Development of wheat genomic resources in WheatCAP

Eduard Akhunov Wheat Genetics Resource Center Kansas State University





United States Department of Agriculture National Institute of Food and Agriculture

WheatCAP meeting, PAG30, San Diego, Jan. 15, 2023

Constructing genome-to-phenome maps is critical for detecting functional variants affecting agronomic traits



KANSAS STATE



Develop eQTL maps for wheat

200 genetically diversity wheat lines (worldwide sample) ~ 3 million SNPs mapped to CS RefSeq genome

2-week seedlings

8568 cis-eQTL -> 8837 transcripts 14,645 trans-eQTL -> 8500 transcripts **Post-anthesis spikes** 3172 cis-eQTL -> 3476 transcripts 9891 trans- eQTL -> 7250 transcripts







eQTL maps for wheat seedlings and spikes in GrainGenes

G-quadruplexes 1	🗟 GrainGenes 🚮 🕺 Geno	me Track View	Wheat Chinese Spring IWGSC RefSeq v1.0 genome a
G4 Motifs	0 100,000,000	200,000,000	61 Hel 300,000 400,000,000
MNase Chromatin States 4	\leftarrow	$ \supset Q Q Q G$	chr1A - chr1A:294565001299340000 (4.78 l
MNase sensitive footprint (MSF)	295,000,000	296,250,000	297,500,000
 MSF intersets w/ imputed 1000 Wheat Exomes SNPs MNase resistant footprint (MRF) MRF intersets w/ imputed 1000 Wheat Exomes SNPs 	chr1A_294860748	A_295459044	chr1A_297380842
10+ Wheat Genome Project 3			
 Genes Transposons LTR Retrotransposons 		hr1A_295459044 ssociated Transcripts: TraesCS1A02G164400.1 TraesCS1A02G164600.1	
Expression QTLs 3			
 HC annotations v1.1 w/ eQTL matches eQTL Seedling SNPs eQTL Spike SNPs 			
Varietal SNPs (Akhunov & Dubcovsky Labs) 61			
1. All Accessions with SNPEff			
▼ Hexaploid 51	TraesCS1402G163900 1	esCS1A02G164400 1	TracsCS1402G164600 1 TracsCS1402G165500 1
Berkut RAC875 Duster CHOTEAU	HC annotations v1.1 w/ eQTL ma TraesCS1A02G164000.1 + TraesCS1A02G164000 TraesCS1A02G164200.1 + TraesCS1A02G164200.2 +	teres1A02G164400	TraesCS1A02G164600 TraesCS1A02G1655 TraesCS1A02G164800.1 ← TraesCS1A02G164800.2 ← TraesCS1A02G164800 TraesCS1A02G164800



https://graingenes.org/jb/?data=%2Fggds%2Fwhe-iwgsc2018



Genomic resources for characterizing regulatory diversity

- 200 accessions for studying regulatory diversity in the wheat genome
 - US wheat breeding programs
 - Founders of NAM populations
 - Primary synthetics and wild wheat from introgression mapping projects
- Characterization of regulatory diversity in the wheat genome
 - Tiered sequencing strategy
 - Few genomes at whole genome level
 - Regulatory sequence capture
 - Low-pass sequencing (Skim-Seq)
 - Updating PHG and Imputation
- Tissues and developmental stages selected for Quant-seq analyses
 - 10 tissues; developmental stages related to biomass production and grain yield
- eQTL mapping and integration with other OMICs data (epigenome, chromatin acc.)
- Joint analysis to identify drivers of trait variation





Epigenomic variation in wheat

- 96 winter wheat cultivars from the US Great Planes breeding programs
- 6.8 Tb bisulfite-treated sequencing data
- 1.5x coverage of each genome
- Call methylated positions in the CHH, CHG or CpG contexts







Coefficient of variation in methylation levels across lines is substantially higher in gene space than in the repetitive portion of genome



CV, methylation%





Updating Practical Haplotype Graph

Version 1 (G3, 2022)



Whole-exome capture sequencing data from a diverse set of 65 wheat accessions

Imputation accuracy: 92-97%

Version 2 (developed and will be shared with T3 for database-wide imputations)

Whole exome capture data (Genotyping Labs + KSU group)

Alignment of additional 250 wheat exomes to CS RefSeq v.2. (U. Yunusbaev, KSU)

~470 wheat lines representing majority of the US breeding programs



Construction of PHG v.2 (K. Jordan, USDA ARS)



PHGv2 > 5 million segregating variants







Expanding Wheat PanGenome for mapping regulatory variants

Genome assembly statistics (HiFi data)

Line	Contigs			
	Total length (bp)	#	N50 (bp)	
GF150	14,485,612,752.00	4,249	14,547,818.00	
GF197	14,441,884,101.00	4,747	9,831,389.00	
GF263	14,356,511,097.00	8,892	3,909,479.00	
GF304	14,513,435,385.00	4,169	17,171,782.00	
GF313	14,591,017,792.00	4,277	17,385,993.00	
GF336	14,465,408,968.00	3,654	17,467,334.00	
GF361	14,395,750,022.00	9,698	3,219,032.00	
GF371	14,477,383,491.00	4,174	9,462,587.00	
GF372	14,486,991,796.00	3,611	14,283,023.00	
GF389	14,471,573,713.00	3,353	16,790,708.00	





Expanding Wheat PanGenome for mapping regulatory variants

Number of genes predicted by Liftoff

Lino	Total # genes		
Line	TFMM	TFPM	
GF150	109,172	96,563	
GF197	105,840	94,295	
GF263	104,875	93,651	
GF304	106,294	94,627	
GF313	107,993	95,592	
GF336	106,166	94,403	
GF361	105,612	94,078	
GF371	106,912	94,939	
GF372	106,562	94,660	
GF389	107,191	95,211	
CS_2.1*	106,313	94,904	
Kariega*	113,879	100,173	

Number of predicted NLRs

Lino	Total		
Line	Complete	Partial	
GF150	2641	911	
GF313	2631	858	
GF336	2550	831	
GF361	2281	801	
GF371	2614	857	
GF372	2576	903	
CS_2.1*	2559	899	
Kariega*	2602	902	
Fielder*	2548	868	
Norin61*	2643	935	
Julius*	2496	883	
Mace*	2493	957	
Arina*	2497	1015	
Jagger*	2487	952	
Lancer*	2424	1043	
Landmark*	2466	890	







Assembled genomes capture broad range of structural diversity



ted Agricultural P

ATE

Acknowledgements

This research was supported by the Agriculture and Food Research Initiative Competitive Grants 2022-68013-36439 (WheatCAP) and 2019-67013-29017 from the USDA National Institute of Food and Agriculture and grant INV-004430 from Bill and Melinda Gates Foundation.



United States Department of Agriculture National Institute of Food and Agriculture BILL& MELINDA GATES foundation