Sequencing 4.3 Million Mutations in Wheat Promoters to Understand and Modify Gene Expression

Junli Zhang

Dubcovsky Lab

IWGSC Workshop @PAG31, 2024-01-13



Outline

I. Exome and promoter capture sequencing of wheat TILLING populations

II. A new regulatory capture design for wheat

Exome capture sequencing of wheat TILLING population

Uncovering hidden variation in polyploid wheat

Andrew L. Phillips^e, Cristobal Uauy^{d,1,2}, and Jorge Dubcovsky^{a,f,1,2}

Ksenia V. Krasileva^{a.b.c}, Hans A. Vasquez-Gross^a, Tyson Howell^a, Paul Bailey^c, Francine Paraiso^a, Leah Clissold^c, James Simmonds^d, Ricardo H. Ramirez-Gonzalez^{c,d}, Xiaodong Wang^a, Philippa Borrill^d, Christine Fosker^c, Sarah Ayling^c,

The wheat genome is too big for whole genome sequencing of mutant populations (genes = 2%)







Sequencing 4.3 million mutations in wheat promoters to understand and modify gene expression

Junli Zhang^{a,1}, Hongchun Xiong^{a,b,1}, Germán F. Burguener^{a,c}, Hans Vasquez-Gross^{a,d}, Qiujie Liu^{a,c}, Juan M. Debernardi^{a,c}, Alina Akhunova^e, Kimberly Garland-Campbell^f, Shahryar F. Kianian^g, Gina Brown-Guedira^h, Curtis Pozniakⁱ, Justin D. Faris^j, Eduard Akhunov^e, and Jorge Dubcovsky^{a,c,2}

Contributed by Jorge Dubcovsky; received April 24, 2023; accepted August 3, 2023; reviewed by Graham Moore, Thorsten Schnurbusch, and Ben Trevaskis

Remapping exome capture to RefSeq v1.1



Resequence Kronos TILLING mutants with promoter capture

- *Cis*-regulatory variants often cause subtle phenotypic change
 - by modifying the timing, pattern, or level of gene expression (Wittkopp and Kalay, 2011)
- Dr. Anthony Hall's lab developed one global wheat promoter capture assay covering 2 Kb promoter region of all high-confidence annotated genes (Gardiner et al., 2019).





Konishi et al., 2006

A SNP in the 5' regulatory region of the *qSH1* gene caused loss of seed shattering due to loss of expression at at the provisional abscission layer

Promoter capture summary

Summary	Promoter	Exome
Number of lines used (overlap 1,465)	1,513	1,521
Uniquely mapped EMS-type mutations	4.3 M	4.7 M
Avg. EMS-type mutations per kb (population)	42	36
Avg. EMS-type mutations/line	2.834	3.084
Mapping space	102 Mb	131 Mb



Distribution of mutation types in 15 mutants

Data availability

- The data is now public available
 - Dubcovsky Lab website: https://dubcovskylab.ucdavis.edu/wheat_blast
 - GrainGenes genome browser: <u>https://wheat.pw.usda.gov/jb?data=/ggds/whe-iwgsc2018</u>
 - Next release of Ensembl Plants
- Kronos mutant seeds can be requested from the Germplasm Resources Unit at the John Innes Centre and from Dr. Dubcovsky Laboratory Tilling Distribution

https://dubcovskylab.ucdavis.edu/wheat-tilling.

