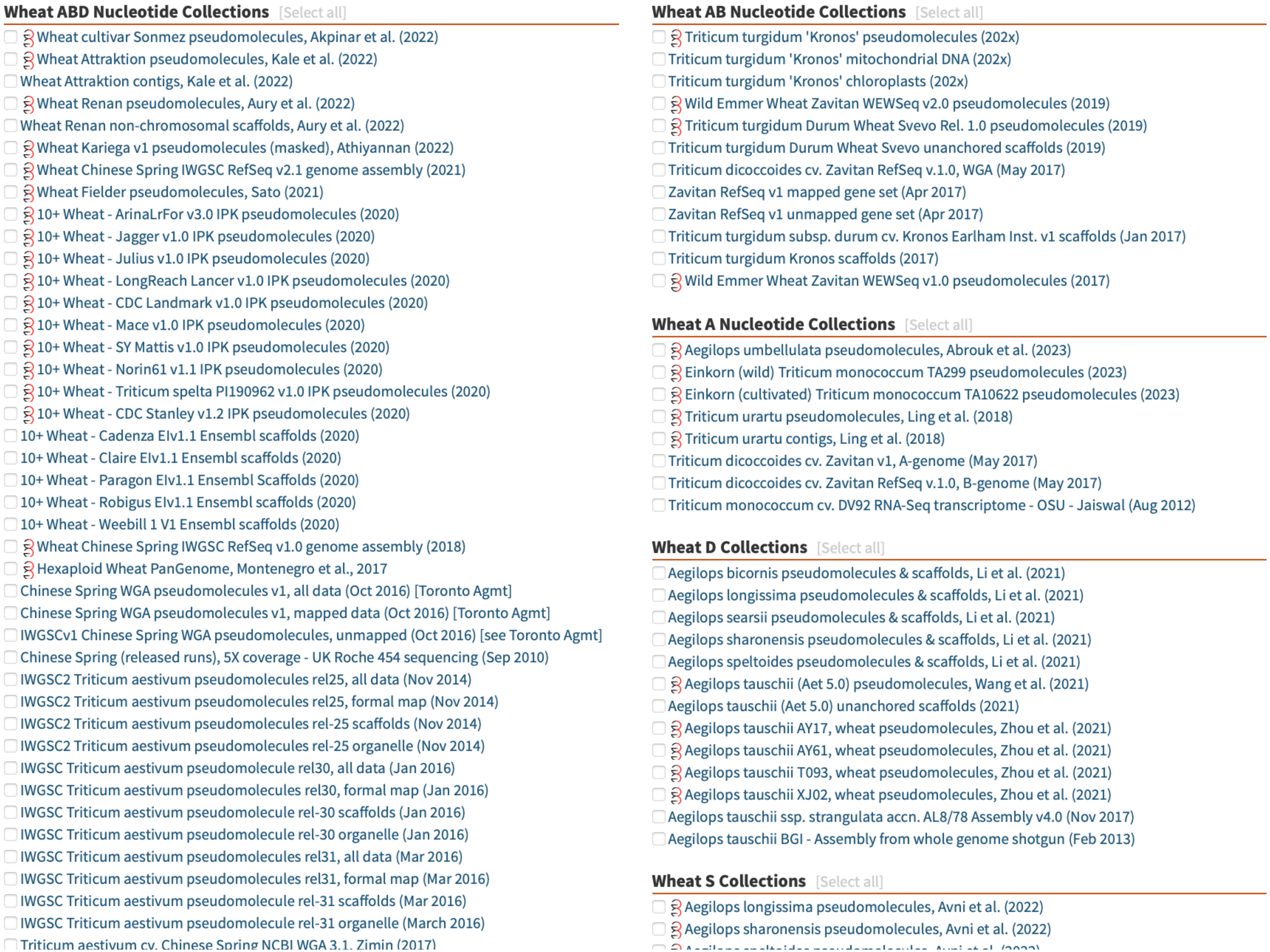
Oat Protein Collections [Select all]

- ☐ § PepsiCo OT3098 Hexaploid Oat v2 proteins (2021)
- ☐ Avena atlantica Diploid Oat proteins (2019)
- ☐ Avena eriantha Diploid Oat proteins (2019)

TREP Protein Collections

- ☐ TREP protein sequences rel-19, Schlagenhauf et al. (2019)

BLAST – Nucleotide - wheat



Wheat ABD Nucleotide Collections [Select all]

- ☐ 🧬 Wheat cultivar Sonmez pseudomolecules, Akpinar et al. (2022)
- ☐ 🧬 Wheat Attraktion pseudomolecules, Kale et al. (2022)
- ☐ Wheat Attraktion contigs, Kale et al. (2022)
- ☐ 🧬 Wheat Renan pseudomolecules, Aury et al. (2022)
- ☐ Wheat Renan non-chromosomal scaffolds, Aury et al. (2022)
- ☐ 🧬 Wheat Kariega v1 pseudomolecules (masked), Athiyannan (2022)
- ☐ 🧬 Wheat Chinese Spring IWGSC RefSeq v2.1 genome assembly (2021)
- ☐ 🧬 Wheat Fielder pseudomolecules, Sato (2021)
- ☐ 🧬 10+ Wheat - ArinaLrFor v3.0 IPK pseudomolecules (2020)
- ☐ 🧬 10+ Wheat - Jagger v1.0 IPK pseudomolecules (2020)
- ☐ 🧬 10+ Wheat - Julius v1.0 IPK pseudomolecules (2020)
- ☐ 🧬 10+ Wheat - LongReach Lancer v1.0 IPK pseudomolecules (2020)
- ☐ 🧬 10+ Wheat - CDC Landmark v1.0 IPK pseudomolecules (2020)
- ☐ 🧬 10+ Wheat - Mace v1.0 IPK pseudomolecules (2020)
- ☐ 🧬 10+ Wheat - SY Mattis v1.0 IPK pseudomolecules (2020)
- ☐ 🧬 10+ Wheat - Norin61 v1.1 IPK pseudomolecules (2020)
- ☐ 🧬 10+ Wheat - Triticum spelta PI190962 v1.0 IPK pseudomolecules (2020)
- ☐ 🧬 10+ Wheat - CDC Stanley v1.2 IPK pseudomolecules (2020)
- ☐ 10+ Wheat - Cadenza EIv1.1 Ensembl scaffolds (2020)
- ☐ 10+ Wheat - Claire EIv1.1 Ensembl scaffolds (2020)
- ☐ 10+ Wheat - Paragon EIv1.1 Ensembl Scaffolds (2020)
- ☐ 10+ Wheat - Robigus EIv1.1 Ensembl scaffolds (2020)
- ☐ 10+ Wheat - Weebill 1 V1 Ensembl scaffolds (2020)
- ☐ 🧬 Wheat Chinese Spring IWGSC RefSeq v1.0 genome assembly (2018)
- ☐ 🧬 Hexaploid Wheat PanGenome, Montenegro et al., 2017
- ☐ Chinese Spring WGA pseudomolecules v1, all data (Oct 2016) [Toronto Agmt]
- ☐ Chinese Spring WGA pseudomolecules v1, mapped data (Oct 2016) [Toronto Agmt]
- ☐ IWGSCv1 Chinese Spring WGA pseudomolecules, unmapped (Oct 2016) [see Toronto Agmt]
- ☐ Chinese Spring (released runs), 5X coverage - UK Roche 454 sequencing (Sep 2010)
- ☐ IWGSC2 Triticum aestivum pseudomolecules rel25, all data (Nov 2014)
- ☐ IWGSC2 Triticum aestivum pseudomolecules rel25, formal map (Nov 2014)
- ☐ IWGSC2 Triticum aestivum pseudomolecules rel-25 scaffolds (Nov 2014)
- ☐ IWGSC2 Triticum aestivum pseudomolecules rel-25 organelle (Nov 2014)
- ☐ IWGSC Triticum aestivum pseudomolecule rel30, all data (Jan 2016)
- ☐ IWGSC Triticum aestivum pseudomolecules rel30, formal map (Jan 2016)
- ☐ IWGSC Triticum aestivum pseudomolecule rel-30 scaffolds (Jan 2016)
- ☐ IWGSC Triticum aestivum pseudomolecule rel-30 organelle (Jan 2016)
- ☐ IWGSC Triticum aestivum pseudomolecules rel31, all data (Mar 2016)
- ☐ IWGSC Triticum aestivum pseudomolecules rel31, formal map (Mar 2016)
- ☐ IWGSC Triticum aestivum pseudomolecule rel-31 scaffolds (Mar 2016)
- ☐ IWGSC Triticum aestivum pseudomolecule rel-31 organelle (March 2016)
- ☐ Triticum aestivum cv. Chinese Spring NCBI WGA 3.1, Zimin (2017)

