

Adaptive introgression from wild emmer into bread wheat revealed by sequencing 1000 wheat exomes

Fei He

e-mail: feihe@ksu.edu

Postdoc at Akhunov lab

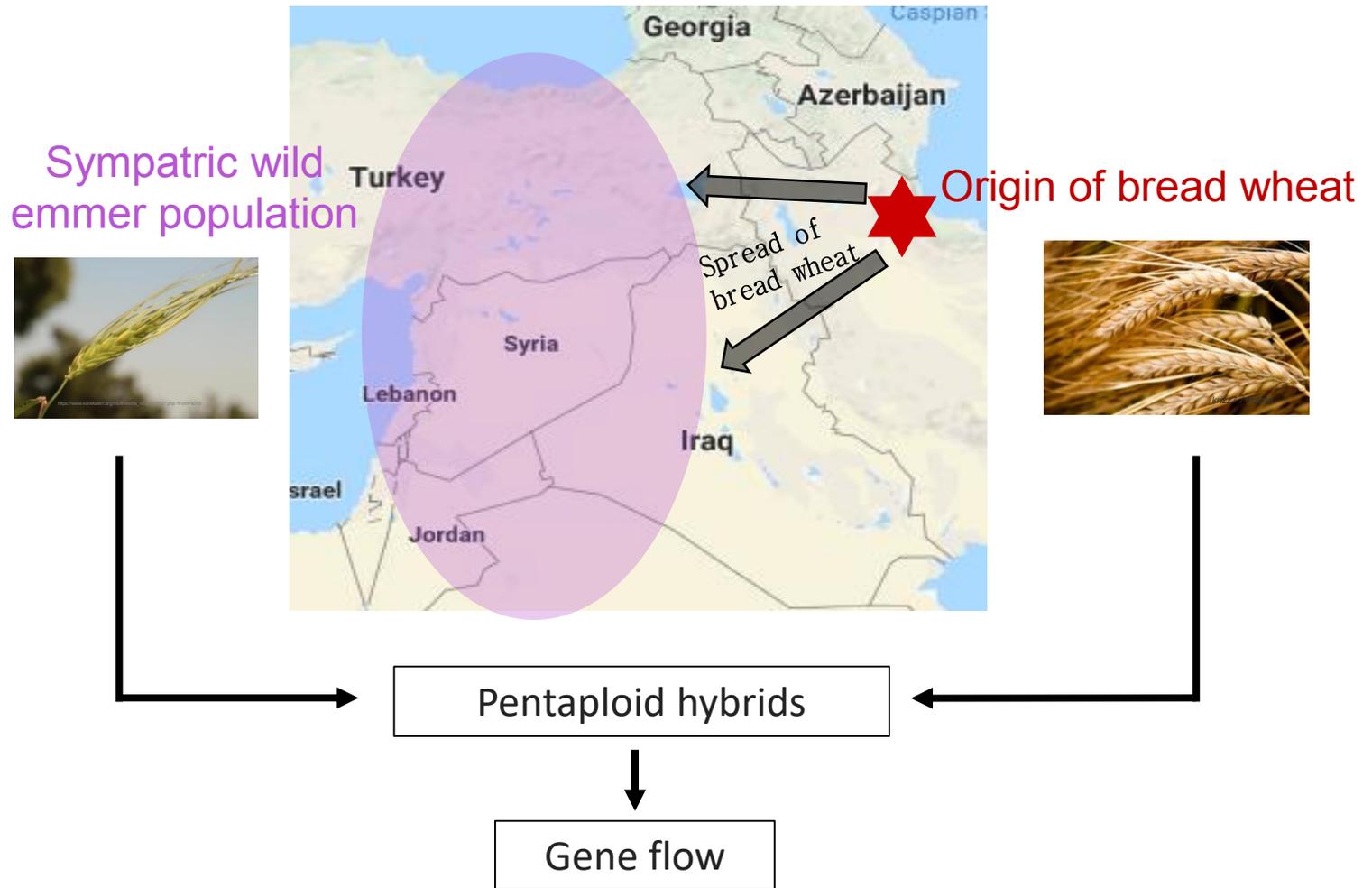
Department of Plant Pathology

Kansas State University

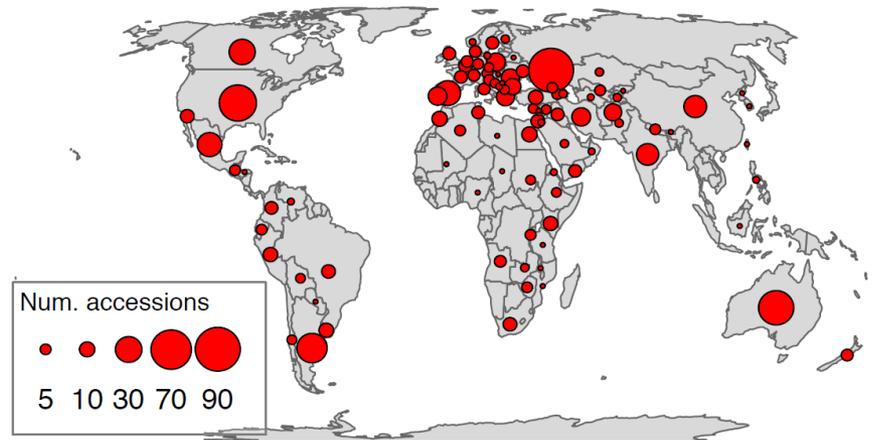
2020-1-11

IWGSC workshop at the Plant and Animal Genome Conference