Use the promoter capture data

Promoters are still not well annotated, so users need to find the conserved promoter regions

- Comparisons of promoter region among grass species to identify conserved regions
- Align promoter region (include about 50 bp of the exons)
- Within conserved regions focus on open chromatin regions (ATAC-seq)



VRN1 Promoter

			SPL						
FUL2 VRN1 Barley Brachy Oryza Zea	GTGAAGAGGGGCTGGCCT TGATTTGGGGAAAGCAAA TGATTTGGGGAAAGCAAT TTAAATGAGGAAAGCGTT CCAAAAAAGGATCGGCCC TCATGCCAGGCAAGGTGT	GC GTT - TGG AT CGG - GAG AT GGGGGAG AT CGG - GAG AG AAA - AGC GTGAGAGAAAGC - CGG	GTTC GACCGTACGA ATT GCACGTACGA ATT GCGCGTACGA ATT GGGCGTACGA CCACGAGAGTACGA CTTCGACCGTACGA	GTGCAGACCCATCACAC CGTTCGACACGTCGACG CGTCCGACACGTCGACA CGGTCGACACGTCGACA GCACGCACGCGTCGACA GCACGCACGCGTCGCCG GAAAGACCCTG	A TOGGAAGCGGGC CCCGCGG CCGGCGG CGGGCCG GGGCCGG GAGCCGG CAGCGGC CAGCGGC	GATGGATGG G G G G C	AGGAGTTCAACCT	GCGCTGCCATTTTTTTATG	GCTTGCCGCCGC CCCGCGG CCCGCGG GCCCGGG GCCCGGG
				•					
			3 - C		L	_FY	ARF	?	
FUL2 VRN1 Barley Brachy Oryza Zea	TGCTGCTACCACTATAAA TGGGGCATCGTGTGGCTG TGGGGCATCGTGTGGCTG CAGGCCATCGTGTGGCTC C TGAGAGATGCCGT	ICACAATAGAGAGAGAGAGA ICAGGAC ICAGGAC ICAGGAC IATGGTC IGGGGCC	GAGAGACTGAGACT	AGCGTGTGGGGGCCATTT CGCGGGGGCCCCGCG CGCGGGGGCCCCGCG CACTGGGCCCCGGG AGGTGGGCCCCGGGC	T AGGCGTCTCA CC G GGCGGGCCGG CC G CGCGGGCCGG CC G CGCGGCCTGA CC G CGCGCCTGA CC A CGCCGCATCG CC	CAAT AGACG CAATGGGTGA CAATGGGTGA CAATGGGTGA CAATGGCCG CAATGCCAG	TCGACAGC TCGACAGC GCTA TCGACAGC GGACA TCGACAGC GGACA TCGACAGC GGAGC TCGACAGC GGCGG	ATGCTCCAGACCAGCCCGGT ATGCTCCAGACCAGCCCGGT ATGCTCCAGACCAGCCCGGT SCGGTCGAAACCACCCCGGT 5CGGGC-AAACCATCCCGGT	ATTGC - ATACCG ATTGC - ATACCG ATTGC - ATACCG ATCGACATACCA ATCGC - AAACC TTCGC - TATACC
				VRN box	Contraction ()			CArG box	
