Challenges and opportunities in positional cloning and structural variation in polyploid crops

Ellie Taagen, PhD Candidate, Dr. Mark Sorrells Lab

Cornell University, Section of Plant Breeding and Genetics

IWGSC webinar, November 18th, 2021

@etaagen

Outline

- **How IWGSC tools enable positional cloning**
- Case study: Positional cloning overpowered by structural variation
- Challenges and opportunities for polyploid crops
- GitHub learning resource

Received: 25 January 2021 | Accepted: 19 April 2021
DOI: 10.1002/tpg2.20106

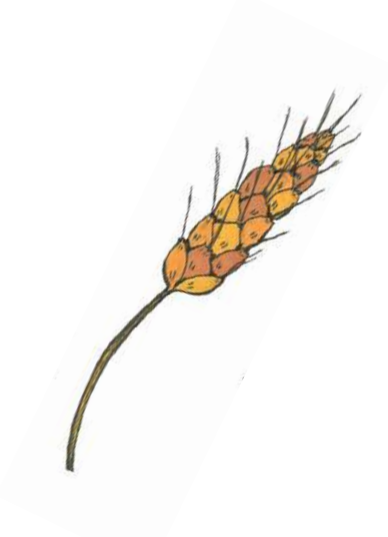
The Plant Genome

ORIGINAL RESEARCH

Positional-based cloning 'fail-safe' approach is overpowered by wheat chromosome structural variation

Ella Taagen[1] | James Tanaka[1] | Alvina Gul[2] | Mark E. Sorrells[1]

Yield improvement



grains per spike

Causal variant id

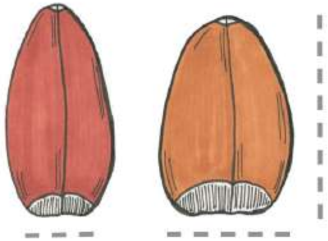
Trait balancing act

Genetic gain

polyploidy

spikes per area^2

weight per grain



Roadmap to positional cloning

Fine-mapping

Reference genome

Sequencing

Functional validation

Genomic structural variations are
common among polyploids

Outline

- How IWGSC tools enable positional cloning
- **Case study: Positional cloning overpowered by structural variation**
- Challenges and opportunities for polyploid crops
- GitHub learning resource

Received: 25 January 2021 | Accepted: 19 April 2021
DOI: 10.1002/tpg2.20106

ORIGINAL RESEARCH

Positional-based cloning 'fail-safe' approach is overpowered by wheat chromosome structural variation

Ella Taagen[1] | James Tanaka[1] | Alvina Gul[2] | Mark E. Sorrells[1]

QTL on chromosome 5A is associated with increased grain weight

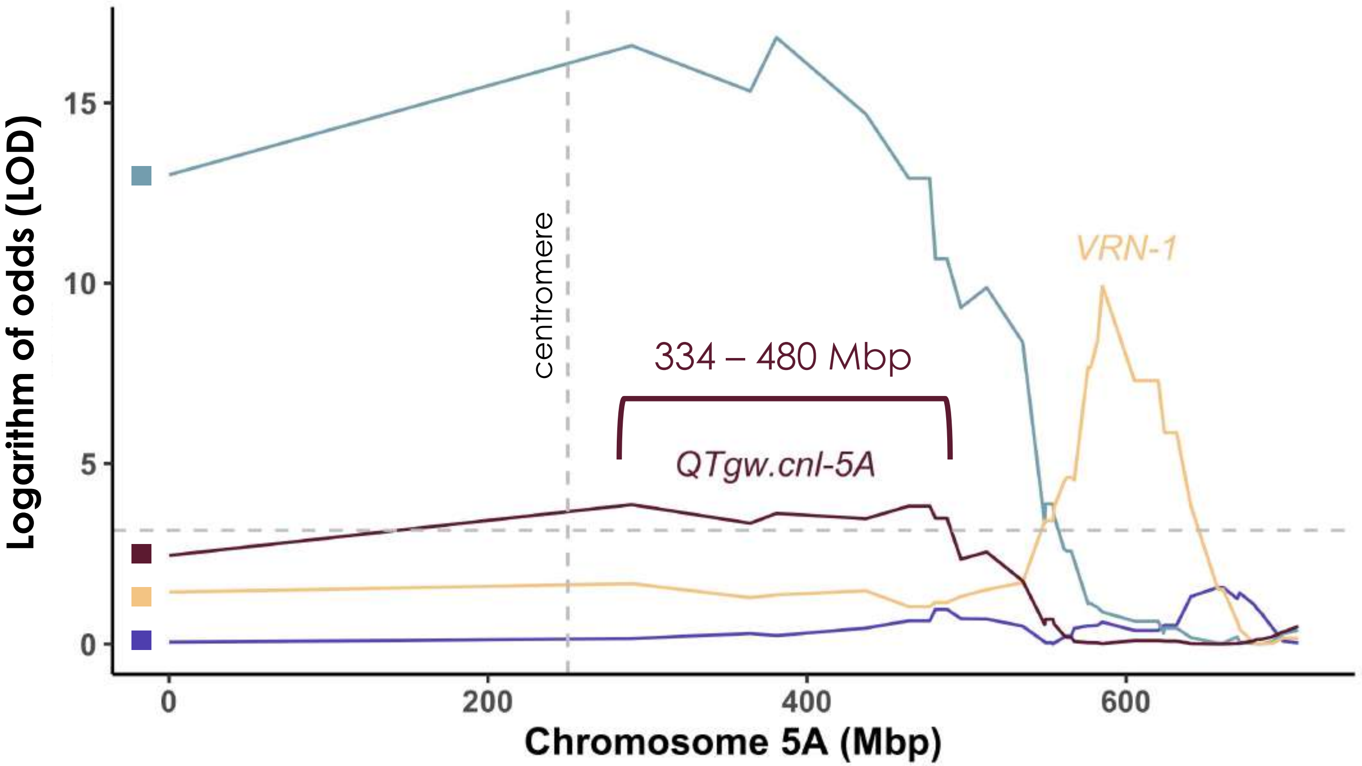
Population:

Synthetic W7984 x Opata M85

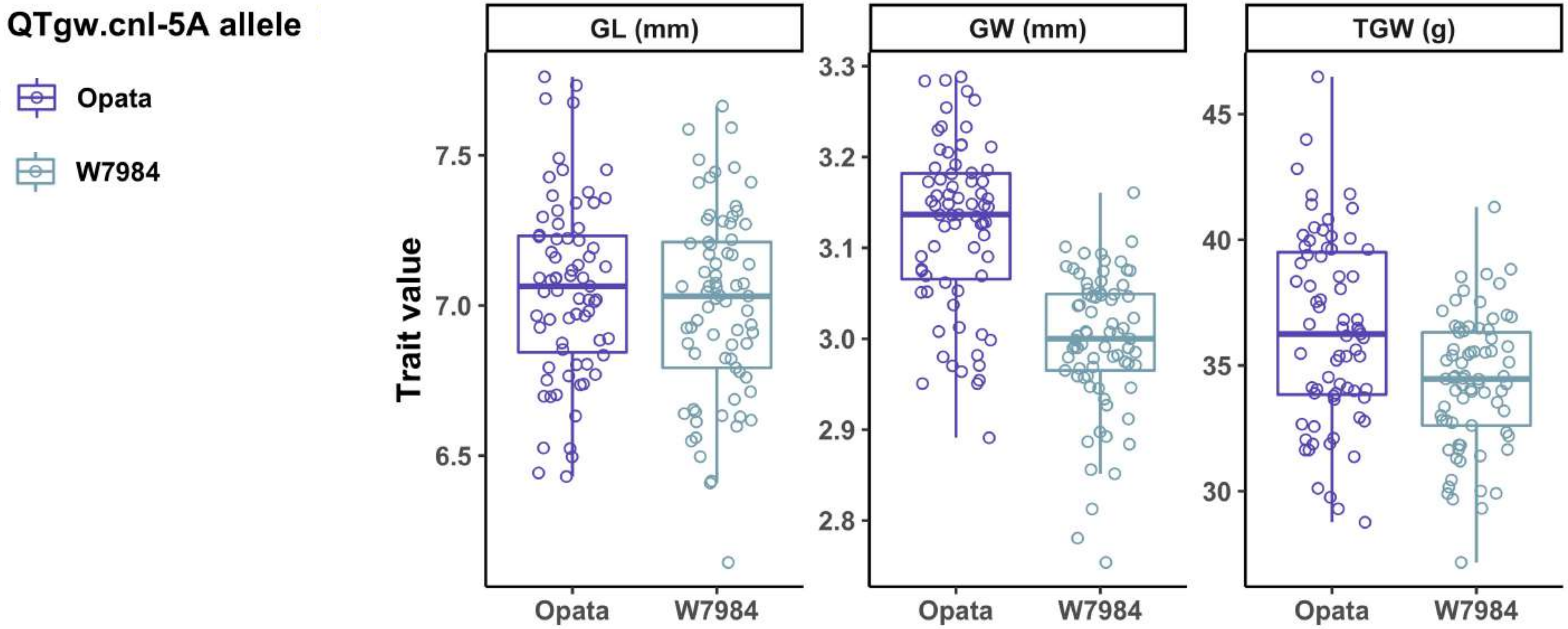
162 doubled-haploid lines, spring

4 year-environment observations (BLUP)

- 🟦 Grain width (GW)
- 🟫 1000-grain weight (TGW)
- 🟧 Heading date (HD)
- 🟪 Grain length (GL)



Opata provides the increased GW and TGW allele

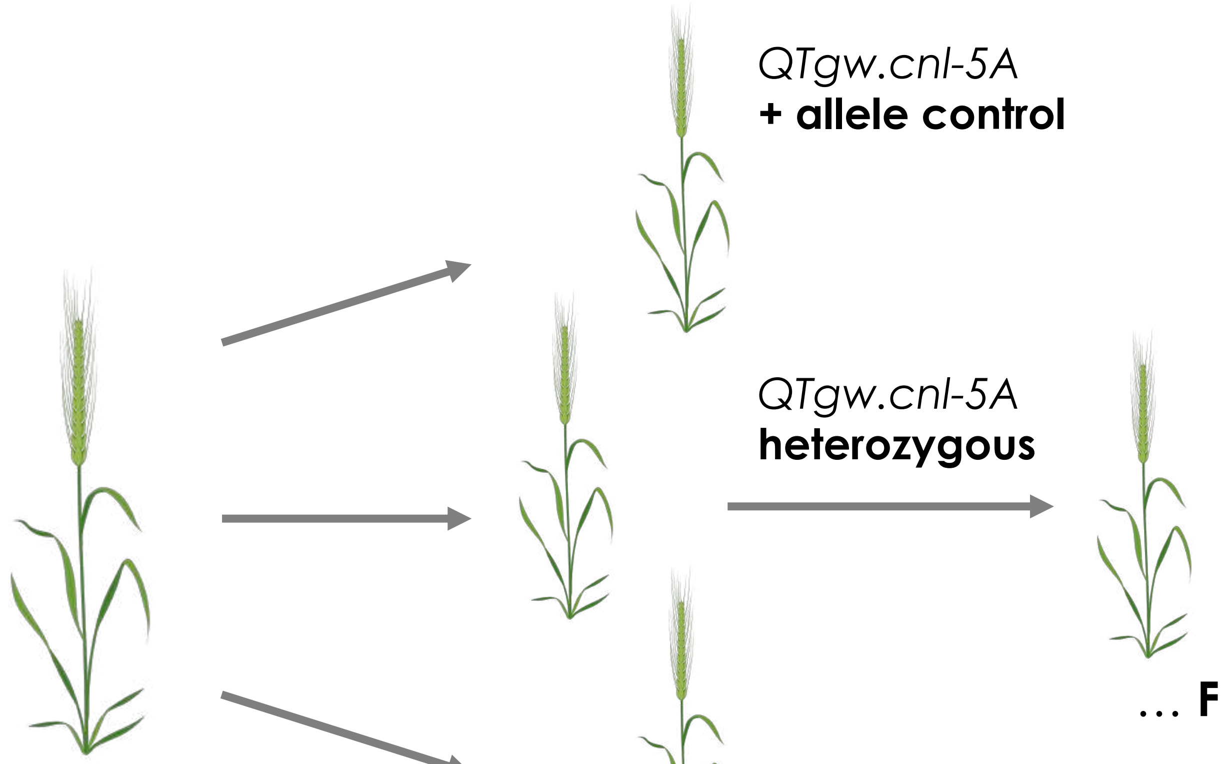


Trait value gained (%)	0.92^{ns}	4.2 ***	6.4 ***
H^2	0.75	0.81	0.68

p-value ns: non-significant, ***: $p \leq 0.001$

Fine-mapping population development

*Heterogenous inbred family (**HIF**)*



QTgw.cnl-5A
+ allele control

QTgw.cnl-5A
heterozygous

… **$F_{6:5}$**

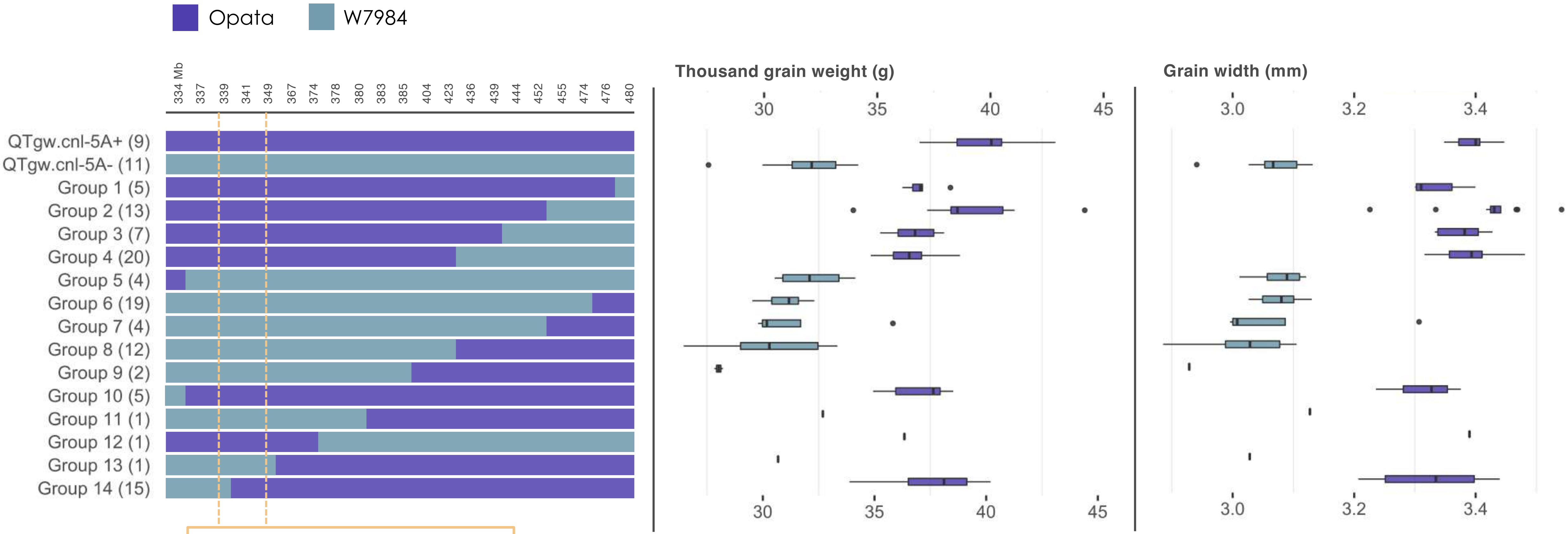
QTgw.cnl-5A
W7984 x Opata RILs $F_{6:1}$