Prior studies provided evidence for gene flow between wild emmer and bread wheat. The scope and impact of gene flow on wheat was not known.



Data collected in this project



of accessions/lines:

- bread wheat = 803 (747 spring wheat)
- wild emmer = 34
- domesticated emmer = 40
- durum = 39
- spelta = 14
- other relatives and synthetic = 170

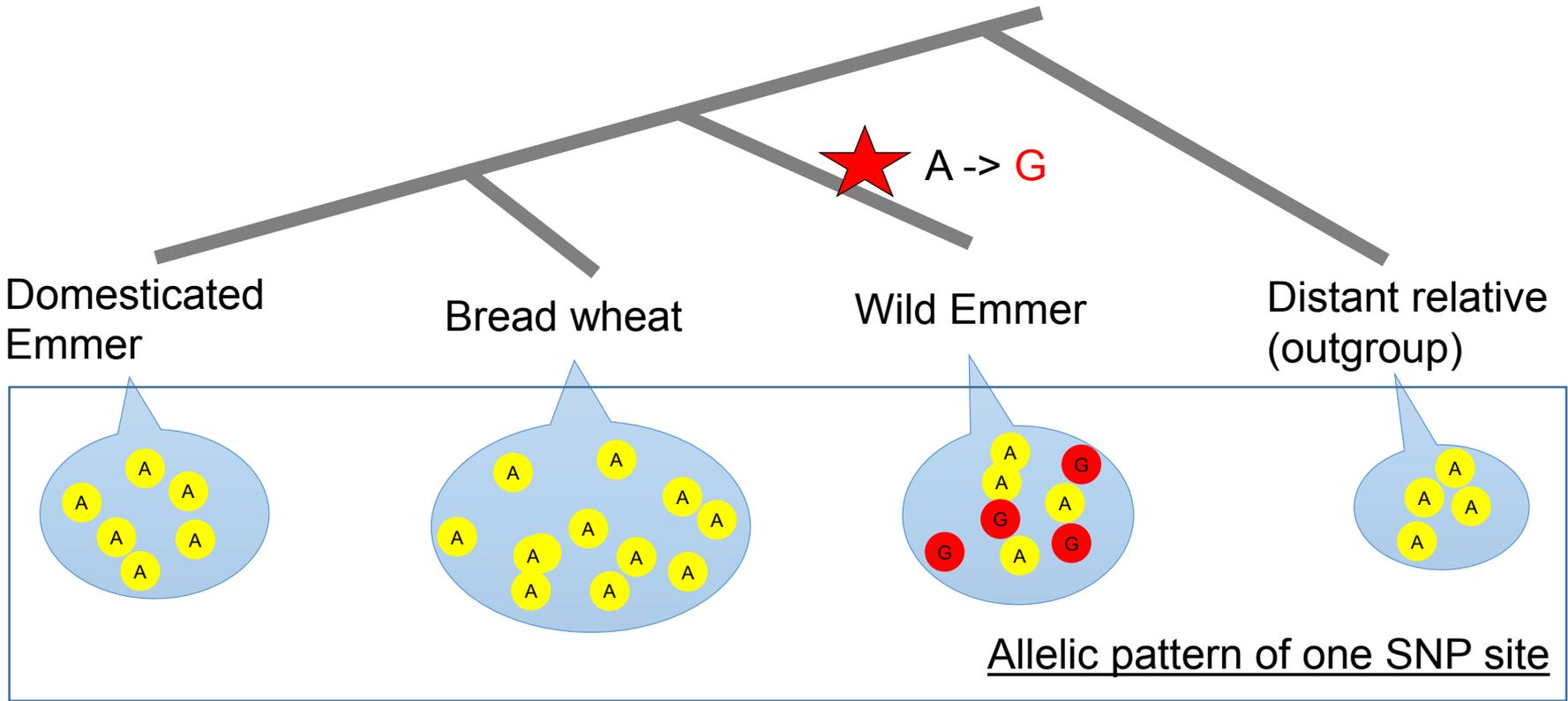
Field phenotyping data:

- Grain filling period
- Heading date
- Plant height
- Harvest weight
- Harvest weight stability (two year repeats)

Exome captured sequencing:

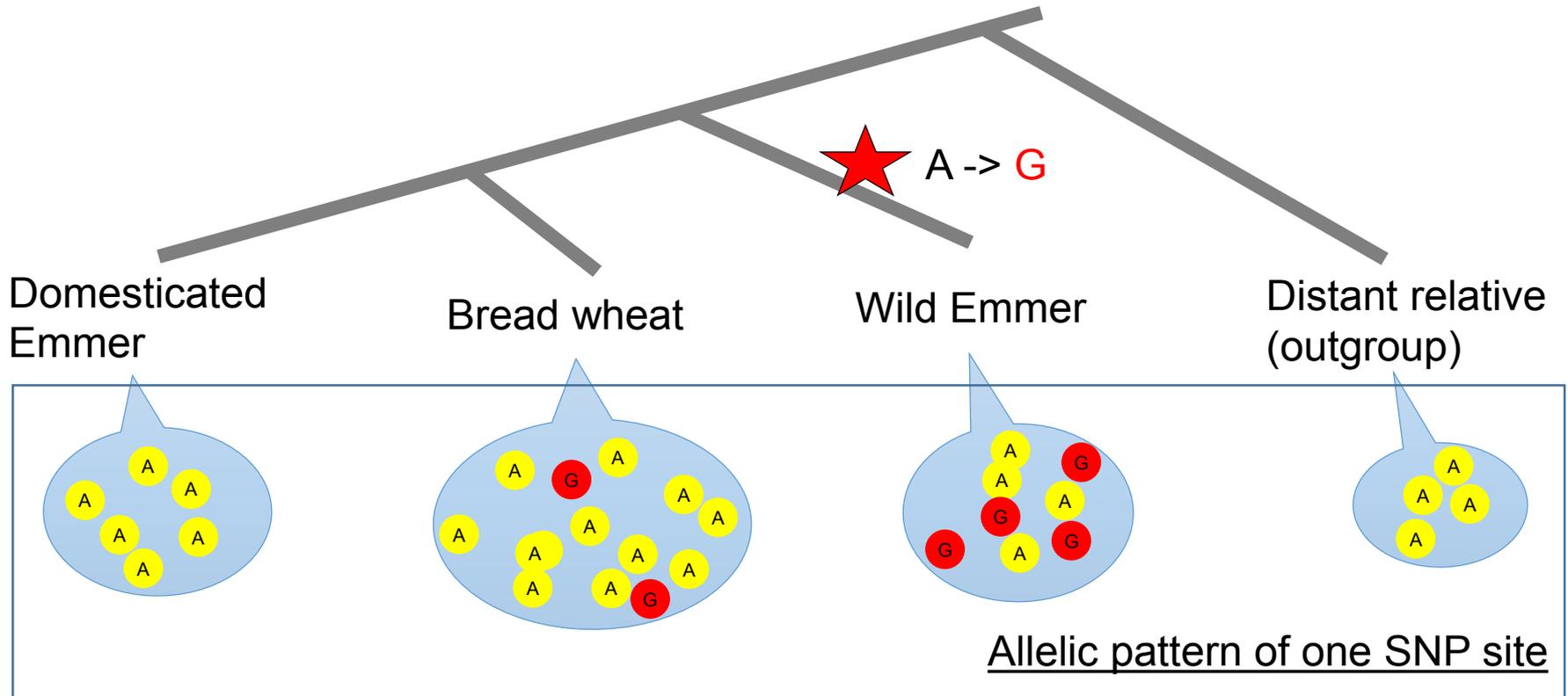
8 million SNP are discovered based on CS reference genome version 1

