NC STATE UNIVERSITY Department of Crop and Soil Sciences

USDA Agricultural Research Service

Development of a Pangenome Reference Model for United States Soft Red Winter Wheat

Matthew Willman[1*], Daniela Miller[1], Luis Rivera-Burgos,[2,1], Vijay Tiwari[3], Amanda Hulse-Kemp[2,1], Gina Brown-Guedira[2,1]

1. North Carolina State University, College of Agriculture and Life Sciences
2. United States Department of Agriculture, Agricultural Research Service
3. University of Maryland, College of Agriculture and Natural Resources

* mrwillma@ncsu.edu

Introduction

Reference alignment is complicated by hexaploidy in wheat (*Triticum aestivum*). Misalignment to homoeologous chromosomes leads to inaccurate genotyping (Figure 1). Development of a pangenome reference model may address this issue by increasing representation of sampled alleles.

Soft red winter wheat (SRWW) is a market class of hexaploid wheat and is the predominant wheat class grown in the United States (US) Eastern region (Figure 2).

Short-read exome-capture data were previously generated for a panel of SRWW and related accessions selected by genetic diversity and use as breeding program founders (Figures 3). Reference-quality assemblies were generated for select accessions.

Hypothesis

Alignment of short reads to SRWW genome reference models will reduce misalignment and improve genotyping accuracy in US SRWW germplasm compared to alignment to the standard 'Chinese Spring' genome reference.