QTgw.cnl-5A
- allele control

Across **5** generations, screened **> 5,000** progeny for *QTgw.cnl-5A* recombination

Fine-mapping population development (HIFs)

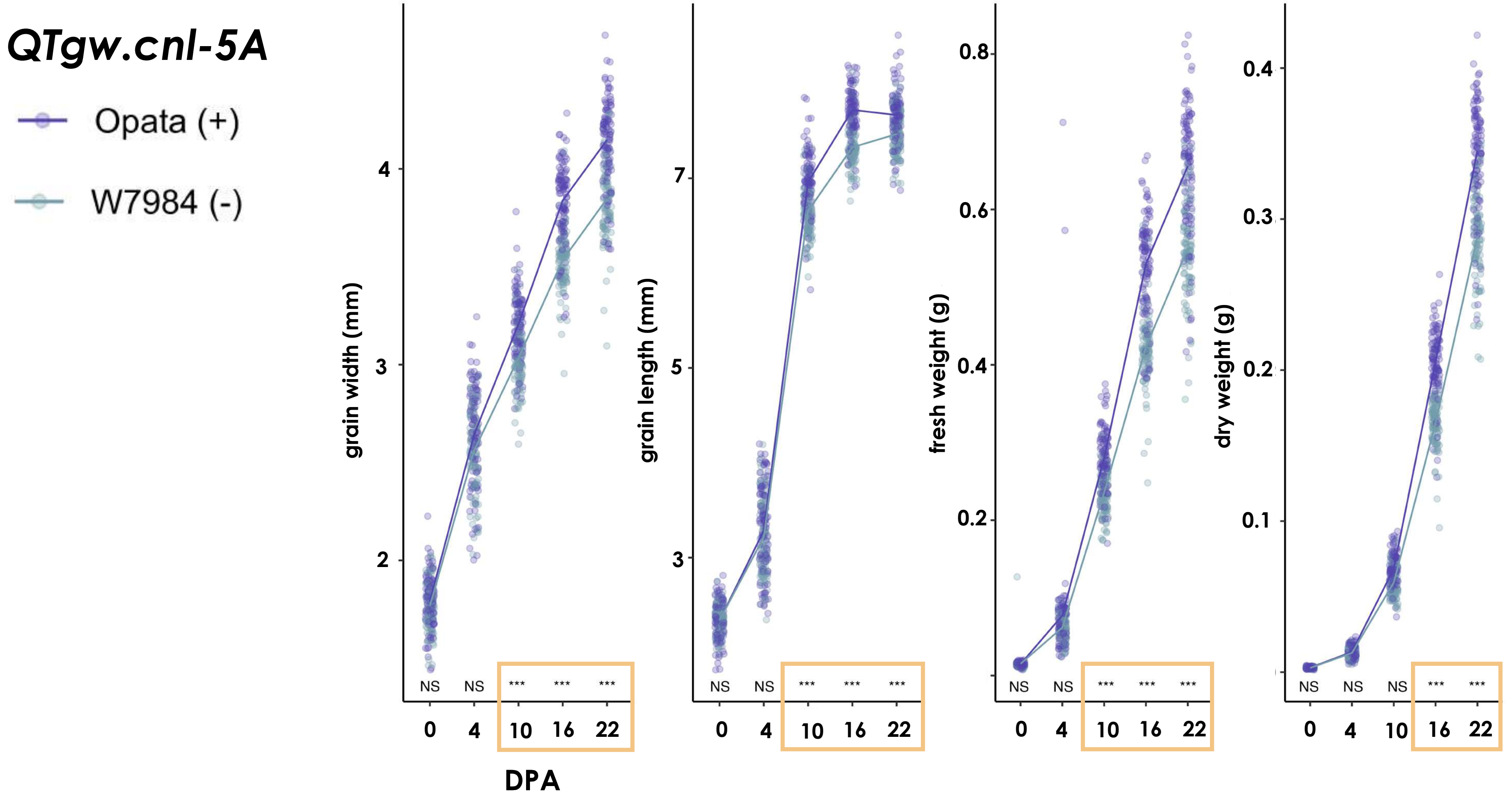
Reduce to **109** lines with *QTgw.cnl-5A* recombination



10 Mbp
37 high confidence genes

QTgw.cnl-5A+ **21.3%** heavier and **10.3%** wider grains than *QTgw.cnl-5A-*

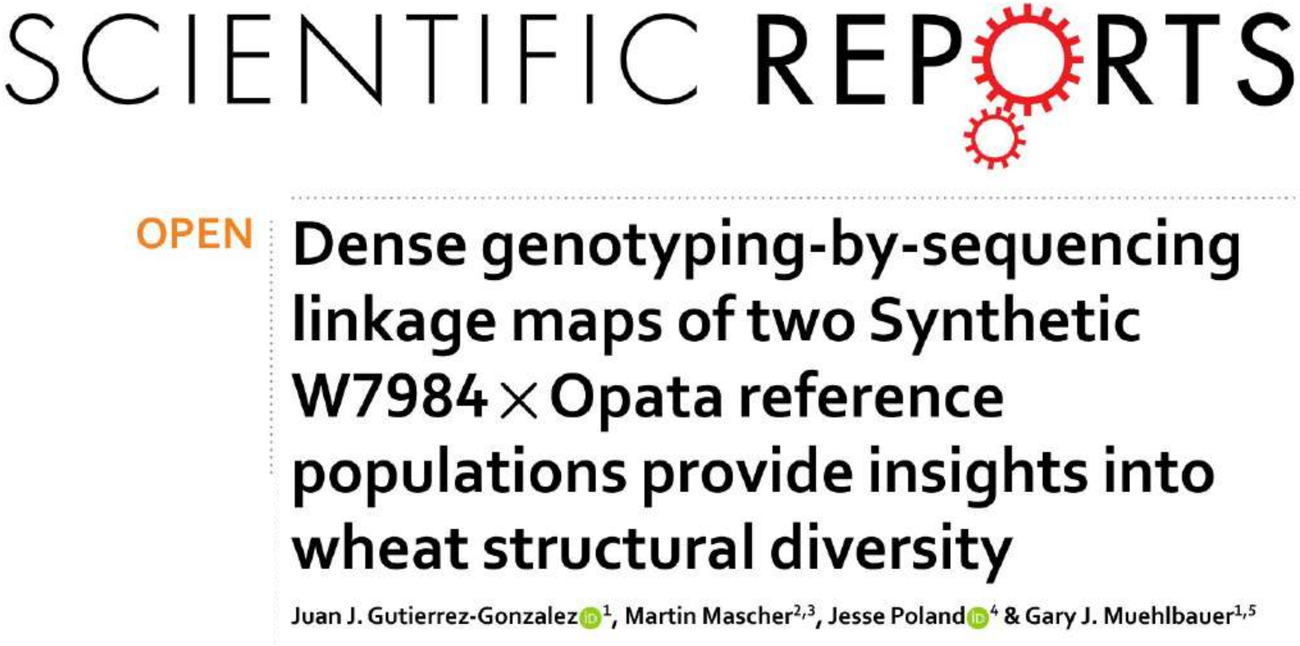
Variation in grain weight and morphology significantly associated with early grain development