Tracing wild emmer introgression using SNP data



G is a wild emmer derived and private allele

Gene flow or incomplete lineage sorting ?

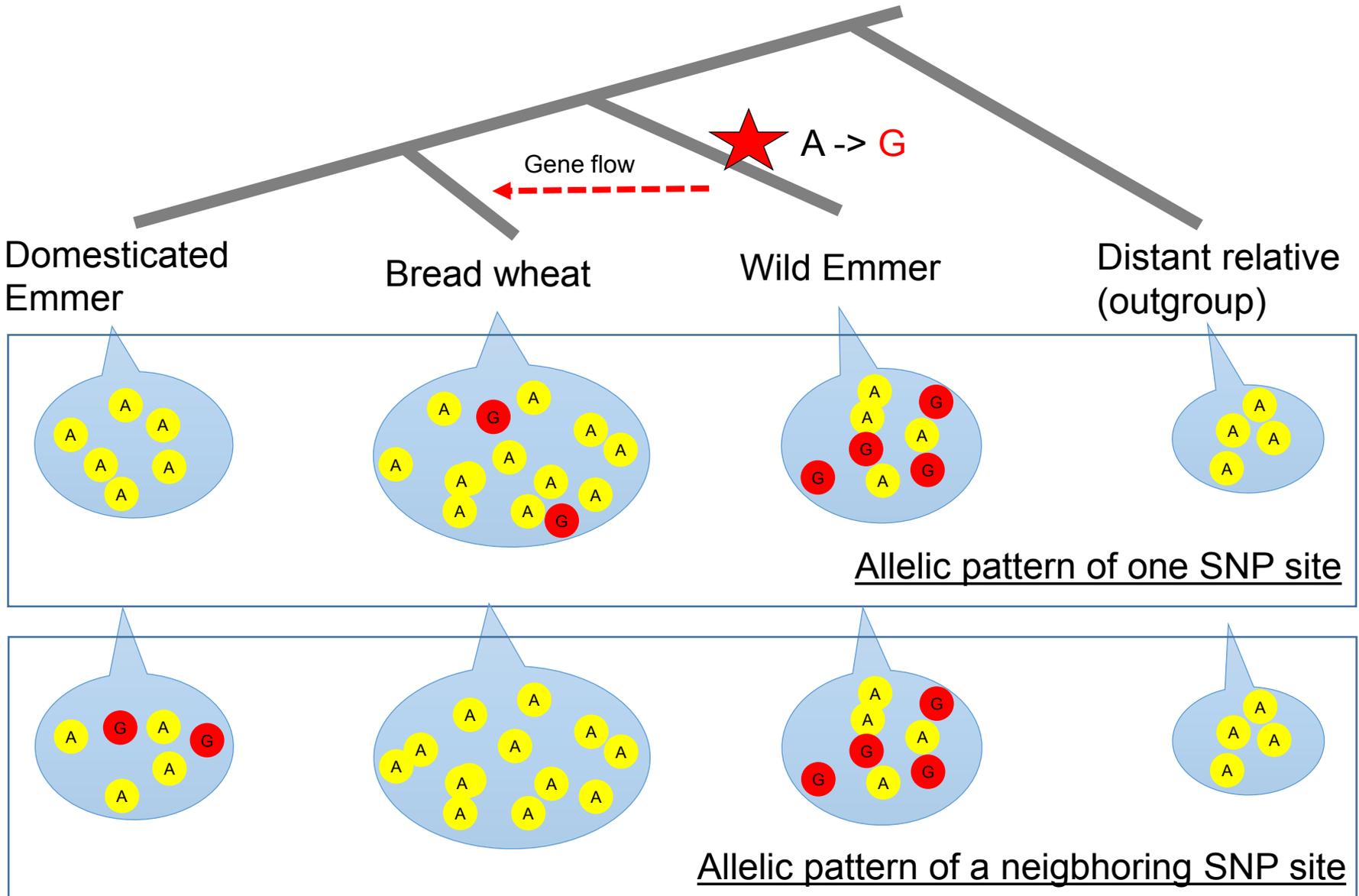