Wheat AB Nucleotide Collections [Select all]

- ☐ 🧬 Triticum turgidum 'Kronos' pseudomolecules (202x)
- ☐ Triticum turgidum 'Kronos' mitochondrial DNA (202x)
- ☐ Triticum turgidum 'Kronos' chloroplasts (202x)
- ☐ 🧬 Wild Emmer Wheat Zavitan WEWSeq v2.0 pseudomolecules (2019)
- ☐ 🧬 Triticum turgidum Durum Wheat Svevo Rel. 1.0 pseudomolecules (2019)
- ☐ Triticum turgidum Durum Wheat Svevo unanchored scaffolds (2019)
- ☐ Triticum dicoccoides cv. Zavitan RefSeq v.1.0, WGA (May 2017)
- ☐ Zavitan RefSeq v1 mapped gene set (Apr 2017)
- ☐ Zavitan RefSeq v1 unmapped gene set (Apr 2017)
- ☐ Triticum turgidum subsp. durum cv. Kronos Earlham Inst. v1 scaffolds (Jan 2017)
- ☐ Triticum turgidum Kronos scaffolds (2017)
- ☐ 🧬 Wild Emmer Wheat Zavitan WEWSeq v1.0 pseudomolecules (2017)

Wheat A Nucleotide Collections [Select all]

- ☐ 🧬 Aegilops umbellulata pseudomolecules, Abrouk et al. (2023)
- ☐ 🧬 Einkorn (wild) Triticum monococcum TA299 pseudomolecules (2023)
- ☐ 🧬 Einkorn (cultivated) Triticum monococcum TA10622 pseudomolecules (2023)
- ☐ 🧬 Triticum urartu pseudomolecules, Ling et al. (2018)
- ☐ 🧬 Triticum urartu contigs, Ling et al. (2018)
- ☐ Triticum dicoccoides cv. Zavitan v1, A-genome (May 2017)
- ☐ Triticum dicoccoides cv. Zavitan RefSeq v.1.0, B-genome (May 2017)
- ☐ Triticum monococcum cv. DV92 RNA-Seq transcriptome - OSU - Jaiswal (Aug 2012)

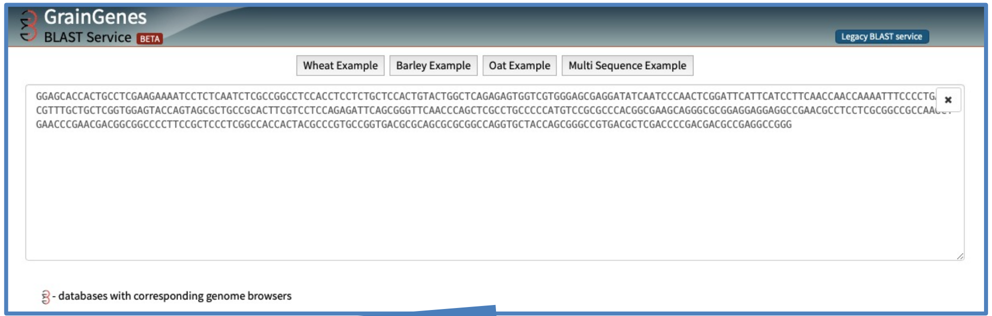
Wheat D Collections [Select all]

- ☐ Aegilops bicornis pseudomolecules & scaffolds, Li et al. (2021)
- ☐ Aegilops longissima pseudomolecules & scaffolds, Li et al. (2021)
- ☐ Aegilops searsii pseudomolecules & scaffolds, Li et al. (2021)
- ☐ Aegilops sharonensis pseudomolecules & scaffolds, Li et al. (2021)
- ☐ Aegilops speltoides pseudomolecules & scaffolds, Li et al. (2021)
- ☐ 🧬 Aegilops tauschii (Aet 5.0) pseudomolecules, Wang et al. (2021)
- ☐ Aegilops tauschii (Aet 5.0) unanchored scaffolds (2021)
- ☐ 🧬 Aegilops tauschii AY17, wheat pseudomolecules, Zhou et al. (2021)
- ☐ 🧬 Aegilops tauschii AY61, wheat pseudomolecules, Zhou et al. (2021)
- ☐ 🧬 Aegilops tauschii T093, wheat pseudomolecules, Zhou et al. (2021)
- ☐ 🧬 Aegilops tauschii XJ02, wheat pseudomolecules, Zhou et al. (2021)
- ☐ Aegilops tauschii ssp. strangulata accn. AL8/78 Assembly v4.0 (Nov 2017)
- ☐ Aegilops tauschii BGI - Assembly from whole genome shotgun (Feb 2013)

Wheat S Collections [Select all]