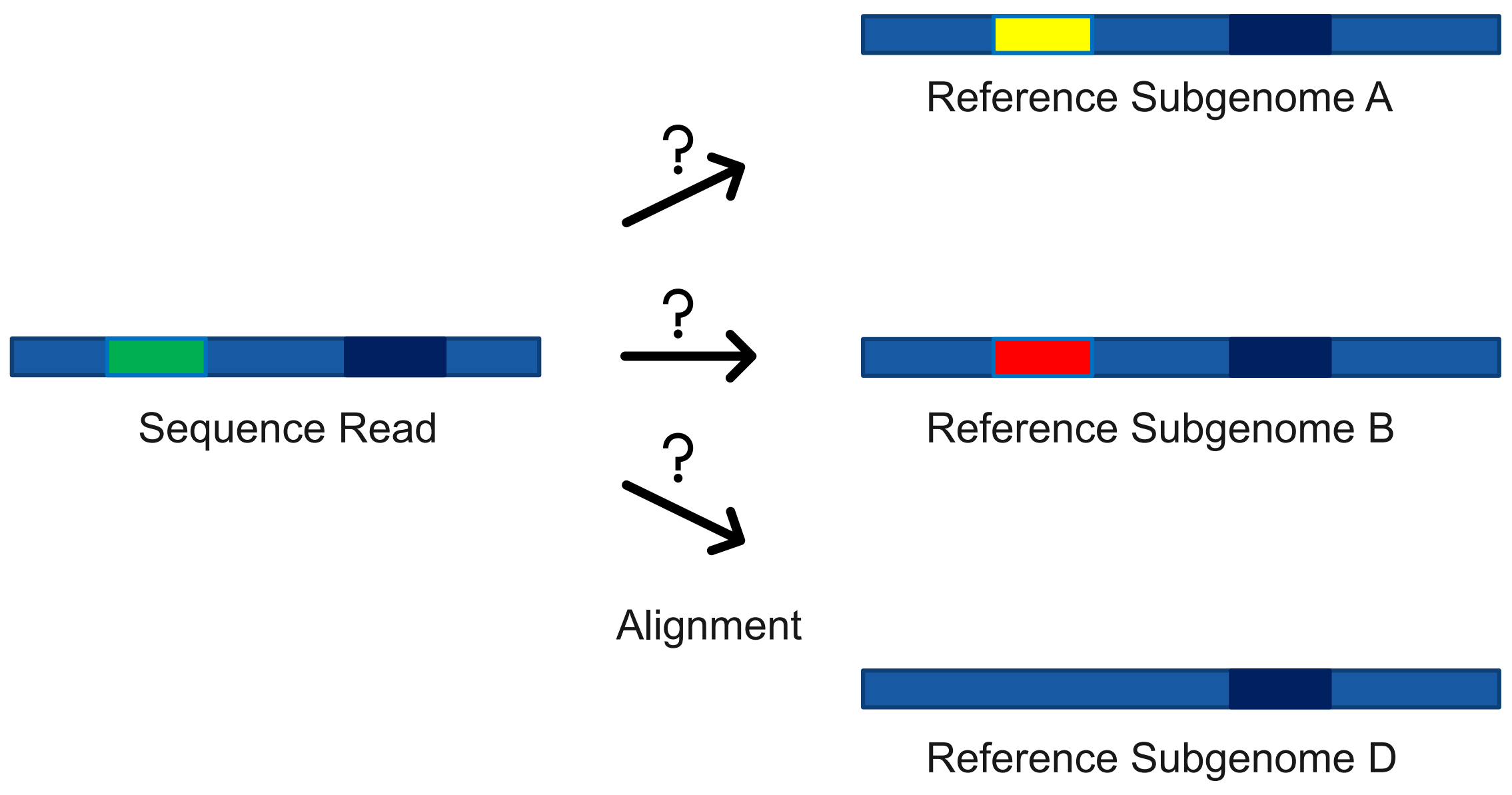

Figure 1. Homoeology between wheat's three subgenomes can complicate alignment and allele discovery. Misalignment will result in erroneous genotyping.