If we observe G in bread wheat but not in domesticated emmer,

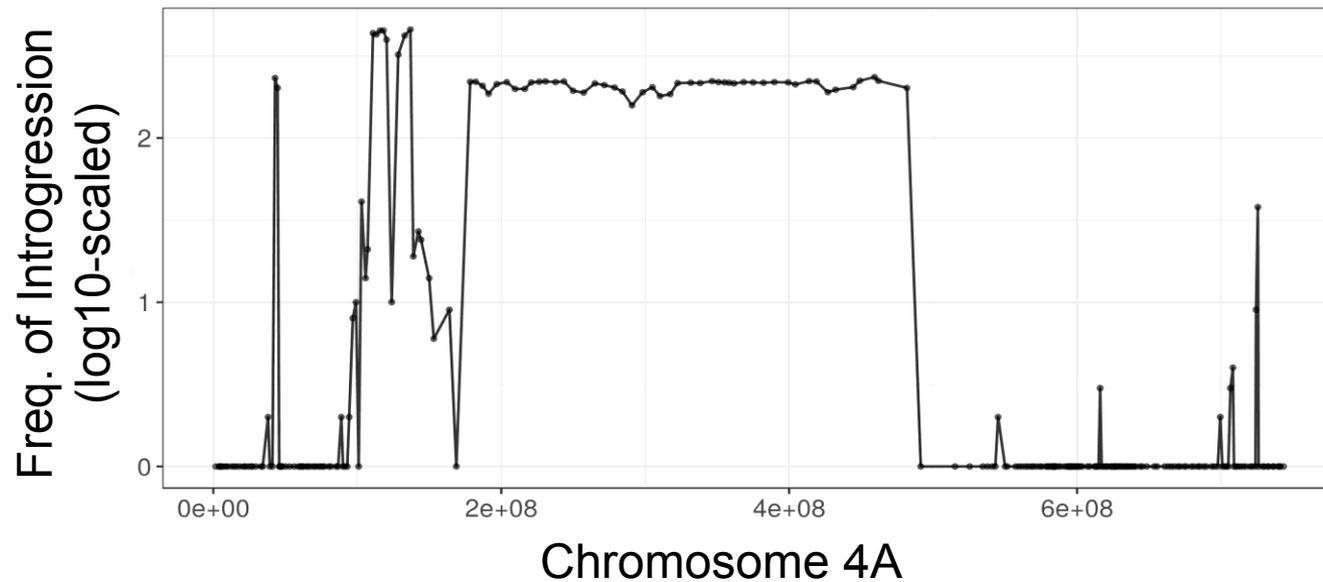
=>

either **gene flow**, or, **incomplete lineage sorting**.