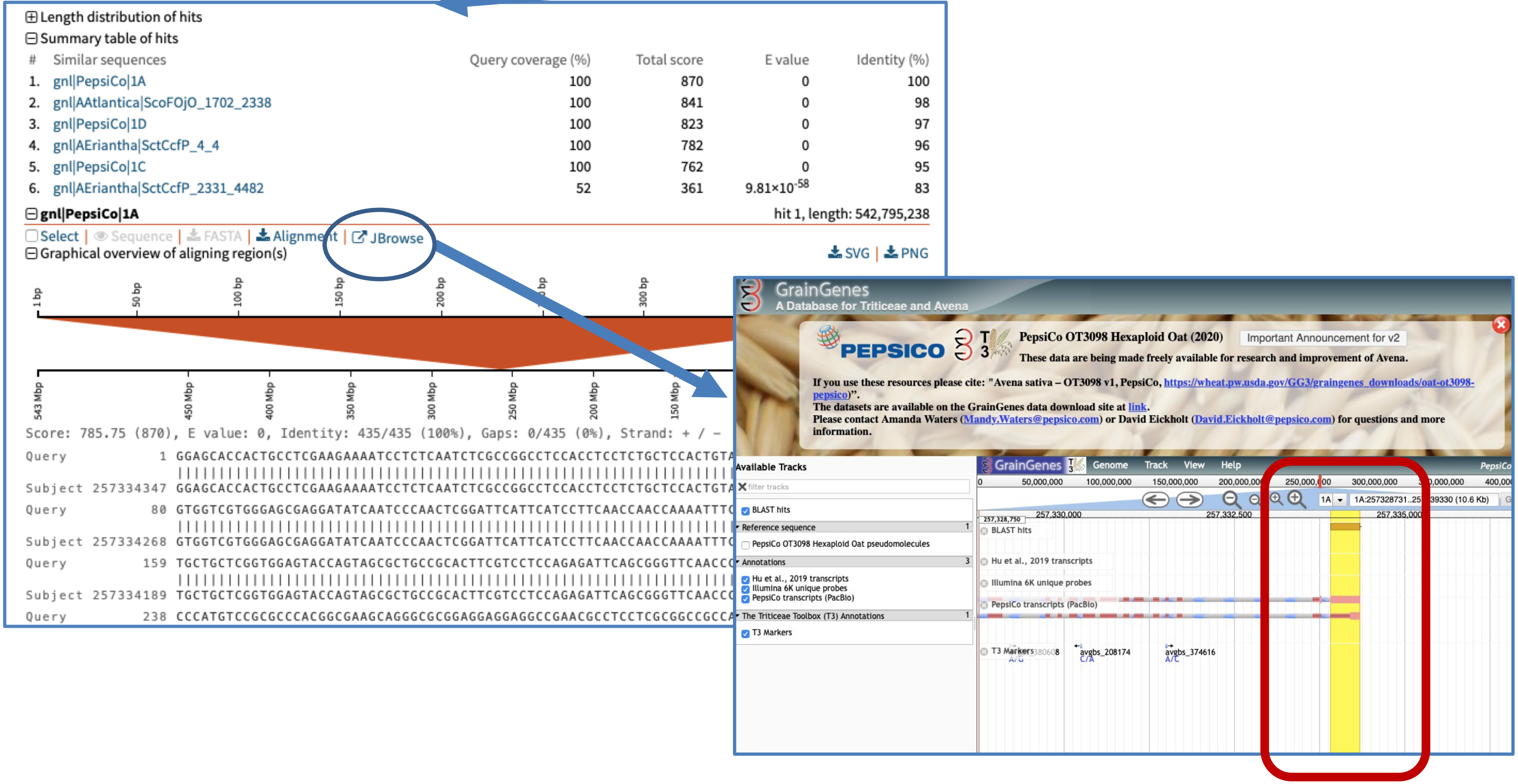
- ☐ 🧬 Aegilops longissima pseudomolecules, Avni et al. (2022)
- ☐ 🧬 Aegilops sharonensis pseudomolecules, Avni et al. (2022)
- ☐ 🧬 Aegilops speltoides pseudomolecules, Avni et al. (2022)

BLAST - Nucleotide



```
GGAGCACCACTGCCTCGAAGAAAATCCTCTCAATCTCGCCGGCCTCCACCTCCTCTGCTCCACTGTACTGGCTCAGAGAGTGGTCGTGGGAGCGAGGATATCAATCCCAACTCGGATTCATTCATCCTTCAACCAACCAAAATTTCCCCTG
CGTTTGCTGCTCGGTGGAGTACCAGTAGCGCTGCCGCACTTCGTCCTCCAGAGATTCAGCGGGTTCAACCCAGCTCGCCTGCCCCCATGTCCGCGCCCACGGCGAAGCAGGGCGCGGAGGAGGAGGCCGAACGCCTCCTCGCGGCCGCCAA
GAACCCGAACGACGGCGGCCCCTTCCGCTCCCTCGGCCACCACTACGCCCGTGCCGGTGACGCGCAGCGCGCGGCCAGGTGCTACCAGCGGGCCGTGACGCTCGACCCCGACGACGCCGAGGCCGGG
```

- databases with corresponding genome browsers



⊞ Length distribution of hits

⊟ Summary table of hits

#	Similar sequences	Query coverage (%)	Total score	E value	Identity (%)
1.	gnl|PepsiCo|1A	100	870	0	100
2.	gnl|AAtlantica|ScoFOjO_1702_2338	100	841	0	98
3.	gnl|PepsiCo|1D	100	823	0	97
4.	gnl|AEriantha|SctCcfP_4_4	100	782	0	96
5.	gnl|PepsiCo|1C	100	762	0	95
6.	gnl|AEriantha|SctCcfP_2331_4482	52	361	9.81×10^{-58}	83

⊟ **gnl|PepsiCo|1A** hit 1, length: 542,795,238

⊟ Graphical overview of aligning region(s)

```
Score: 785.75 (870), E value: 0, Identity: 435/435 (100%), Gaps: 0/435 (0%), Strand: + / -
Query           1 GGAGCACCACTGCCTCGAAGAAAATCCTCTCAATCTCGCCGGCCTCCACCTCCTCTGCTCCACTGTA
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Subject 257334347 GGAGCACCACTGCCTCGAAGAAAATCCTCTCAATCTCGCCGGCCTCCACCTCCTCTGCTCCACTGTA
Query          80 GTGGTCGTGGGAGCGAGGATATCAATCCCAACTCGGATTCATTCATCCTTCAACCAACCAAAATTTC
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Subject 257334268 GTGGTCGTGGGAGCGAGGATATCAATCCCAACTCGGATTCATTCATCCTTCAACCAACCAAAATTTC
Query         159 TGCTGCTCGGTGGAGTACCAGTAGCGCTGCCGCACTTCGTCCTCCAGAGATTCAGCGGGTTCAACCC
                  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Subject 257334189 TGCTGCTCGGTGGAGTACCAGTAGCGCTGCCGCACTTCGTCCTCCAGAGATTCAGCGGGTTCAACCC
Query         238 CCCATGTCCGCGCCCACGGCGAAGCAGGGCGCGGAGGAGGCCGAACGCCTCCTCGCGGCCGCCA
```

PepsiCo OT3098 Hexaploid Oat (2020)

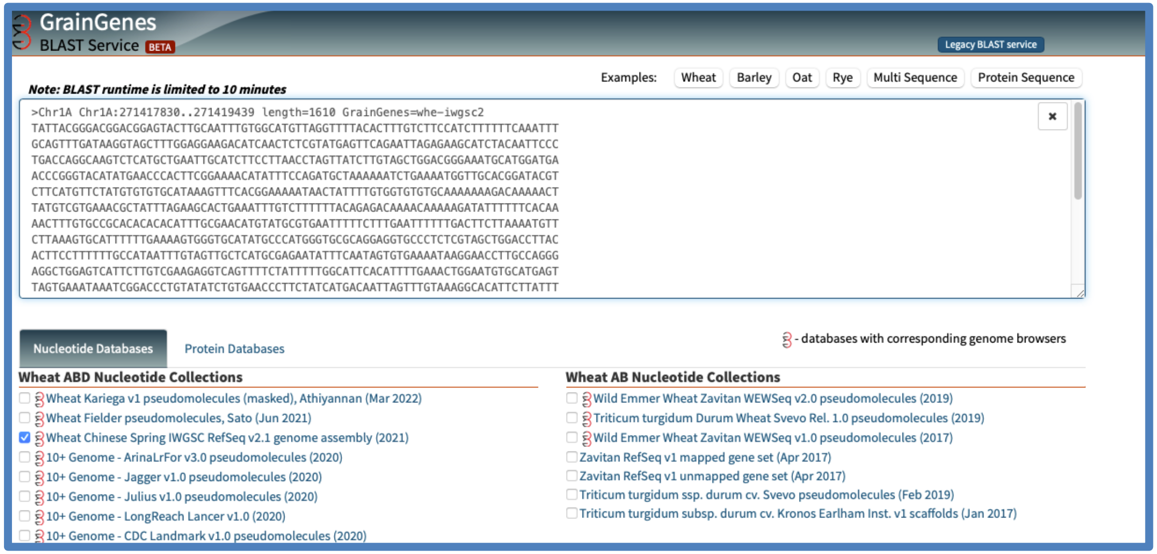
These data are being made freely available for research and improvement of Avena.