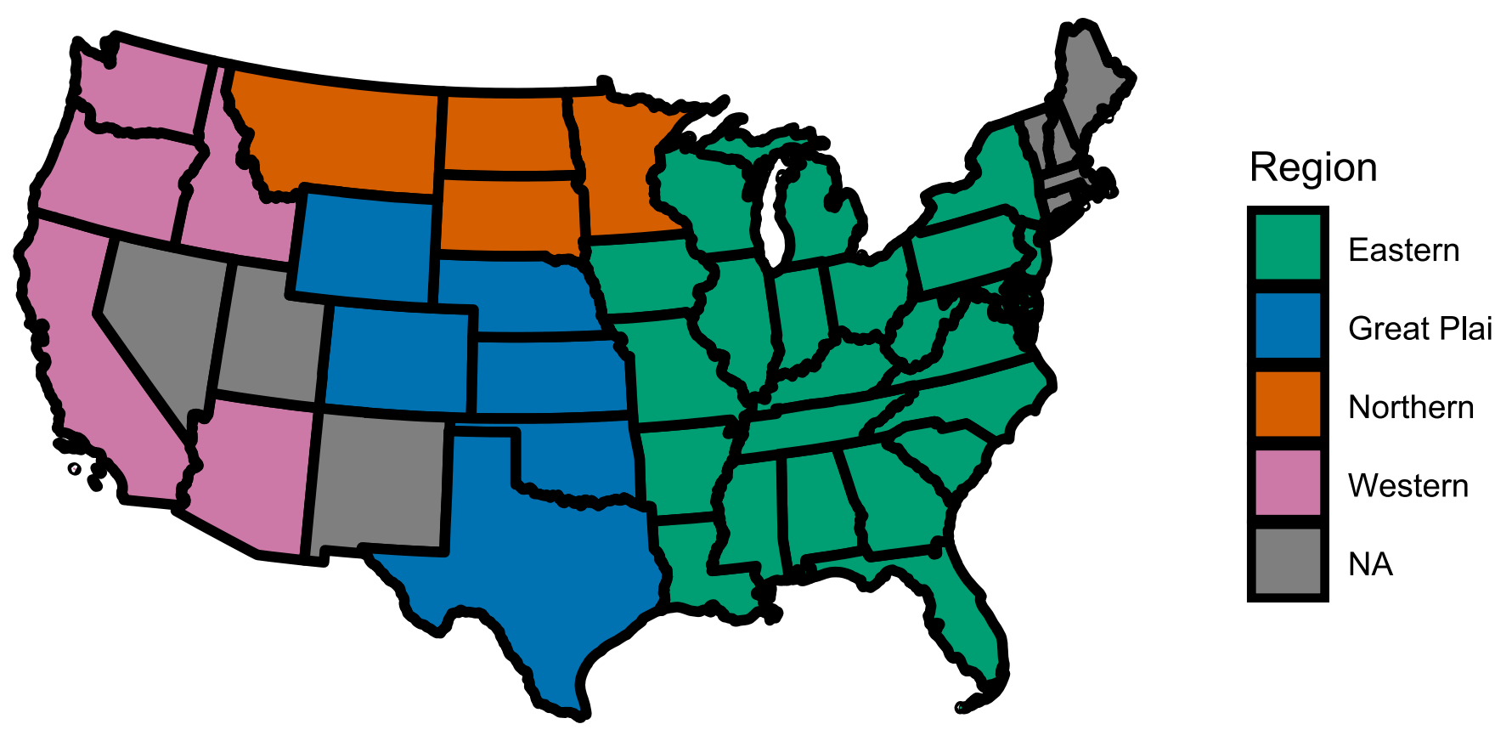

Figure 2. SRWW is the predominant wheat class grown in the US Eastern region.