Use f_d test to compare two allelic pattern in a genomic window

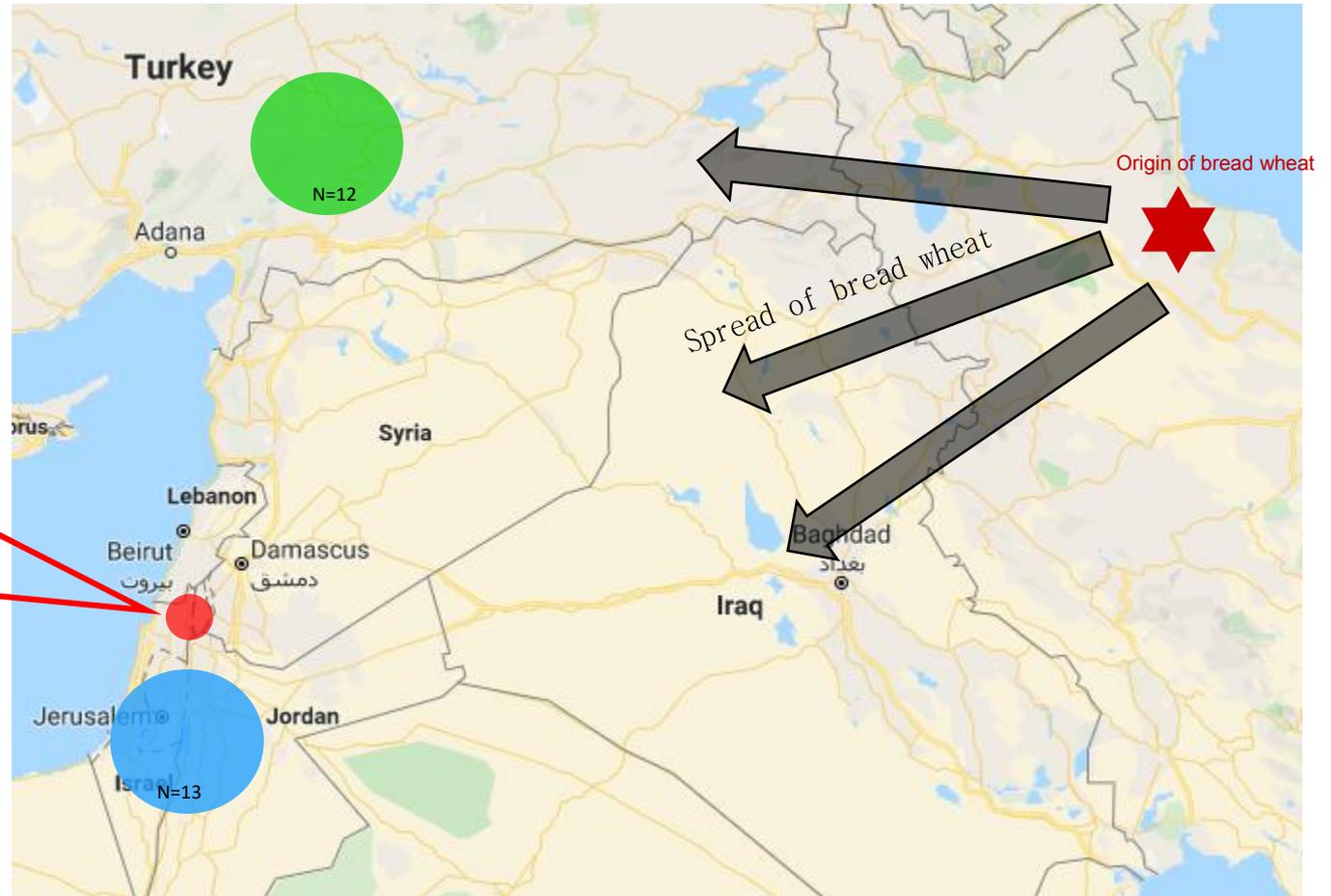


Running f_d test for wild emmer introgression in a population of 800 bread wheat



We estimated gene flow from three wild emmer populations into bread wheat

Wild emmer near the Sea of Galilee

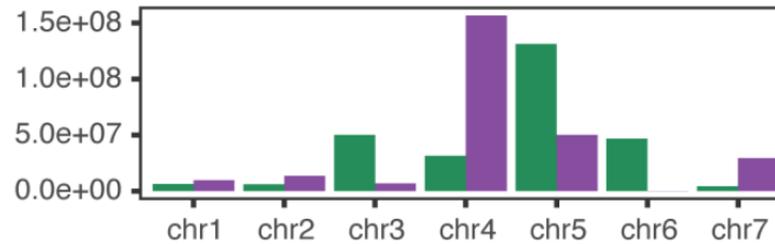


What is the level of gene flow from different wild emmer populations?