If you use these resources please cite: "Avena sativa – OT3098 v1, PepsiCo, https://wheat.pw.usda.gov/GG3/graingenes_downloads/oat-ot3098-pepsico)".
The datasets are available on the GrainGenes data download site at link.
Please contact Amanda Waters (Mandy.Waters@pepsico.com) or David Eickholt (David.Eickholt@pepsico.com) for questions and more information.

Available Tracks

- ☑ BLAST hits
- Reference sequence 1
  - ☐ PepsiCo OT3098 Hexaploid Oat pseudomolecules
- Annotations 3
  - ☑ Hu et al., 2019 transcripts
  - ☑ Illumina 6K unique probes
  - ☑ PepsiCo transcripts (PacBio)
- The Triticeae Toolbox (T3) Annotations 1
  - ☑ T3 Markers

1.3. JBLAST-light



Note: BLAST runtime is limited to 10 minutes

```
>Chr1A Chr1A:271417830..271419439 length=1610 GrainGenes=whe-iwgsc2
TATTACGGGACGGACGGAGTACTTGCAATTTGTGGCATGTTAGGTTTTACACTTTGTCTTCCATCTTTTTTCAAATTT
GCAGTTTGATAAGGTAGCTTTGGAGGAAGACATCAACTCTCGTATGAGTTCAGAATTAGAGAAGCATCTACAATTCCC
TGACCAGGCAAGTCTCATGCTGAATTGCATCTTCCTTAACCTAGTTATCTTGTAGCTGGACGGGAAATGCATGGATGA
ACCCGGGTACATATGAACCCACTTCGGAAAACATATTTCCAGATGCTAAAAAATCTGAAAATGGTTGCACGGATACGT
CTTCATGTTCTATGTGTGTGCATAAAGTTTCACGGAAAAATAACTATTTTGTGGTGTGTGCAAAAAAAGACAAAAACT
TATGTCGTGAAACGCTATTTAGAAGCACTGAAATTTGTCTTTTTTACAGAGACAAAACAAAAAGATATTTTTTCACAA
AACTTTGTGCCGCACACACACATTTGCGAACATGTATGCGTGAATTTTTCTTTGAATTTTTTGACTTCTTAAAATGTT
CTTAAAGTGCATTTTTTGAAAAGTGGGTGCATATGCCCATGGGTGCGCAGGAGGTGCCCTCTCGTAGCTGGACCTTAC
ACTTCCTTTTTTGCCATAATTTGTAGTTGCTCATGCGAGAATATTTCAATAGTGTGAAAATAAGGAACCTTGCCAGGG
AGGCTGGAGTCATTCTTGTCGAAGAGGTCAGTTTTCTATTTTTGGCATTCACATTTTGAAACTGGAATGTGCATGAGT
TAGTGAAATAAATCGGACCCTGTATATCTGTGAACCCTTCTATCATGACAATTAGTTTGTAAAGGCACATTCTTATTT
```

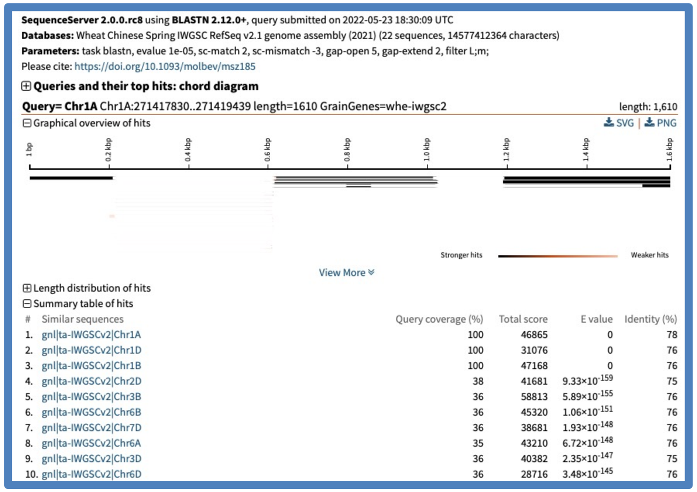
- databases with corresponding genome browsers

Wheat ABD Nucleotide Collections

- ☐ Wheat Kariega v1 pseudomolecules (masked), Athiyannan (Mar 2022)
- ☐ Wheat Fielder pseudomolecules, Sato (Jun 2021)
- ☑ Wheat Chinese Spring IWGSC RefSeq v2.1 genome assembly (2021)
- ☐ 10+ Genome - ArinaLrFor v3.0 pseudomolecules (2020)
- ☐ 10+ Genome - Jagger v1.0 pseudomolecules (2020)
- ☐ 10+ Genome - Julius v1.0 pseudomolecules (2020)
- ☐ 10+ Genome - LongReach Lancer v1.0 (2020)
- ☐ 10+ Genome - CDC Landmark v1.0 pseudomolecules (2020)

Wheat AB Nucleotide Collections

- ☐ Wild Emmer Wheat Zavitan WEWSeq v2.0 pseudomolecules (2019)
- ☐ Triticum turgidum Durum Wheat Svevo Rel. 1.0 pseudomolecules (2019)
- ☐ Wild Emmer Wheat Zavitan WEWSeq v1.0 pseudomolecules (2017)
- ☐ Zavitan RefSeq v1 mapped gene set (Apr 2017)
- ☐ Zavitan RefSeq v1 unmapped gene set (Apr 2017)
- ☐ Triticum turgidum ssp. durum cv. Svevo pseudomolecules (Feb 2019)
- ☐ Triticum turgidum subsp. durum cv. Kronos Earlham Inst. v1 scaffolds (Jan 2017)



SequenceServer 2.0.0.rc8 using **BLASTN 2.12.0+**, query submitted on 2022-05-23 18:30:09 UTC
Databases: Wheat Chinese Spring IWGSC RefSeq v2.1 genome assembly (2021) (22 sequences, 14577412364 characters)
Parameters: task blastn, evalue 1e-05, sc-match 2, sc-mismatch -3, gap-open 5, gap-extend 2, filter L;m;
Please cite: https://doi.org/10.1093/molbev/msz185

⊞ **Queries and their top hits: chord diagram**

Query= Chr1A Chr1A:271417830..271419439 length=1610 GrainGenes=whe-iwgsc2 length: 1,610