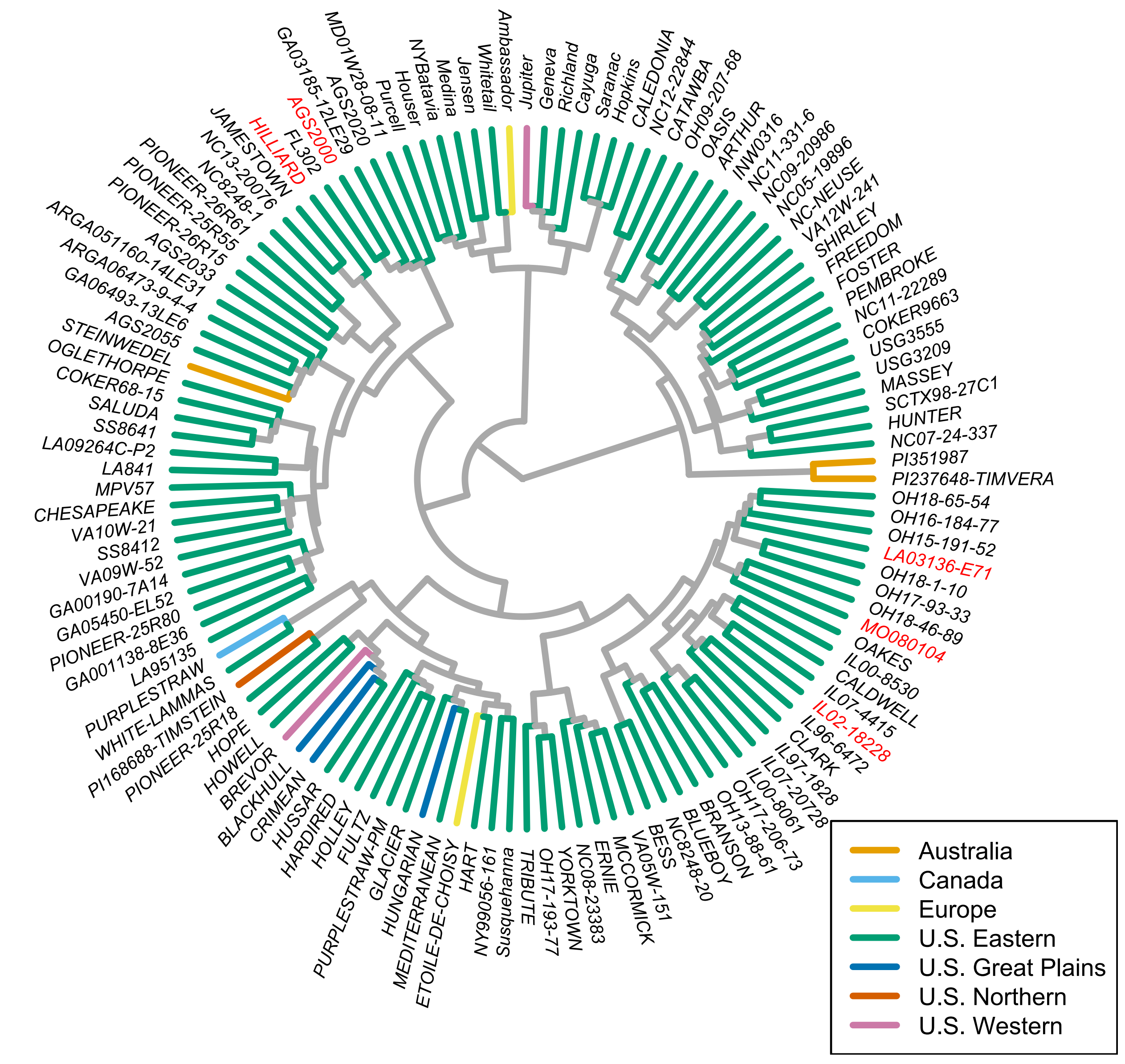

Figure 3. Radial plot of phylogenetic distances of SRWW panel accessions. Edge colors indicate accession origin. Reference-quality assemblies have been generated for accessions listed in red.

Methods

To examine alignment reference affects on mapping quality and genotyping accuracy, short read exome capture data were aligned to 'Hilliard', 'AGS2000', and 'Chinese Spring' reference assemblies by bowtie2, and genotypes were called by GATK. Mean alignment quality (MAPQ) scores were estimated for each accession. Inbreeding coefficient scores were estimated by vcftools.

Introgressions in SRWW assemblies were explored by alignment of short read genotyping-by-sequencing (GBS) data from related species[1]. Structural variation between SRWW assemblies was examined through whole chromosome alignment by MUMmer.

Results

Alignment of panel short reads to local references increased MAPQ scores for most accessions compared to alignment to the standard 'Chinese Spring' assembly (Figure 4).

Alignment to 'Hilliard' increased inbreeding coefficient (F) estimates for most panel individuals when compared to 'Chinese Spring'. Alignment to 'AGS2000' reduced F estimates for most panel individuals while greatly increasing F estimates for closely related individuals (Figure 4).

Alignment of GBS short reads from related species to SRWW references revealed two introgressions on chromosome 6A possibly originating from *T. monococcum*. One introgression appeared in both accessions, and one appeared in 'Hilliard' only. Whole-chromosome alignment revealed a gap in synteny at the location of the 'Hilliard'-specific introgression (Figure 5).