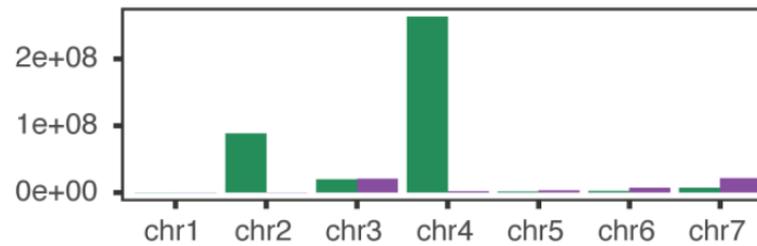
Introgression from different wild emmer source populations differs across wheat genome



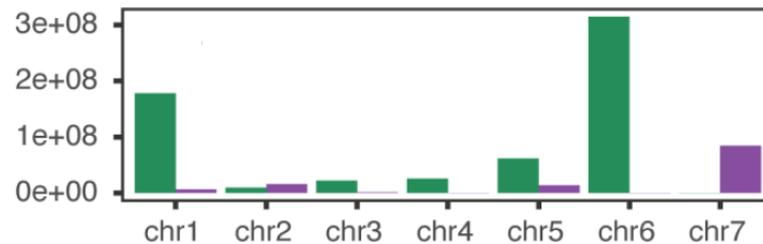
Source wild emmer population: Turkey



Source wild emmer population: Galilee

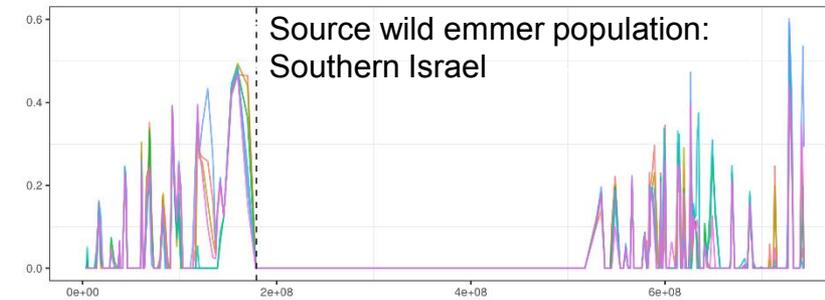
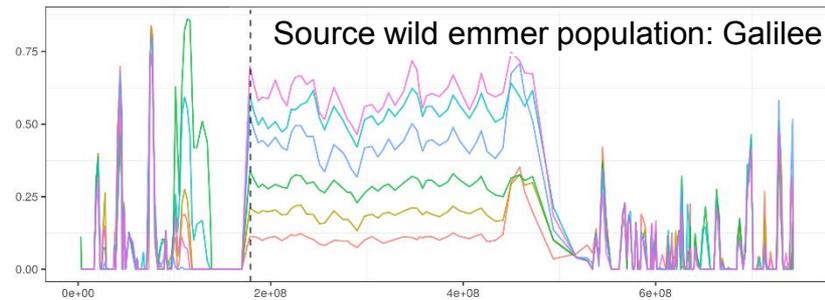
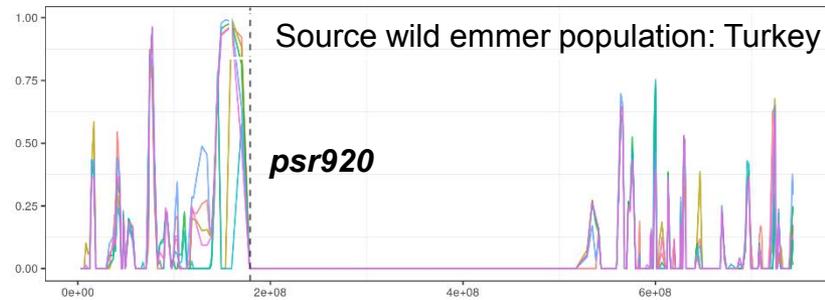