⊟ Graphical overview of hits

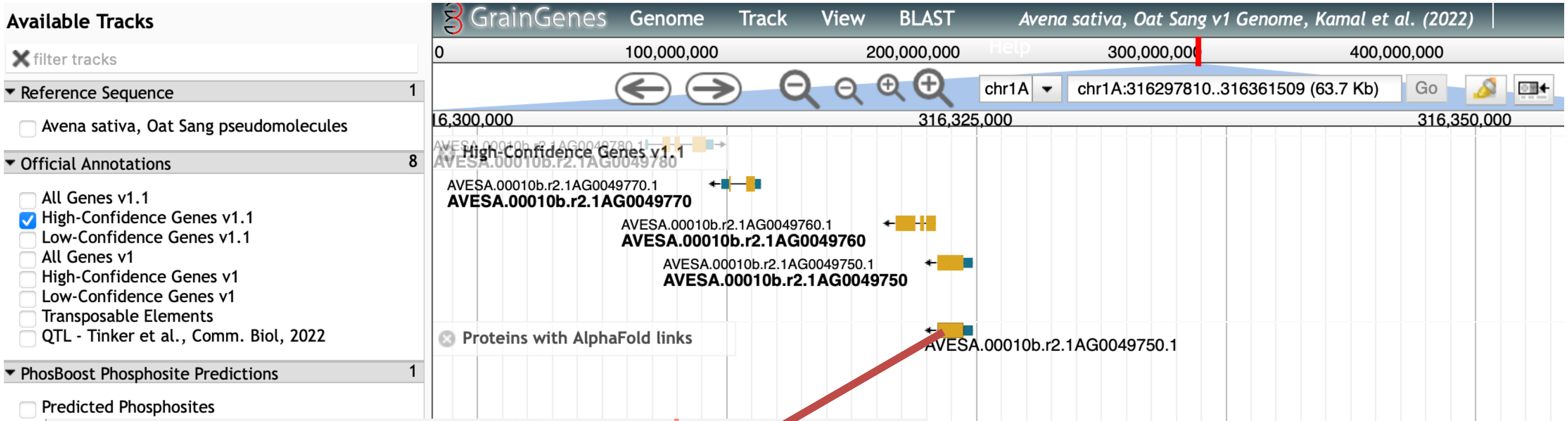
⊞ Length distribution of hits

⊟ Summary table of hits

#	Similar sequences	Query coverage (%)	Total score	E value	Identity (%)
1.	gnl|ta-IWGSCv2|Chr1A	100	46865	0	78
2.	gnl|ta-IWGSCv2|Chr1D	100	31076	0	76
3.	gnl|ta-IWGSCv2|Chr1B	100	47168	0	76
4.	gnl|ta-IWGSCv2|Chr2D	38	41681	9.33×10^{-159}	75
5.	gnl|ta-IWGSCv2|Chr3B	36	58813	5.89×10^{-155}	76
6.	gnl|ta-IWGSCv2|Chr6B	36	45320	1.06×10^{-151}	76
7.	gnl|ta-IWGSCv2|Chr7D	36	38681	1.93×10^{-148}	76
8.	gnl|ta-IWGSCv2|Chr6A	35	43210	6.72×10^{-148}	76
9.	gnl|ta-IWGSCv2|Chr3D	36	40382	2.35×10^{-147}	75
10.	gnl|ta-IWGSCv2|Chr6D	36	28716	3.48×10^{-145}	76

Genome Browser – AlphaFold/Protein 3D structure links

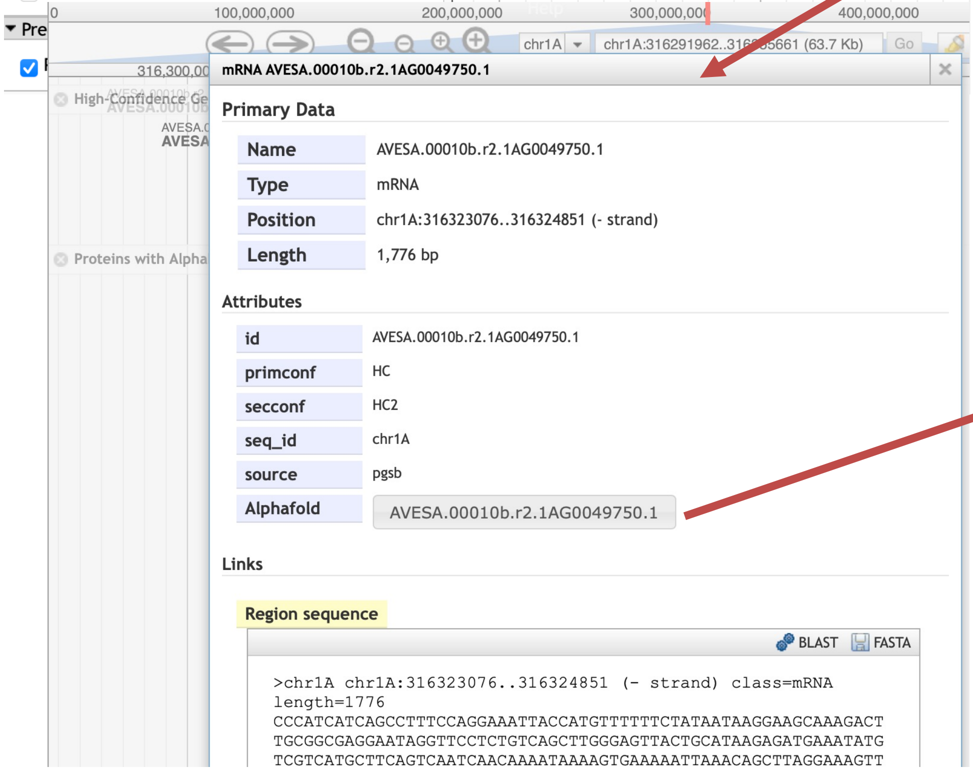
(Chinese Spring IWGSC v1 | Durum Wheat | Aegilops v4 | Triticum urartu | Barley Morex v3 | Oat Sang)



Available Tracks

filter tracks

- Reference Sequence 1
  - ☐ Avena sativa, Oat Sang pseudomolecules
- Official Annotations 8
  - ☐ All Genes v1.1
  - ☑ High-Confidence Genes v1.1
  - ☐ Low-Confidence Genes v1.1
  - ☐ All Genes v1
  - ☐ High-Confidence Genes v1
  - ☐ Low-Confidence Genes v1
  - ☐ Transposable Elements
  - ☐ QTL - Tinker et al., Comm. Biol, 2022
- PhosBoost Phosphosite Predictions 1
  - ☐ Predicted Phosphosites



mRNA AVESA.00010b.r2.1AG0049750.1

Primary Data

Name	AVESA.00010b.r2.1AG0049750.1
Type	mRNA
Position	chr1A:316323076..316324851 (- strand)
Length	1,776 bp

Attributes

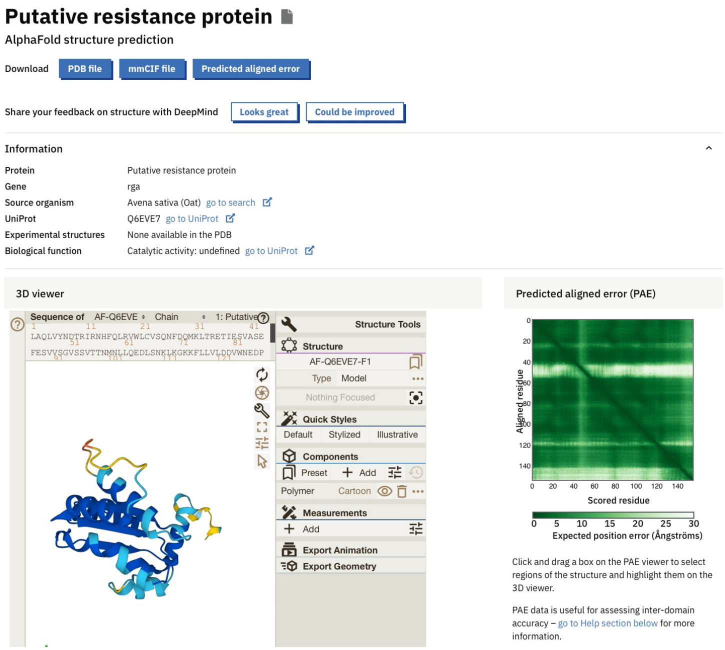
id	AVESA.00010b.r2.1AG0049750.1
primconf	HC
secconf	HC2
seq_id	chr1A
source	pgsb
Alphafold	AVESA.00010b.r2.1AG0049750.1

