

Whealbi

Wheat and barley Legacy for Breeding Improvement

Grant agreement number: FP7-613556

**Collaborative Project
SEVENTH FRAMEWORK PROGRAMME**

Deliverable D4.1

**Database containing sequence variation
in >500 wheat and >500 barley accessions
across exome capture design space**

Due date: M24

Actual submission date: M32

Project start date: January 1st, 2014 **Duration:** 60 months

Workpackage concerned: WP4

Concerned workpackage leader: WU/DLO

Dissemination level: PU

Summary

Objectives: The WP2 generated a massive amount of resequencing and genotyping data which is stored on the information systems located at HMGU and has to be integrated into an externally accessible data retrieval system.

Rationale: The variation data and associated meta information describing individual accessions were integrated into an indexed data structure to allow rapid access. Access to the data structures is enabled through a RESTful web application which allows querying the database and gives a tabular and graphical representation of the results.

Teams involved: HMGU, INRA

Database

The variation data derived from 500 barley and 500 wheat accessions (WP2) and the related metadata (Taxonomy, cultivar, biological state) were integrated into an indexed data structure to enable rapid access and searches for variation data in defined genomic regions (e.g. gene). The data structure is optimized for position-based queries but additional indexes can be created to support unanticipated use-cases and to improve run time of complex queries. The data structures are designed to be easily extendible. Once the data has been accessed, queries for those identifiers will also be enabled.

Web application

An intuitive web application has been developed and made available to allow querying the available variation data through a web browser. The query results are enriched with a summarizing chart showing the distribution of genotype calls across biological states and the genomic context of the variant. Details of variants are augmented with meta data (taxonomy, cultivar, biological condition). Links to external resources can easily be enabled to integrate this resource with existing databases.

The web application follows the RESTful design pattern to enable standardized interoperation between different sites and allow the integration of data into other systems using a common data exchange format (e.g. JSON). An access control system has been implemented to protect unpublished variation data from unauthorized access.

The web application is available at the following URL

<http://pgsb.helmholtz-muenchen.de/whealbi>

URL, login credentials and further updates will be announced to the partners with according instructions.

Overview of results from a range search.

Wheatbi ☰

Note: This site will be continuously updated and improved.

Search results

Contig	Position	Reference base	Alternative base	Link
Traes_1AL_00530B6DB.3.mrna1	229	G	[T]	details
Traes_1AL_00530B6DB.3.mrna1	414	C	[T]	details
Traes_1AL_00530B6DB.3.mrna1	621	C	[T]	details
Traes_1AL_00530B6DB.3.mrna1	774	G	[A]	details

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°FP7-613556

Details for a particular variant in wheat.

Search... [Home](#) [Wheatbi project](#)

Note: This site will be continuously updated and improved.

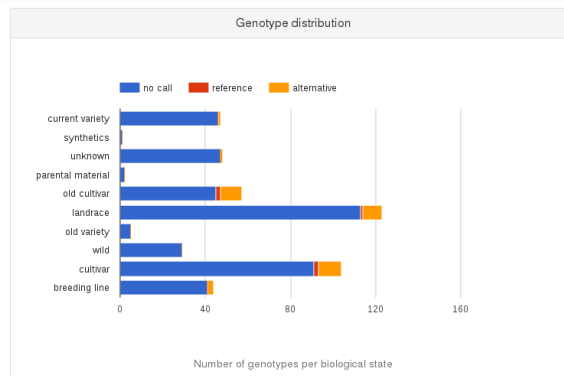
Traes_1AL_00530B6DB.3.mrna1 at position 229

Summary

Genomic context

```

GCCCGGAT CCCGATCC CATCACGG CTCACGGGT TCCGGTCTGT CTGCGGACA TGGCGCGCG CGTCAGGGCA
AGCGCTGCG CTGCTGTC CCSCCCCTG GAACCGCTC GCTCCCGTG CCGGCTAGC GACGCGCTC GTCTCCCT
TCCC89CG 6T6TCCGTC 6TCCGATG CCGCGACGC GTCTGCGCG CTCTGCGACC TGACCACCG CCGCCACBCC
GCTCTGCCA ABATCAGGC GTCCACGAC GGCAGACCA TGTCCGACAT CAGGTACGTG CACCATACGG CTGACGABG
CGCGCGAAG GTCCAGATCC GCACCCCTT CTATGATGA TCCTTCGAA CATACCTGA ATTCTCAGT CACTTCGCT
TGCCATTCA TGCCCGGCC GGTCTTTGT CAGCAAGCAC ABACGACATG ATGATCATCA TGTCCACCG CTGCTTGTG
GAACGABAGA GCATAGTGT CTGCGCGCG GAGACGCGG CCTGGACAA GCTGCCAAA CCGCCGCTT CTGCTTCTGA
CTATTTATC GATGTAAGCT CTTTTCAGG CAAGATTTAT G6CTTGBAAT CTAATG9TGC CACATTTG9T TTTGACCAA
CCACACTGA GTTCTCTGC TGBATGATG C6CCACAGT GAGTCCAA CTTTACTCCA T6ATATTTG GCTTCAAAA
CATGATGATC CTGTGATTT C6ACTACTC CATCTGTTG CATTGCCAAG CAAGTACTT TTGTCTGTC TAAAGGTGAA
ATCATTGAG CCAGCAGTT TTGCTTCTT TGAATTAAG TCTGGATCTA G6GATGGCCA GCTCTCCTGG CGTAAGGTAA
CAGGAGTGG AATCGGTGG AACACGATG TGTTCATGGA TTGCCACCAT GCGACCTTTA GCGACAATA T6GTGTTGGA
ACCGAGCTC G6ATATATTA TGTCTTAC C6GACGTGG ATCCCAAAG ATCGCGTAC TACTATGAG T6CATGACGG
CAAGATGGAG TGCGTGA CAACACTGA CGATAATGT GAATACTCAA CTAAGCTAG TTGTTTGT CCTAG
    
```



Sample data

Sample ID	Taxonomy	Genome	Variety	Biological state	Genotype	Depth	Genotype Quality
WW-001	Triticum aestivum	BAD	LEEDS	Current variety	no call	1	4
WW-002	Triticum aestivum	BAD	STADIUM	Current variety	no call	2	7
WW-003	Triticum aestivum	BAD	CALISOL	Current variety	no call	1	4

Conclusion

We provide a database containing the variation data generated in WP2 and the associated web application using current technology including intuitive search capabilities. Regarding the dynamic nature of the available data (e.g. new reference sequences, additional data sources, unanticipated use-cases/queries), the database and web interface will receive further improvements and additional features in the course of the project. Possible extensions include additional querying capabilities (e.g. by gene id or SNP accessions when those become available) or integration into the PlantsDB framework.