Bonferroni corrected p-value NS: non-significant, ***: p ≤ 0.001

No difference in grain fill duration

Gutierrez-Gonzalez et al., 2019:


SCIENTIFIC REPORTS

OPEN

Dense genotyping-by-sequencing linkage maps of two Synthetic W7984 × Opata reference populations provide insights into wheat structural diversity

Juan J. Gutierrez-Gonzalez[1], Martin Mascher[2,3], Jesse Poland[4] & Gary J. Muehlbauer[1,5]

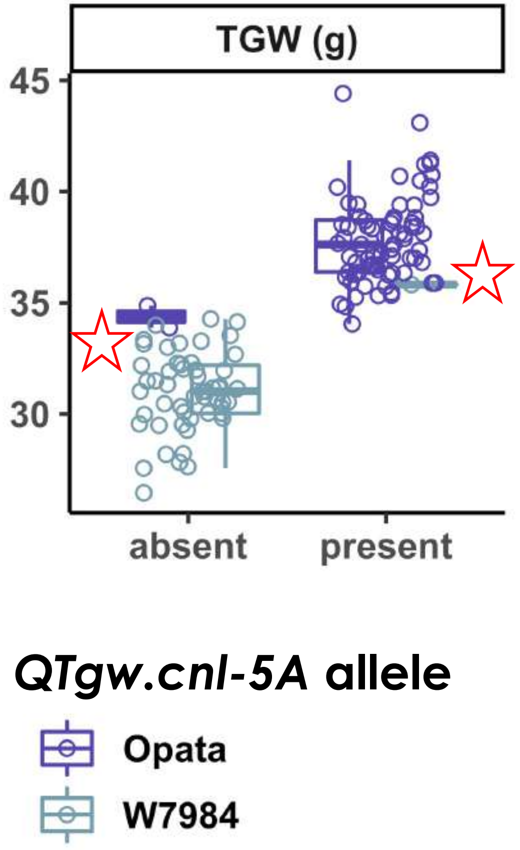
"We found that chromosome arm 5AS is ***missing*** *from W7984."*

QTgw.cnl-5A is in linkage with chromosome arm 5AS structural variation

	QTgw.cnl-5A **HIFs**	***QTgw.cnl-5A+*** **Opata**	***QTgw.cnl-5A-*** **W7984**
Chr arm 5AS, absent	54	**2**	52
Chr arm 5AS, present	73	72	**1**

Correlation coefficient **0.91**

No significant interaction



We used **RNA-seq** to investigate differentially expressed genes on chromosome arm 5AS and in the *QTgw.cnl-5A* candidate region

Chromosome arm 5AS present : *QTgw.cnl-5A+*

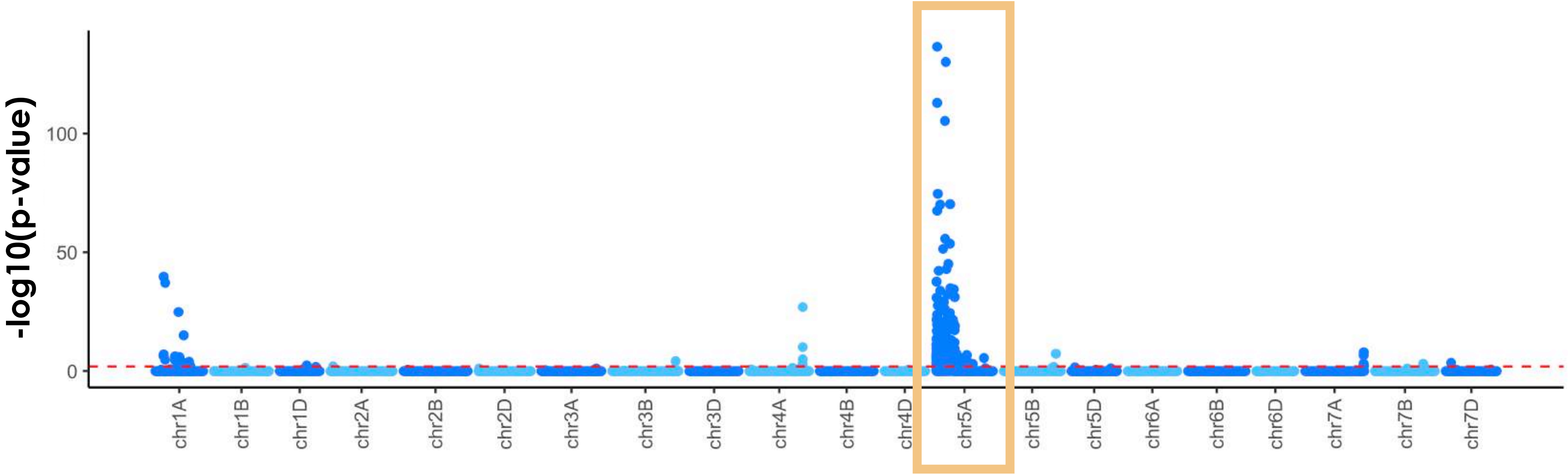
vs

Chromosome arm 5AS absent : *QTgw.cnl-5A-*

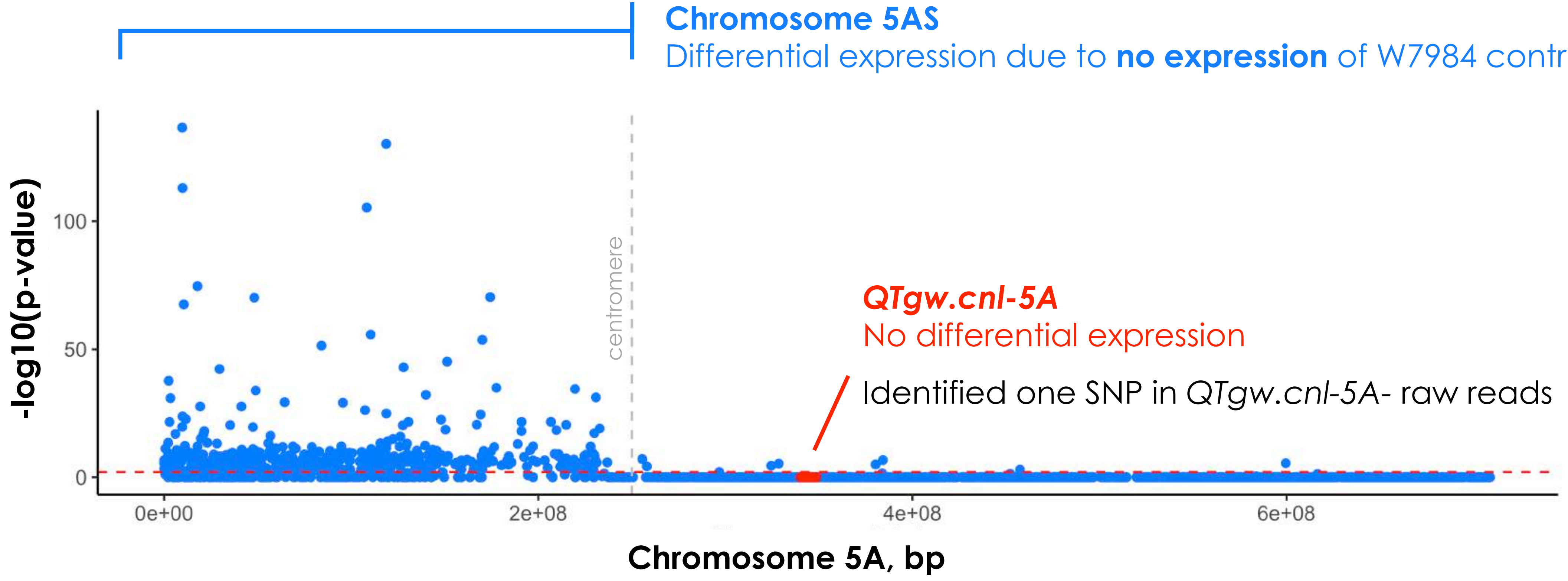
RNA-seq of *QTgw.cnl-5A* confirmed the significance of chromosome arm 5AS structural variation