Links

Region sequence

BLAST FASTA

```
>chr1A chr1A:316323076..316324851 (- strand) class=mRNA
length=1776
CCCATCATCAGCCTTTCCAGGAAATTACCATGTTTTTTCTATAATAAGGAAGCAAAGACT
TGCGGCGAGGAATAGGTTCCTCTGTCAGCTTGGGAGTTACTGCATAAGAGATGAAATATG
TCGTCATGCTTCAGTCAATCAACAAAATAAAAGTGAAAAATTAAACAGCTTAGGAAAGTT
```



Putative resistance protein

AlphaFold structure prediction

Share your feedback on structure with DeepMind

Information

Protein	Putative resistance protein
Gene	rga
Source organism	Avena sativa (Oat) go to search
UniProt	Q6EVE7 go to UniProt
Experimental structures	None available in the PDB
Biological function	Catalytic activity: undefined go to UniProt

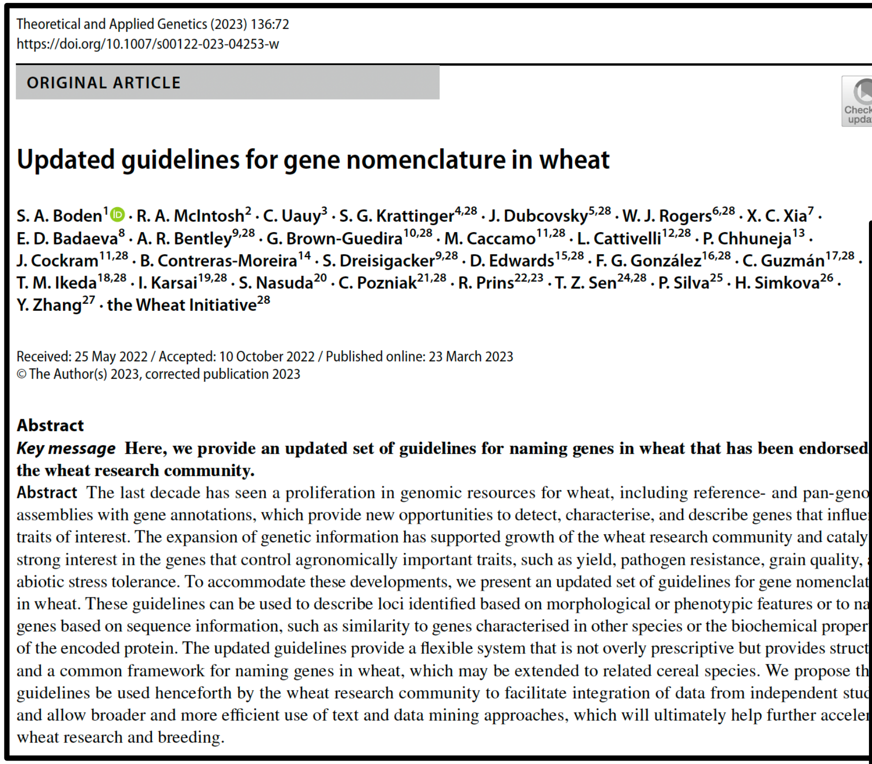
3D viewer

Predicted aligned error (PAE)

Click and drag a box on the PAE viewer to select regions of the structure and highlight them on the 3D viewer.

PAE data is useful for assessing inter-domain accuracy – go to Help section below for more information.

Nomenclature & Data Standards



Theoretical and Applied Genetics (2023) 136:72
https://doi.org/10.1007/s00122-023-04253-w

ORIGINAL ARTICLE

Updated guidelines for gene nomenclature in wheat

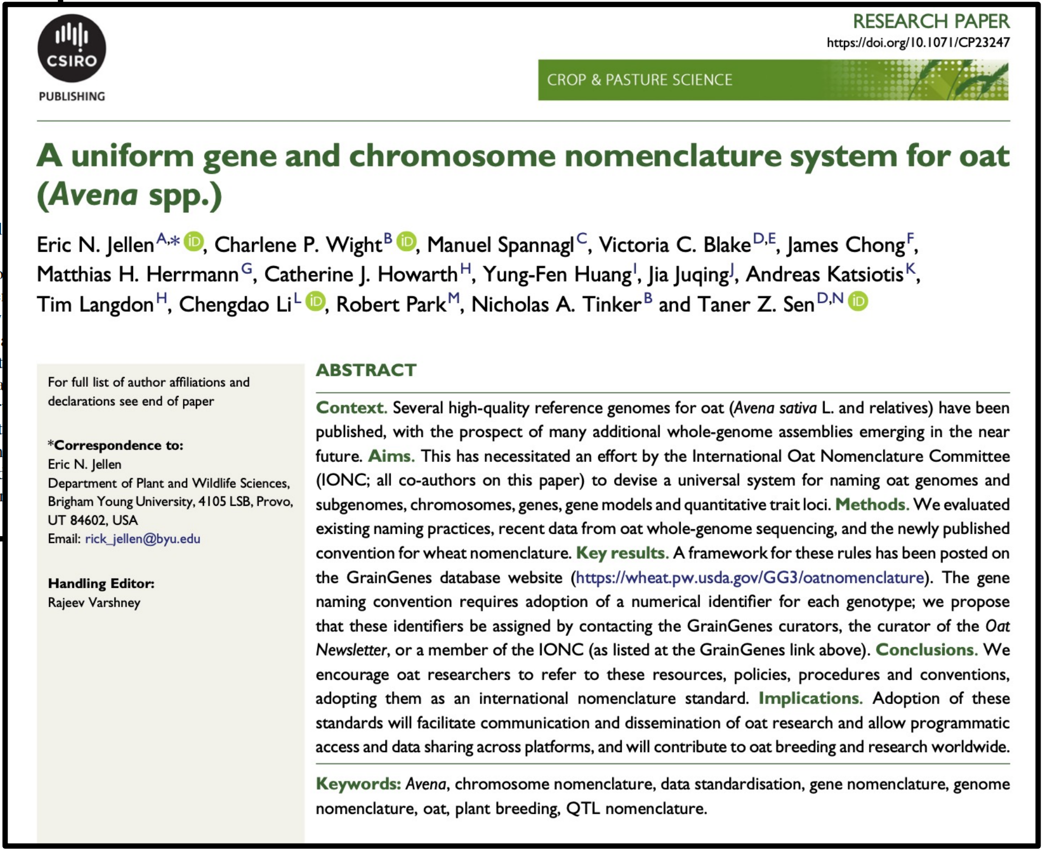
S. A. Boden[1] · R. A. McIntosh[2] · C. Uauy[3] · S. G. Krattinger[4,28] · J. Dubcovsky[5,28] · W. J. Rogers[6,28] · X. C. Xia[7] · E. D. Badaeva[8] · A. R. Bentley[9,28] · G. Brown-Guedira[10,28] · M. Caccamo[11,28] · L. Cattivelli[12,28] · P. Chhuneja[13] · J. Cockram[11,28] · B. Contreras-Moreira[14] · S. Dreisigacker[9,28] · D. Edwards[15,28] · F. G. González[16,28] · C. Guzmán[17,28] · T. M. Ikeda[18,28] · I. Karsai[19,28] · S. Nasuda[20] · C. Pozniak[21,28] · R. Prins[22,23] · T. Z. Sen[24,28] · P. Silva[25] · H. Simkova[26] · Y. Zhang[27] · the Wheat Initiative[28]