FUL2 VRN1 Barley Brachy Oryza Zea	CGGC - C- CGCT - C- CGCT - C- CGCT - C- GCCTCCTCCTCCTCCGAC CCCT - C-	GGAAC - CAC GGGGC - CAG GGGGC - CAG CGTCCTCAGA GATCCGGCGC - GCGC - CCCCT - CCGA	CTCG TCC	ATCGC A CTITAAAAA CTITAAAAA CTITAAAAA CTITAAAAA CTITAAAAAA CTITAAAAAA CTITAAAAAAAAAA	C CCTCCC CCA C CCCTCC CCD C CCCTCCCATCGO C CCCCCCATCGO C CCCCCA TTT C CCCCCA TTT T CCGCCT CCT	TTACAC CTGCCC CTGCCC TTGCCC CTCCTCCTC CCCAGCC	CACAAATCCATCO GAACCCTCGTTTTO GAACCCTCGTTTO GAACCCTCGTTO CCTCCTCCTCGTTO TCTCCCCTCCTCCTCC	ATTECTETTECTETECCEATE GGECTGGCCATCCTCCC GGECTGGCCATCCTCCC GGECTGTCC	TTTAAAGCTGCA TCTCCTCC TATCCCCT GACCCACT GTCGTCTC CCTCACCT
FUL2 VRN1 Barley Brachy Oryza Zea	TCCCATCTCCC CCTCTCTTCCA CTCCCCTCCTCCCA CTCCCGCCTCCC CCCCCACTCGAT - CGATCCA CCCCCCCCCCTATCTCTCCA	ATCCATCGATCGATCC ACCGCCGCAGCTAGCT	GTCCCCCCCATCGC GCGACGTCATGCAC	GCGCCGACGCATTCCGCCC TCGCCGGCGCCACCGCC	S CCGTCTCGCCGTG A CCGCATACT	-TC -CCT -CTT TTC -TTC -TCCACGTGA	TGGGGGCCGGGGC	AP2 site	CAGCAGCCACCA
FUL2 VRN1 Barley Brachy Oryza Zea	ACCGCCGCGGCGAGC CACCCAACCACCTGA CACCCAACCACCTGA CCCCTAATCCGCCTA CCACCAGTAGTTGCCGTG TACCTATCCTGTGTCAAG	CAGCGACGTCTCGCCT TAGCCATGGCTCCGCC CAGCCATGGCTCCGCC ATGGCACAGGCT TGGGGATAGGT CAAGCCTCTCG	CCTCCCACCCCGG GCCTCGCCTCCGCC ACCTCGCCTCCGCC	CCGCCGTGCGCGCCA TGCGCCAGTCGG AG CGCGCCTCTGGG AG CCACGGCGGG AG - GTGGCTATAGG GC - CA	T AGACACAAGAGCC T AGCCGTCGCGGGTC T AGCCGTCGCGGGTC T AACCGTCGCGCGGTC T AGTGGTCGCGCGC C AACAAGGAAGGA	CG ACAGC CTCCCGGTGT CGCTCGCTCG CGCTGGCTGG TGATAGCGAG AGCTAGCTAG	TAGATCG TGGAGGGTAG CTCGCTCGCT CTTGTTGCGG GTGGGTAGGG TTTTATAGCTGCT	GTCGGCGGCGGCGGCTGAA	GCGACGTGCCTG
						€ VRN	l promot	er	
FUL2 VRN1 Barley Brachy Oryza Zea	ATTTTGGTTGGAGGTAGA AGCTAGGATTTAG	GAGAGAGAGAGAGGG	AGGGAGGGAGGAGGA	CGAAAT 	CCCGGTTCTCGAGC CCCGGTCCTCGAGC TCTCGATTCCGGGA A GGAAGAACAGGAG A GGGAGAAAAGGCGGC	AGGA ATG G CGGA ATG G CGGA ATG G AGGA ATG G SGAA ATG G CGGC ATG G	TCGCGGCAAGGTG GCGCGGGAAGGTG GCGCGGGAAGGTG GCGCGGGAAGGTG GCGCGGGCAAGGTG GCGCGGGAAGGTG	CAGCTGAAGCGGATAGAGA CAGCTGAAGCGGATCGAGA CAGCTGAAGCGGATCGAGA CAGCTGAAGCGGATCGAGA CAGCTGAAGCGGATCGAGA CAGCTGAAGCGGATCGAGA	ACAAGATAAATC ACAAGATCAACC ACAAGATCAACC ACAAGATCAACC ACAAGATCAACC ACAAGATCAACC ACAAGATCAACC