535 of the 556 differentially expressed genes are on chromosome 5A

None of the homoeologous copies were differentially expressed



RNA-seq of HIFs confirmed the significance of chromosome arm 5AS structural variation

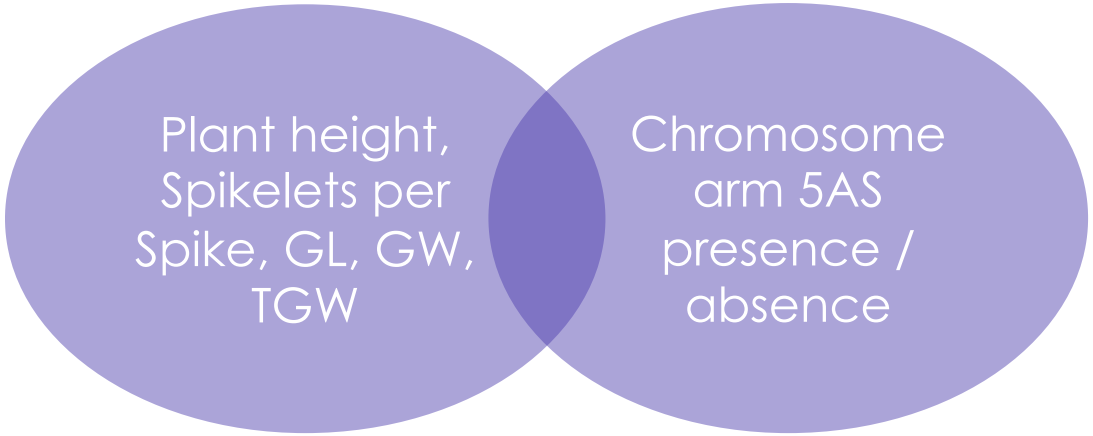


Chromosome 5AS
Differential expression due to **no expression** of W7984 control

QTgw.cnl-5A
No differential expression

Identified one SNP in *QTgw.cnl-5A*- raw reads

Chromosome arm 5AS differential expression may provide insight into candidate genes



Plant height, Spikelets per Spike, GL, GW, TGW

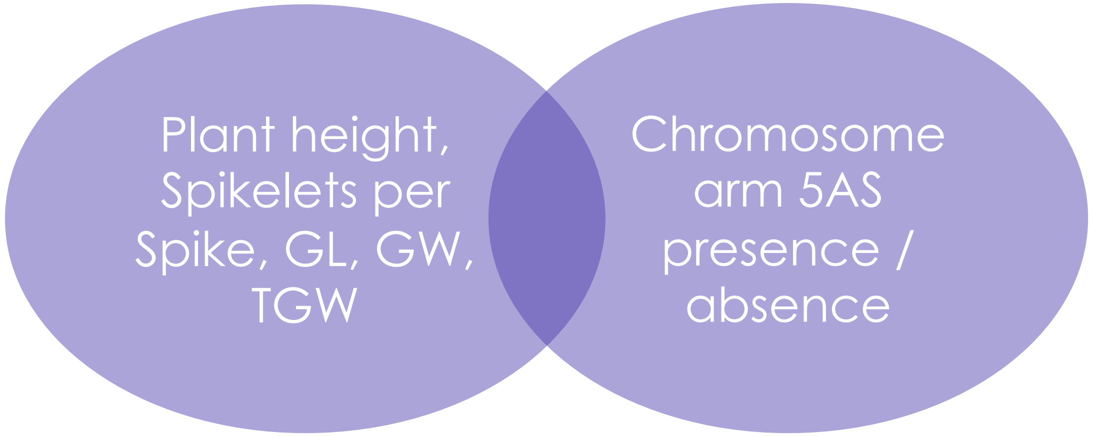
Chromosome arm 5AS presence / absence

Masked variation

Lack of significantly enriched gene ontology (GO) terms

Phenotypic distribution characteristic of single gene, rather than entire chromosome arm

Chromosome arm 5AS differential expression may provide insight into candidate genes



Plant height, Spikelets per Spike, GL, GW, TGW

Chromosome arm 5AS presence / absence

292 differentially expressed genes with GO terms related spike and early grain growth

Identified **9 candidate genes** on chromosome arm 5AS that may impact yield components

Hypothesis: masked variation of homoeolog dosage-dependent or nonadditive functionally redundant genes

Outline

- How IWGSC tools enable positional cloning
- Case study: Positional cloning overpowered by structural variation
- **Challenges and opportunities for polyploid crops**
- GitHub learning resource

Received: 25 January 2021 | Accepted: 19 April 2021

DOI: 10.1002/tpg2.20106

The Plant Genome

ORIGINAL RESEARCH

Positional-based cloning 'fail-safe' approach is overpowered by wheat chromosome structural variation

Ella Taagen[1] | James Tanaka[1] | Alvina Gul[2] | Mark E. Sorrells[1]

Detecting chromosome structural variation

Early 1990s

W7984 x Opata original crosses

2011

Reconstructed and expanded

Poland et al., 2012

W7984 x Opata DH high-density GBS genetic map

2018

No site-specific polymorphisms on 5AS

Gutierrez-Gonzalez et al., 2019

dense GBS linkage map, 5AS structural variation

Brinton et al., 2020 and *Walkowiak et al., 2020,* underscore the structural diversity of wheat

Polyploid positional cloning recommendations

1. Move from a SNP to haplotype-based approach to identify genetic diversity
2. Invest in sequencing to detect structural variants
3. Use the transcriptome to identify candidate genes
4. Traits with broad overlapping QTL may not be pleiotropic

Outline

- How IWGSC tools enable positional cloning
- Case study: Positional cloning overpowered by structural variation
- Challenges and opportunities for polyploid crops
- **GitHub learning resource**