Received: 25 May 2022 / Accepted: 10 October 2022 / Published online: 23 March 2023
© The Author(s) 2023, corrected publication 2023

Abstract

Key message **Here, we provide an updated set of guidelines for naming genes in wheat that has been endorsed the wheat research community.**

Abstract The last decade has seen a proliferation in genomic resources for wheat, including reference- and pan-geno assemblies with gene annotations, which provide new opportunities to detect, characterise, and describe genes that influe traits of interest. The expansion of genetic information has supported growth of the wheat research community and cataly strong interest in the genes that control agronomically important traits, such as yield, pathogen resistance, grain quality, abiotic stress tolerance. To accommodate these developments, we present an updated set of guidelines for gene nomenclat in wheat. These guidelines can be used to describe loci identified based on morphological or phenotypic features or to na genes based on sequence information, such as similarity to genes characterised in other species or the biochemical proper of the encoded protein. The updated guidelines provide a flexible system that is not overly prescriptive but provides struct and a common framework for naming genes in wheat, which may be extended to related cereal species. We propose th guidelines be used henceforth by the wheat research community to facilitate integration of data from independent stud and allow broader and more efficient use of text and data mining approaches, which will ultimately help further acceler wheat research and breeding.



CSIRO PUBLISHING

RESEARCH PAPER
https://doi.org/10.1071/CP23247

CROP & PASTURE SCIENCE

A uniform gene and chromosome nomenclature system for oat (*Avena* spp.)

Eric N. Jellen[A,*], Charlene P. Wight[B], Manuel Spannagl[C], Victoria C. Blake[D,E], James Chong[F], Matthias H. Herrmann[G], Catherine J. Howarth[H], Yung-Fen Huang[I], Jia Juqing[J], Andreas Katsiotis[K], Tim Langdon[H], Chengdao Li[L], Robert Park[M], Nicholas A. Tinker[B] and Taner Z. Sen[D,N]

For full list of author affiliations and declarations see end of paper

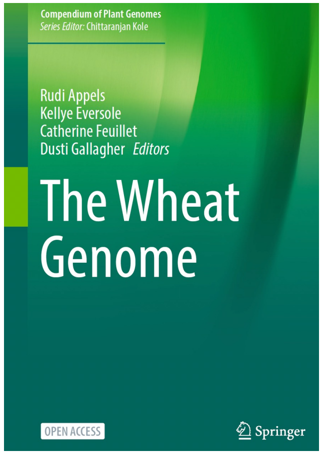
***Correspondence to:**
Eric N. Jellen
Department of Plant and Wildlife Sciences, Brigham Young University, 4105 LSB, Provo, UT 84602, USA
Email: rick_jellen@byu.edu

Handling Editor:
Rajeev Varshney

ABSTRACT

Context. Several high-quality reference genomes for oat (*Avena sativa* L. and relatives) have been published, with the prospect of many additional whole-genome assemblies emerging in the near future. **Aims.** This has necessitated an effort by the International Oat Nomenclature Committee (IONC; all co-authors on this paper) to devise a universal system for naming oat genomes and subgenomes, chromosomes, genes, gene models and quantitative trait loci. **Methods.** We evaluated existing naming practices, recent data from oat whole-genome sequencing, and the newly published convention for wheat nomenclature. **Key results.** A framework for these rules has been posted on the GrainGenes database website (https://wheat.pw.usda.gov/GG3/oatnomenclature). The gene naming convention requires adoption of a numerical identifier for each genotype; we propose that these identifiers be assigned by contacting the GrainGenes curators, the curator of the *Oat Newsletter*, or a member of the IONC (as listed at the GrainGenes link above). **Conclusions.** We encourage oat researchers to refer to these resources, policies, procedures and conventions, adopting them as an international nomenclature standard. **Implications.** Adoption of these standards will facilitate communication and dissemination of oat research and allow programmatic access and data sharing across platforms, and will contribute to oat breeding and research worldwide.

Keywords: *Avena*, chromosome nomenclature, data standardisation, gene nomenclature, genome nomenclature, oat, plant breeding, QTL nomenclature.



Compendium of Plant Genomes
Series Editor: Chittaranjan Kole

Rudi Appels
Kellye Eversole
Catherine Feuillet
Dusti Gallagher *Editors*

The Wheat Genome

OPEN ACCESS

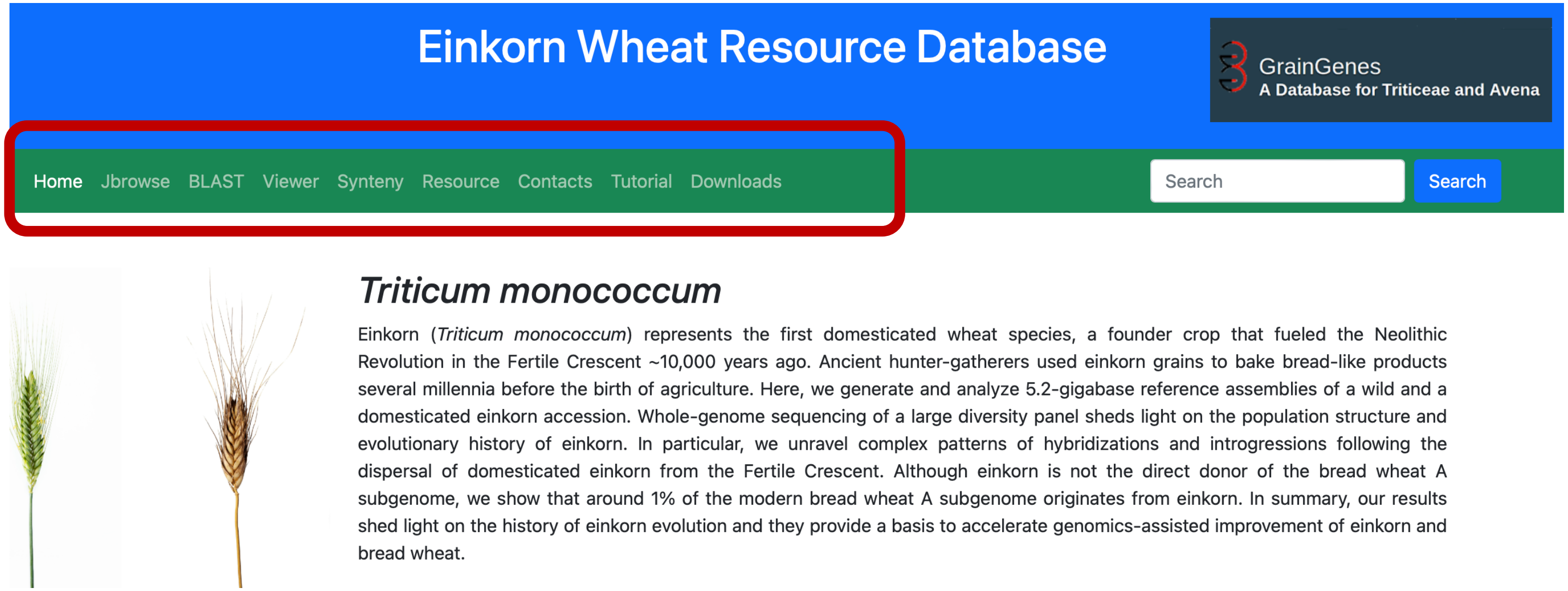
Springer

2

Wheat Data Integration and FAIRification: IWGSC, GrainGenes, Ensembl and Other Data Repositories

Michael Alaux, Sarah Dyer and Taner Z. Sen

Einkorn Pangenome (https://wheat.pw.usda.gov/GG3/pangenome)



Triticum monococcum

Einkorn (*Triticum monococcum*) represents the first domesticated wheat species, a founder crop that fueled the Neolithic Revolution in the Fertile Crescent ~10,000 years ago. Ancient hunter-gatherers used einkorn grains to bake bread-like products several millennia before the birth of agriculture. Here, we generate and analyze 5.2-gigabase reference assemblies of a wild and a domesticated einkorn accession. Whole-genome sequencing of a large diversity panel sheds light on the population structure and evolutionary history of einkorn. In particular, we unravel complex patterns of hybridizations and introgressions following the dispersal of domesticated einkorn from the Fertile Crescent. Although einkorn is not the direct donor of the bread wheat A subgenome, we show that around 1% of the modern bread wheat A subgenome originates from einkorn. In summary, our results shed light on the history of einkorn evolution and they provide a basis to accelerate genomics-assisted improvement of einkorn and bread wheat.