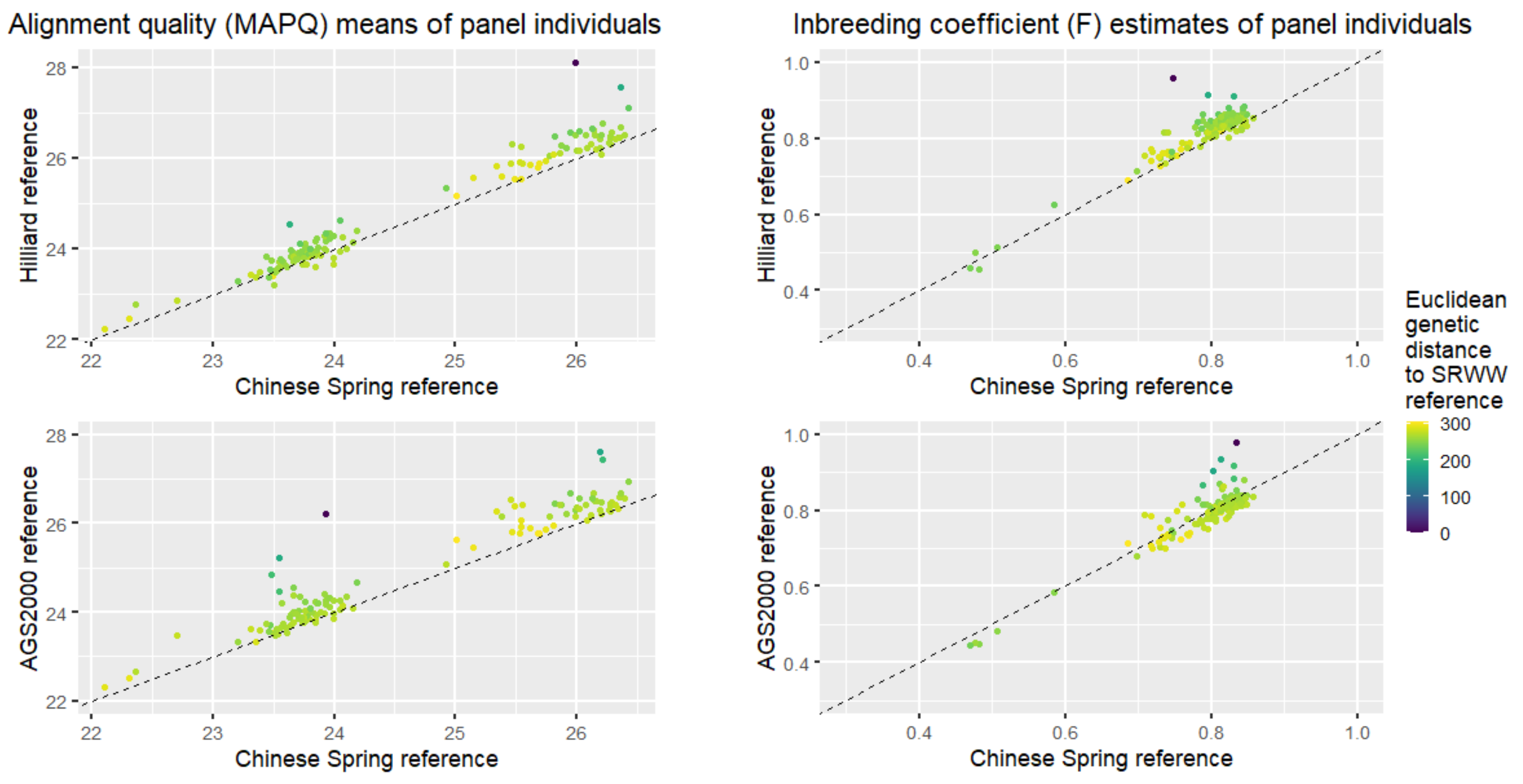

Figure 4. Alignment of panel exome capture data to SRWW assemblies increased overall MAPQ and F values. Dashed lines indicate y = x. Expected F for each line is 1, given that each accession is derived from a highly inbred line. F estimates may be reduced by erroneous heterozygous calls resulting from read misalignments.