Received: 25 January 2021 | Accepted: 19 April 2021

DOI: 10.1002/tpg2.20106

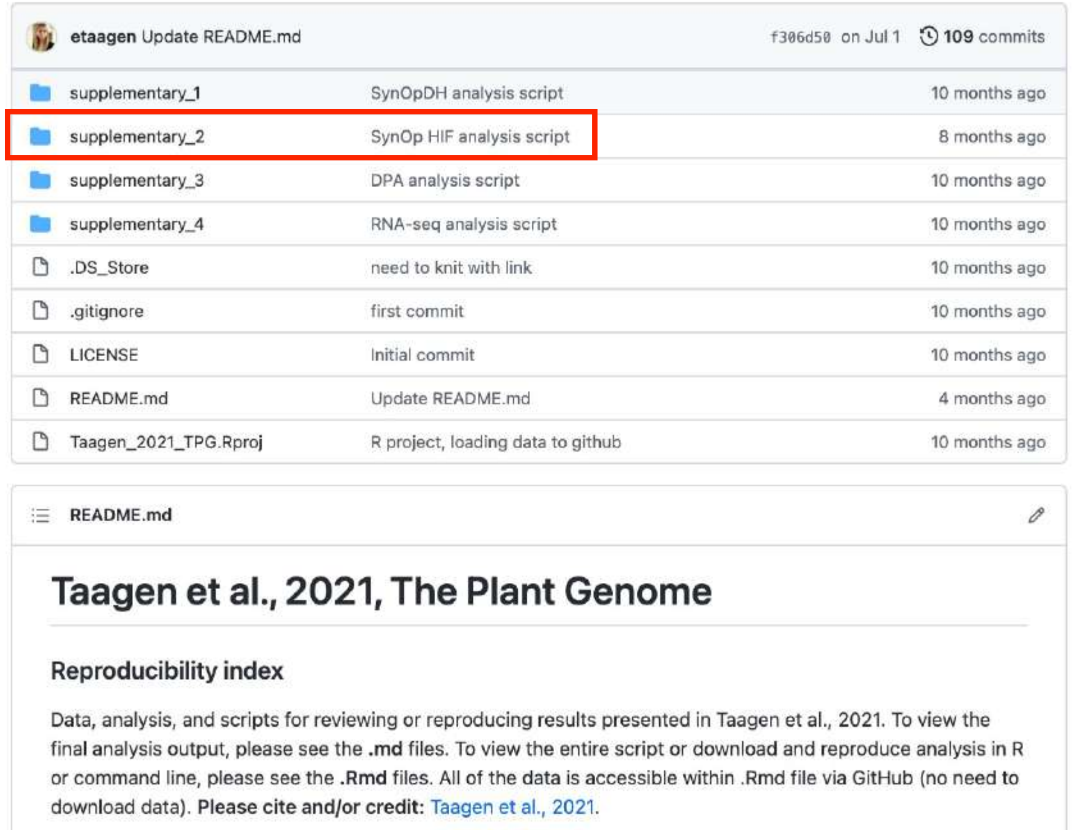
The Plant Genome

ORIGINAL RESEARCH

Positional-based cloning 'fail-safe' approach is overpowered by wheat chromosome structural variation

Ella Taagen[1] | James Tanaka[1] | Alvina Gul[2] | Mark E. Sorrells[1]

github.com/etaagen/Taagen_2021_TPG



etaagen Update README.md f306d50 on Jul 1 **109** commits

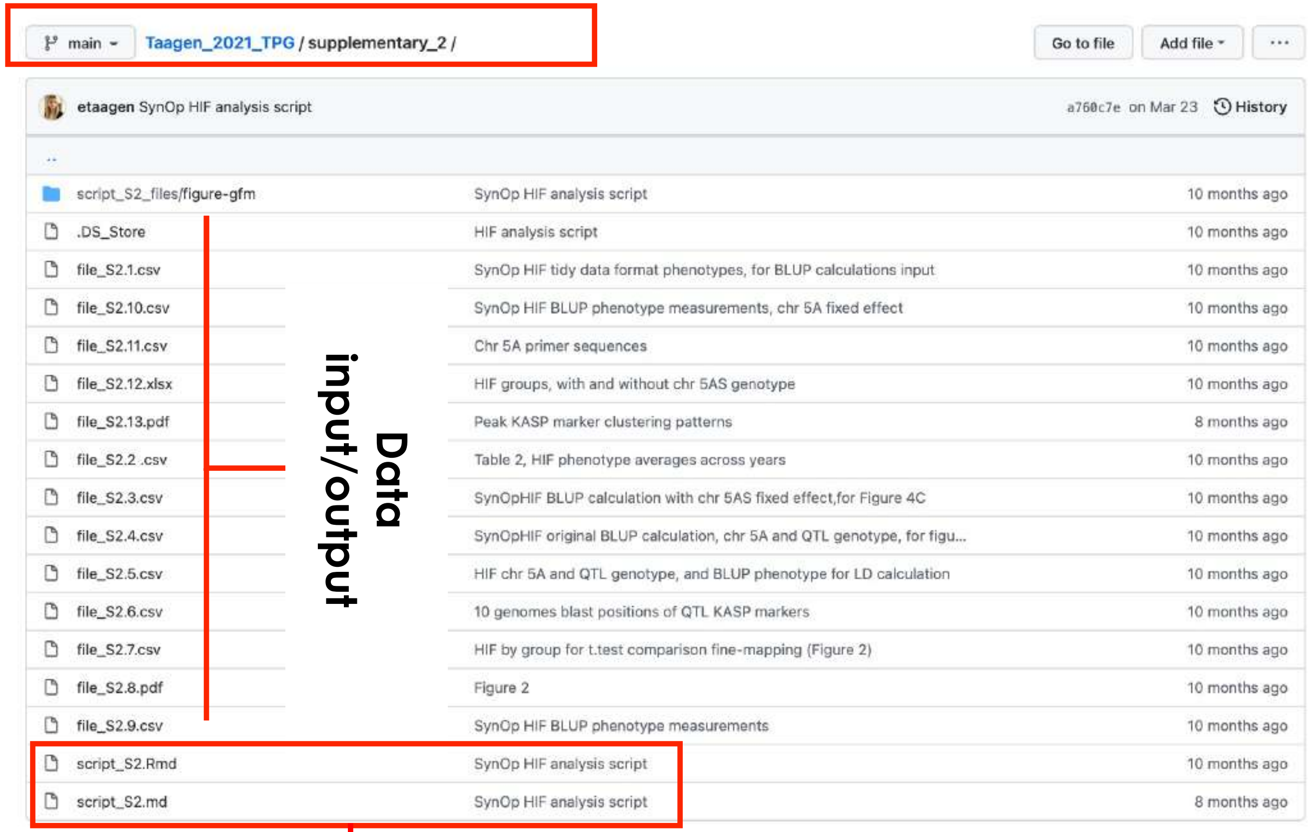
supplementary_1	SynOpDH analysis script	10 months ago
supplementary_2	SynOp HIF analysis script	8 months ago
supplementary_3	DPA analysis script	10 months ago
supplementary_4	RNA-seq analysis script	10 months ago
.DS_Store	need to knit with link	10 months ago
.gitignore	first commit	10 months ago
LICENSE	Initial commit	10 months ago
README.md	Update README.md	4 months ago
Taagen_2021_TPG.Rproj	R project, loading data to github	10 months ago

README.md

Taagen et al., 2021, The Plant Genome

Reproducibility index

Data, analysis, and scripts for reviewing or reproducing results presented in Taagen et al., 2021. To view the final analysis output, please see the **.md** files. To view the entire script or download and reproduce analysis in R or command line, please see the **.Rmd** files. All of the data is accessible within .Rmd file via GitHub (no need to download data). **Please cite and/or credit:** Taagen et al., 2021.