Article

Einkorn genomics sheds light on history of the oldest domesticated wheat

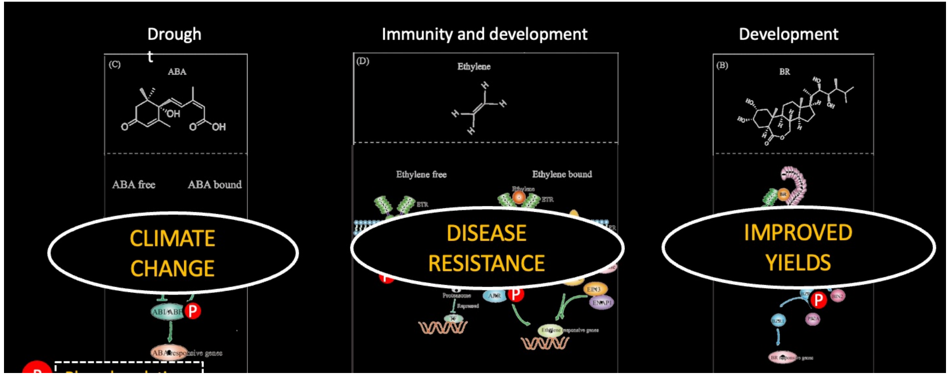
https://doi.org/10.1038/s41586-023-06389-7
Received: 16 October 2022
Accepted: 29 June 2023
Published online: 2 August 2023
Open access
Check for updates

Hanin Ibrahim Ahmed[1,2,11], Matthias Heuberger[3,11], Adam Schoen[4,11], Dal-Hoe Koo[5], Jesus Quiroz-Chavez[6], Laxman Adhikari[1,2], John Raupp[5], Stéphane Cauet[7], Nathalie Rodde[7], Charlotte Cravero[7], Caroline Callot[7], Gerard R. Lazo[8], Nagarajan Kathiresan[9], Parva K. Sharma[4], Ian Moot[4], Inderjit Singh Yadav[4], Lovepreet Singh[4], Gautam Saripalli[4], Nidhi Rawat[4], Raju Datla[10], Naveenkumar Athiyannan[1,2], Ricardo H. Ramirez-Gonzalez[6], Cristobal Uauy[6], Thomas Wicker[3], Vijay K. Tiwari[4]✉, Michael Abrouk[1,2]✉, Jesse Poland[1,2]✉ & Simon G. Krattinger[1,2]✉

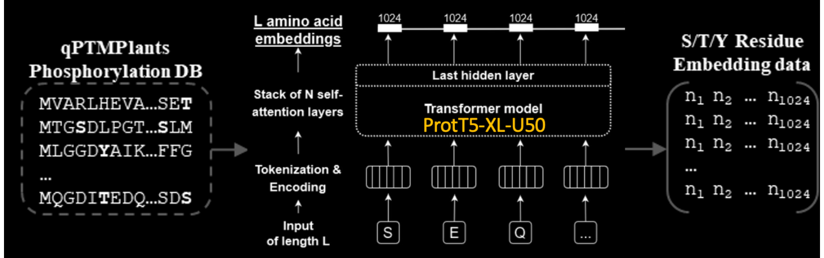
Einkorn (*Triticum monococcum*) was the first domesticated wheat species, and was central to the birth of agriculture and the Neolithic Revolution in the Fertile Crescent around 10,000 years ago[1,2]. Here we generate and analyse 5.2-Gb genome assemblies for wild and domesticated einkorn, including completely assembled centromeres. Einkorn centromeres are highly dynamic, showing evidence of ancient and recent centromere shifts caused by structural rearrangements. Whole-genome sequencing analysis of a diversity panel uncovered the population structure and evolutionary history of einkorn, revealing complex patterns of hybridizations and introgressions after the dispersal of domesticated einkorn from the Fertile Crescent. We also show that around 1% of the modern bread wheat (*Triticum aestivum*) A subgenome originates from einkorn. These resources and findings highlight the history of einkorn evolution and provide a basis to accelerate the genomics-assisted improvement of einkorn and bread wheat.

Research: Phosphorylation Prediction

Protein phosphorylation plays an important role in all aspects of plant development, immunity and stress response



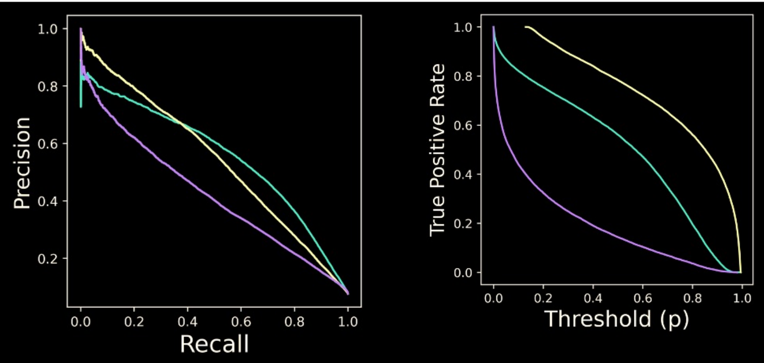
Using Protein Language Models (PLMs) to generate input data for a protein phosphorylation machine learning classifier



Protein Language Models (PLMs) capture functional and structural information for each amino-acid

qPTMplants: Xue (2022) Nucl. Acids Res.
ProtT5: Elnaggar *et al.* (2021) IEEE TPAMI

PhosBoost performs better than DeepPhos and is comparable to PhosphoLingo while providing higher recall



Classifier (AUPRC)
- 🟪 **DeepPhos (0.42)**
- 🟩 **PhosphoLingo (0.56)**
- 🟨 **PhosBoost (0.54)**

Classifier (% Recall)
- 🟪 **DeepPhos (14%)**
- 🟩 **PhosphoLingo (56%)**
- 🟨 **PhosBoost (78%)**

Higher recall comes at the tradeoff of reduced precision (more False Positives)

Poretsky, Andorf, Sen, *Plant Direct*, 2023

GrainGenes
A Database for Triticeae and Avena

- Please cite us: Yao et al., Database 2022

- We appreciate
  - Data generators – contact us for your needs
  - GrainGenes Liaison Committee

USDA
American Malting Barley Association
Berkeley University of California
Montana State University
Agriculture and Agri-Food Canada
IPK Leibniz Institute
International Wheat Genome Sequencing Consortium
AgBioData
Toward enhanced genomics, genetics, and breeding research outcomes through standardization of practices and protocols across agricultural databases
wheatIS
Wheat Information System