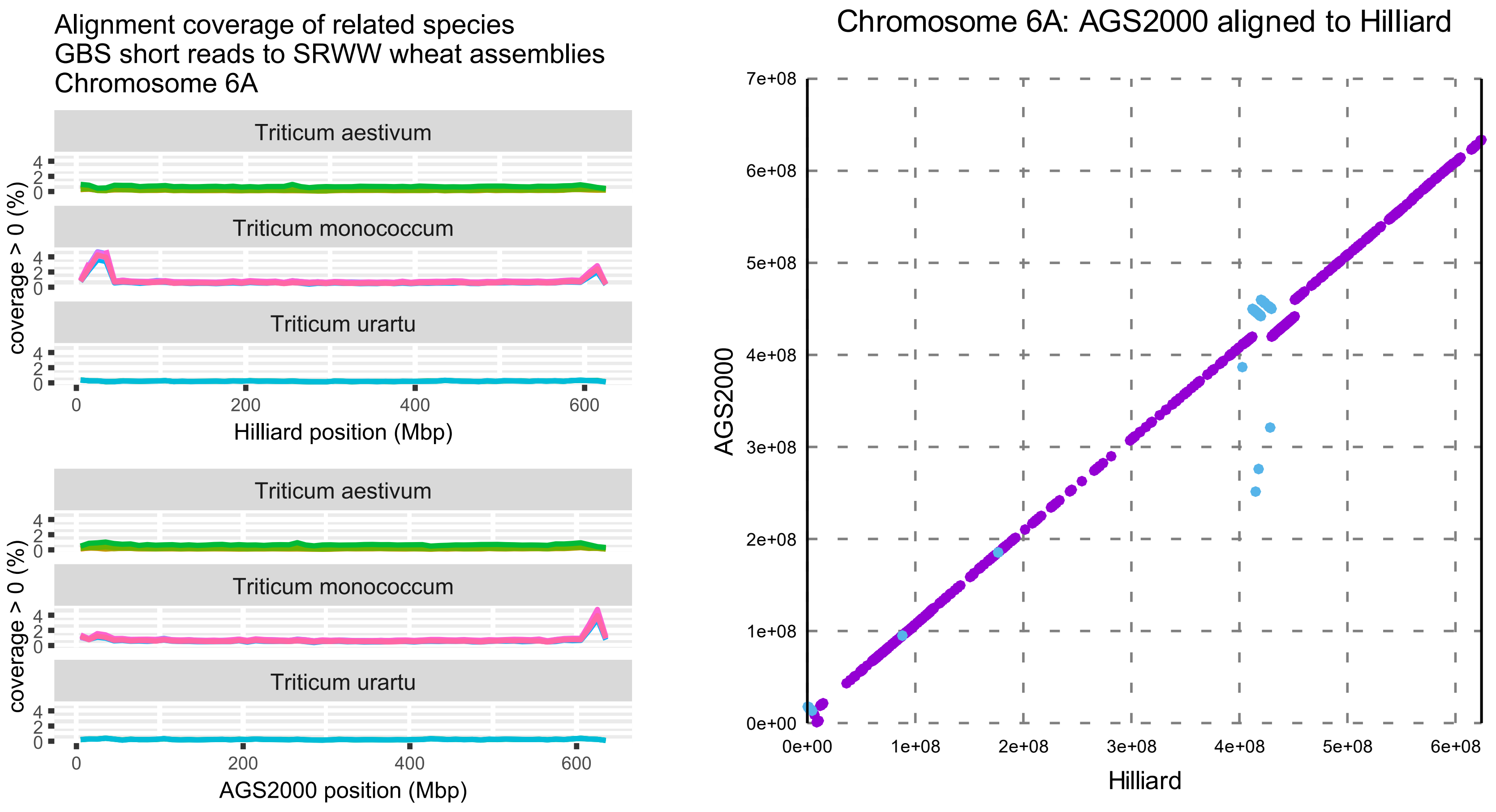

Figure 5. Alignment coverage of related species to SRWW reference assemblies revealed two putative *T. monococcum* introgressions. One introgression unique to 'Hilliard' overlaps a gap in whole genome alignment between 'Hilliard' and 'AGS2000' assemblies.

Conclusions

Alignment of short read sequences to a SRWW reference assembly can improve alignment quality and genotyping accuracy of related germplasm. This highlights the importance of reference selection in genotyping. Use of a poor alignment reference can inflate heterozygous calls due to misalignment.

Interspecies introgressions are a source of variation in wheat that influence alignment of short read data. Failure to model introgressions may lead to misalignment and erroneous genotyping. Increased allelic diversity may be captured by combining individual assemblies in a SRWW pangenome reference, leading to further improvement of genotyping accuracy across diverse accessions.

Future Directions

Pangenome models will be constructed using reference-quality assemblies (20x), long-read sequencing (5x), and exome-capture short-read sequencing data. Additional sequencing efforts will prioritize diverse SRWW accessions. Model efficacy will be tested by trait-linked marker discovery and genomic prediction in diverse US SRWW germplasm.

References

1. Hyun, D. Y. et al. Genotyping-by-Sequencing Derived Single Nucleotide Polymorphisms Provide the First Well-Resolved Phylogeny for the Genus *Triticum* (Poaceae). Front. Plant Sci. 11, 688 (2020).

Acknowledgements

Special thanks to Brian Ward (Forage Genetics International), Katherine Jordan (USDA-ARS), and Nico Lara (NCSU).
Funding was provided by USDA NIFA grant 2022-68013-36439 - Wheat CAP
USDA-ARS CRIS projects 6070-22000-019-000-D and 6066-21310-006-000-D

WHEAT CAP Coordinated Agricultural Project