Taagen_2021_TPG / supplementary_2 /

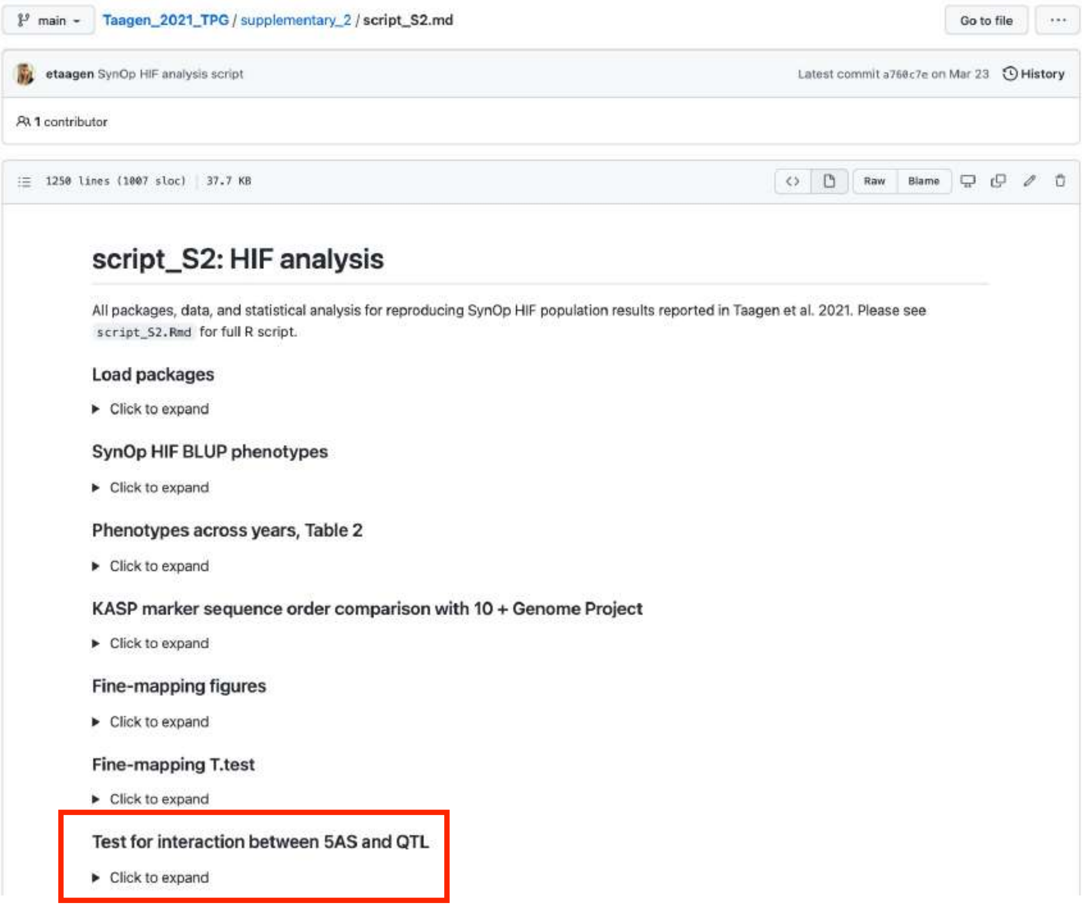
etaagen SynOp HIF analysis script a760c7e on Mar 23

..		
script_S2_files/figure-gfm	SynOp HIF analysis script	10 months ago
.DS_Store	HIF analysis script	10 months ago
file_S2.1.csv	SynOp HIF tidy data format phenotypes, for BLUP calculations input	10 months ago
file_S2.10.csv	SynOp HIF BLUP phenotype measurements, chr 5A fixed effect	10 months ago
file_S2.11.csv	Chr 5A primer sequences	10 months ago
file_S2.12.xlsx	HIF groups, with and without chr 5AS genotype	10 months ago
file_S2.13.pdf	Peak KASP marker clustering patterns	8 months ago
file_S2.2 .csv	Table 2, HIF phenotype averages across years	10 months ago
file_S2.3.csv	SynOpHIF BLUP calculation with chr 5AS fixed effect,for Figure 4C	10 months ago
file_S2.4.csv	SynOpHIF original BLUP calculation, chr 5A and QTL genotype, for figu...	10 months ago
file_S2.5.csv	HIF chr 5A and QTL genotype, and BLUP phenotype for LD calculation	10 months ago
file_S2.6.csv	10 genomes blast positions of QTL KASP markers	10 months ago
file_S2.7.csv	HIF by group for t.test comparison fine-mapping (Figure 2)	10 months ago
file_S2.8.pdf	Figure 2	10 months ago
file_S2.9.csv	SynOp HIF BLUP phenotype measurements	10 months ago
script_S2.Rmd	SynOp HIF analysis script	10 months ago
script_S2.md	SynOp HIF analysis script	8 months ago

Data input/output

Scripts for data analysis

github.com/etaagen/Taagen_2021_TPG



Taagen_2021_TPG / supplementary_2 / script_S2.md

etaagen SynOp HIF analysis script

Latest commit a760c7e on Mar 23 History

1 contributor

1250 lines (1007 sloc) 37.7 KB

script_S2: HIF analysis

All packages, data, and statistical analysis for reproducing SynOp HIF population results reported in Taagen et al. 2021. Please see `script_S2.Rmd` for full R script.