VRN1 Promoter

Milter tracks 0 50,000,000 100,000,000 150,000,000 200,000,000 250,000,000 300,000,000 3	50,000,000 400,000,0
• Reference sequence 1 $\bigcirc \bigcirc $	5A ▼ chr5A:58742125658
✓ IWGSC RefSeq v1.0 Pseudomolecules	
IWGSC Official Annotations 14 IWGSC Official Annotations 14 Zoom in to see sequence Zoom in to see sequence	Zoom in to see sequence
GrainGenes Curated Tracks 4	
The Triticeae Toolbox (T3) Annotations	
▶ 1000 Wheat Exomes 3 Trace CS5A02G391700.2	
Ensembl	
HAPMAP 1	30 -
WHEALBI Z C Leaf protoplast ATACseq	30
✓ EMS Mutations in Kronos (Dubcovsky Lab) 4	20 -
▼ Promoter Capture 2	20
	10
High-confidence SNPs	
	E o
▼ Exome Capture 2	90 ¬
S Root ATACseq fresh	80 -
Exome data with annotated variant	70 - 60 -
	50 -
	40 - 30 -
	20 -
▶ G-quadruplexes 1	
MNase Chromatin States	
▶ 10+ Wheat Genome Project 3 High-confidence SNPs	
Expression QTLs 3 SNV C-> T SNV C-> T	V C -> T
▼ ATACseq (Dubcovsky Lab) 2	
Leaf protoplast ATACseq Kronos2306:G587423366A Kronos587:C587424282T	Kronos3160:G587425167A
▼ Root ATACseq fresh	SINV G -> A
Kronos2381:G587423511A Kronos3669:C587424394T	
SNV G -> A SNV C -> T	
Predicted Protein 3D Structures	,
Varietal SNPs (Akhunov & Dubcovsky Labs) OT Kronos2944:C5874233151 Kronos2957:G5874244444 SNV C -> T SNV G -> A	`
Kronos2597:C5874235271 Kronos1146:G5874244457 SNV C -> T SNV G -> A	۱.
Kronos4414:C587423551T Kronos3622:G587424462 SNV C -> T SNV G -> A	4
Kranap564,0597422596A	т
SNV G -> A SNV C -> T	10.
CDI Kronos4679:G587423608A Kronos2157:C587424	556T
JFL SNVG->A SNVC->T	

VRN1 Promoter



A second-generation capture panel for cost-effective sequencing of genome regulatory regions in wheat and relatives

Junli Zhang, Juan M. Debernardi, Germán F. Burguener, Frédéric Choulet, Etienne Paux, Lauren O'Connor, Jacob Enk, Jorge Dubcovsky 🔀

First published: 09 December 2022 | https://doi.org/10.1002/tpg2.20296

A new regulatory capture is needed

- The Gardiner promoter capture assay has some limitations
 - Enrichment efficiency is low: 8x to 32x
 - Still have some repetitive regions
 - Only covered 2 Kb of promoter region: ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) revealed more regulatory regions outside the promoter region
- Gardiner probe set has been discontinued.

ATAC-seq data revealed more regulatory regions

A new improved and expanded regulatory capture design

Durum wheat Kronos Supplement RefSeq v1.0

2-Kb upstream of all highconfidence genes from Gardiner et al. (2019) **ATAC-seq** Supplement

Capture performance evaluation

Comparison between Gardiner and Arbor probes

- 24 Kronos EMSmutagenized lines
- Captured with either Gardiner or Arbor probes
- Both were down-sampled to 40-M pairs

Arbor probes showed significant increases in **specificity** and **coverage** in the target region.

Summary

- Promoter and exome capture of Kronos EMS mutants
 - 4.3 M SNPs in the promoter and 4.7 M in the coding region
 - Comparisons of promoter region among grass species to identify conserved regions
- New regulatory capture
 - Expanded with open chromatin regions from new ATAC-seq data.
 - Increased specificity and coverage in the target region.
 - Versatile and cost-effective for wheat and related *Triticeae* species.

Acknowledgements

Kronos TILLING promoter capture

University of California, Davis Howard Hughes Medical Institute Kansas State University University of Minnesota USDA-ARS Chinese Academy of Agricultural Sciences University of Saskatchewan

Howard Hughes Medical Institute

United States Department of Agriculture National Institute of Food and Agriculture

International Atomic Energy Agency

New regulatory capture design

UC Davis UCA, INRAE, GDEC, France VetAgro Sup, Lempdes, France Daicel Arbor Biosciences