Load packages

▶ Click to expand

SynOp HIF BLUP phenotypes

▶ Click to expand

Phenotypes across years, Table 2

▶ Click to expand

KASP marker sequence order comparison with 10 + Genome Project

▶ Click to expand

Fine-mapping figures

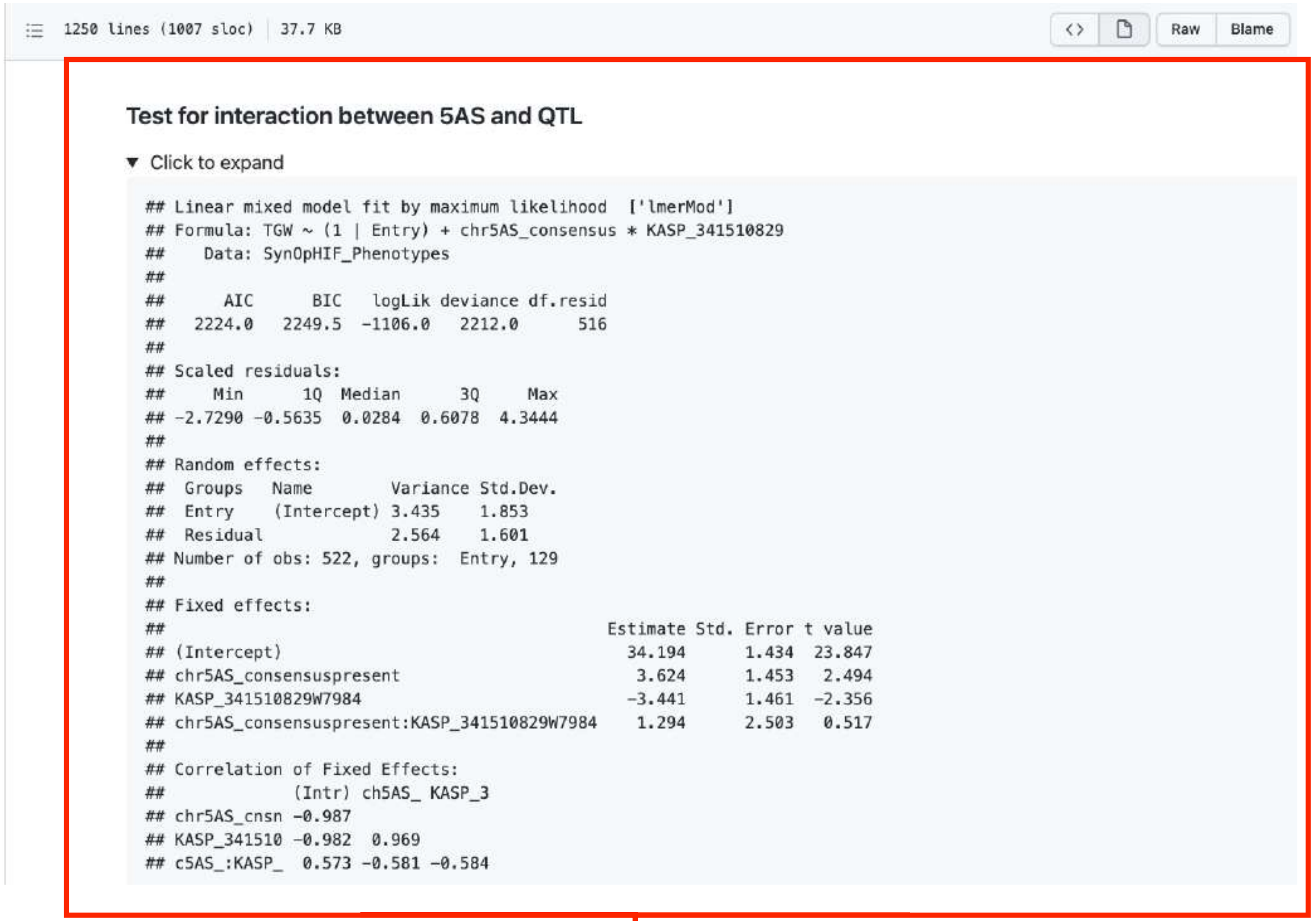
▶ Click to expand

Fine-mapping T.test

▶ Click to expand

Test for interaction between 5AS and QTL

▶ Click to expand



1250 lines (1007 sloc) 37.7 KB

Test for interaction between 5AS and QTL

▼ Click to expand

```
## Linear mixed model fit by maximum likelihood  ['lmerMod']
## Formula: TGW ~ (1 | Entry) + chr5AS_consensus * KASP_341510829
##    Data: SynOpHIF_Phenotypes
##
##      AIC      BIC   logLik deviance df.resid
##   2224.0   2249.5  -1106.0   2212.0      516
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7290 -0.5635  0.0284  0.6078  4.3444
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Entry    (Intercept) 3.435    1.853
##  Residual             2.564    1.601
## Number of obs: 522, groups:  Entry, 129
##
## Fixed effects:
##                                             Estimate Std. Error t value
## (Intercept)                                   34.194      1.434  23.847
## chr5AS_consensuspresent                        3.624      1.453   2.494
## KASP_341510829W7984                           -3.441      1.461  -2.356
## chr5AS_consensuspresent:KASP_341510829W7984    1.294      2.503   0.517
##
## Correlation of Fixed Effects:
##             (Intr) ch5AS_ KASP_3
## chr5AS_cnsn -0.987
## KASP_341510 -0.982  0.969
## c5AS_:KASP_  0.573 -0.581 -0.584
```

Review code and reproduce results

Acknowledgements

Dr. Mark Sorrells lab

Wheat CAP colleagues & collaborators

Eastern Regional Small Grains Genotyping Lab

International Wheat Genome Sequencing Consortium

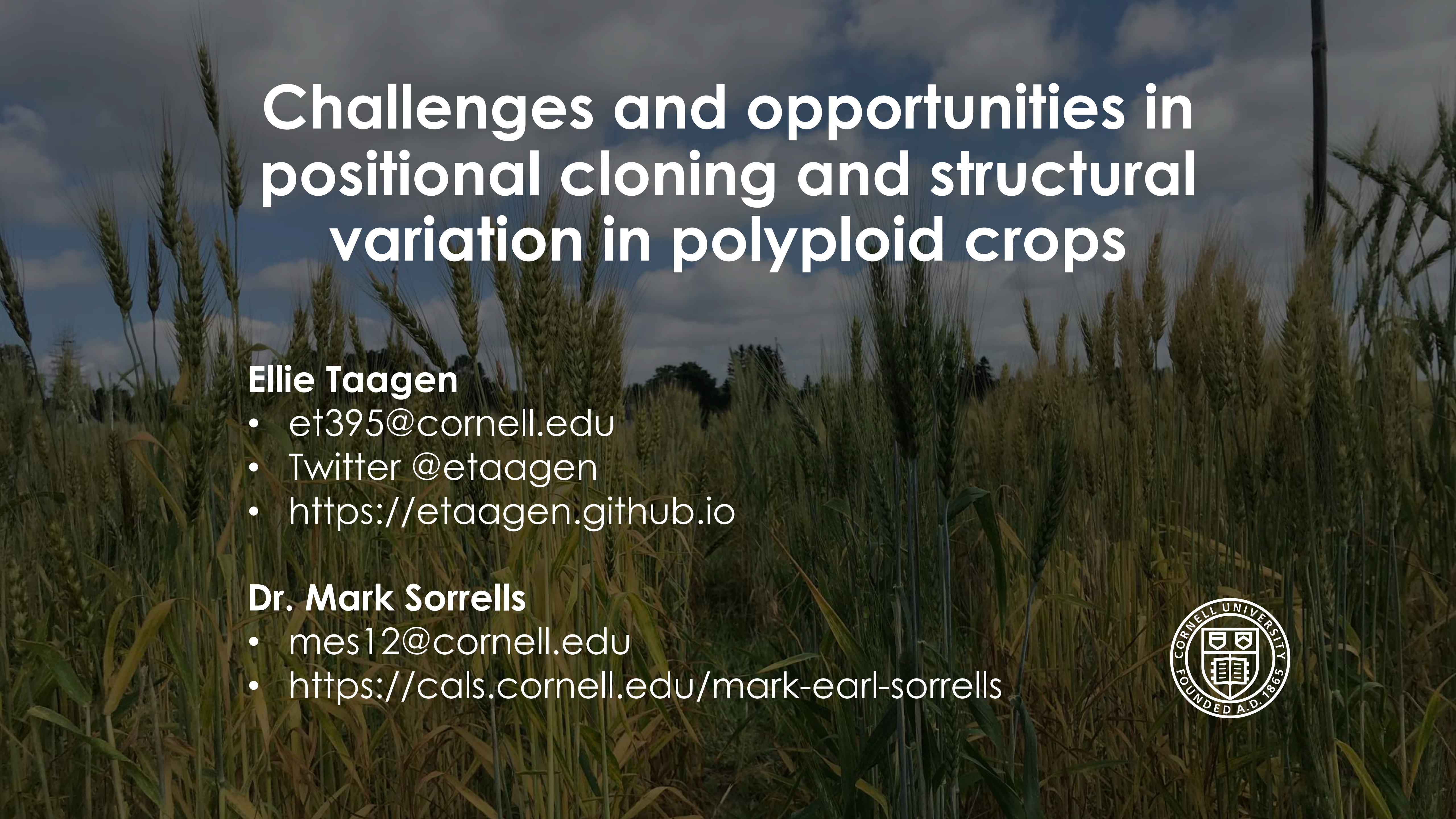
Open-source tools

Cornell CALS
College of Agriculture and Life Sciences

USDA NIFA

Wheat CAP
Coordinated Agricultural Project

IWGSC



Challenges and opportunities in positional cloning and structural variation in polyploid crops

Ellie Taagen

- et395@cornell.edu
- Twitter @etaagen
- https://etaagen.github.io

Dr. Mark Sorrells

- mes12@cornell.edu
- https://cals.cornell.edu/mark-earl-sorrells

Sources

Taagen, E., Tanaka, J., Gul, A., & Sorrells, M. E. (2021). Positional-based cloning 'fail-safe' approach is overpowered by wheat chromosome structural variation. *Plant Genome*. 2021;14:e20106. https://doi.org/10.1002/tpg2.20106

Gutierrez-Gonzalez, J.J., Mascher, M., Poland, J. *et al.* Dense genotyping-by-sequencing linkage maps of two Synthetic W7984×Opata reference populations provide insights into wheat structural diversity. *Sci Rep* **9**, 1793 (2019). https://doi.org/10.1038/s41598-018-38111-3

Brinton, J., Ramirez-Gonzalez, R. H., Simmonds, J., Wingen, L., Orford,S., Griffiths, S., Haberer, G., Spannagl, M., Walkowiak, S., Poz-niak, C., & Uauy, C. (2020). A haplotype-led approach to increase the precision of wheat breeding. *Communications Biology*,3, 712. https://doi.org/10.1038/s42003-020-01413-2

Walkowiak, S., Gao, L., Monat, C., Haberer, G., Kassa, M. T., Brinton,J., Ramirez-Gonzalez, R. H., Kolodziej, M. C., Delorean, E., Tham-bugala, D., Klymiuk, V., Byrns, B., Gundlach, H., Bandi, V., Siri, J.N., Nilsen, K., Aquino, C., Himmelbach, A., Copetti, D., ... Poz-niak, C. J. (2020). Multiple wheat genomes reveal global variation in modern breeding. *Nature*,588, 277–283. https://doi.org/10.1038/s41586-020-2961-x