Source wild emmer population: Southern Israel



■ A - genome ■ B - genome

Previously detected introgression on 4A originated from the Sea of Galilee



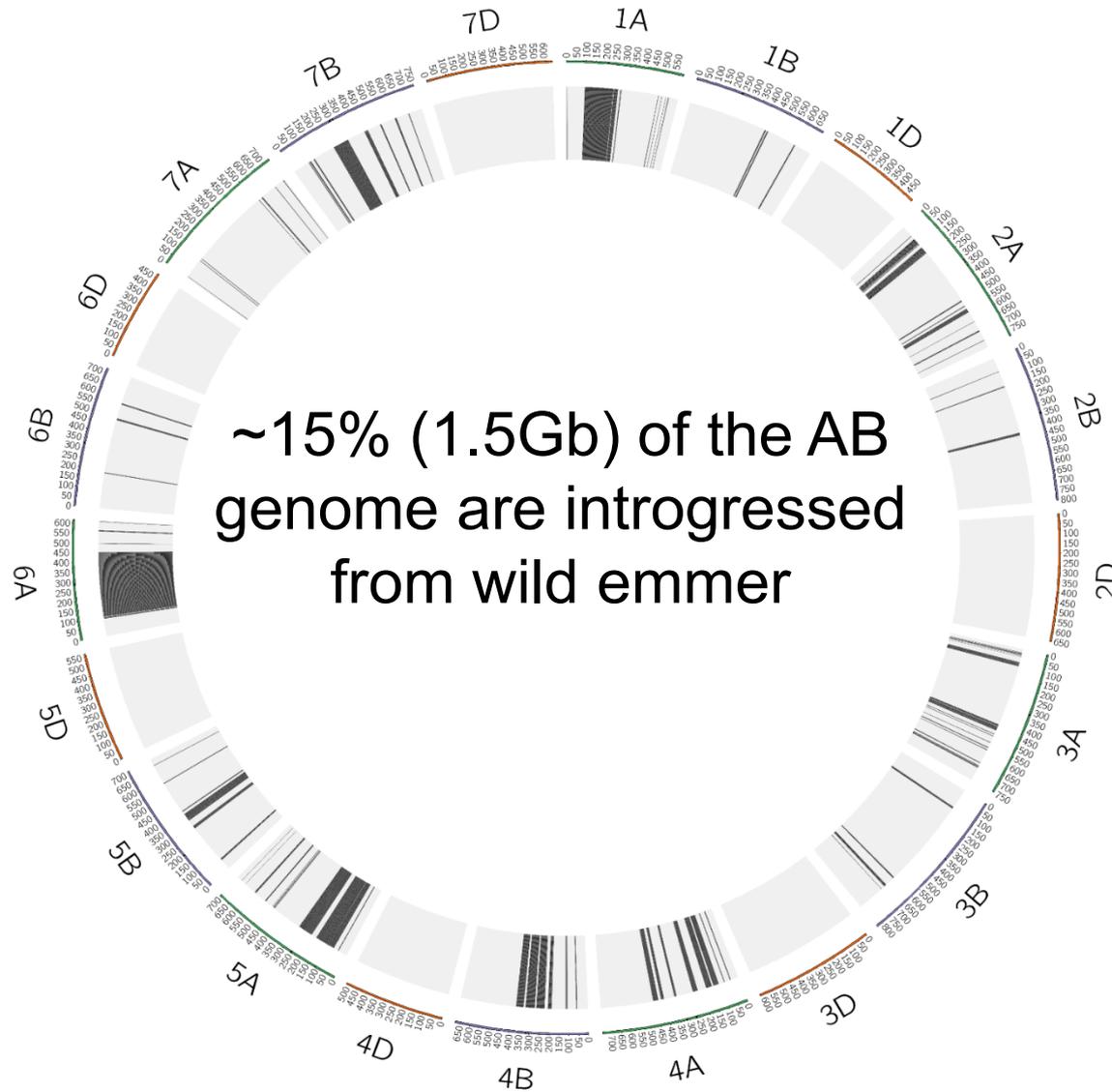
Introgression signal for different landrace population (f_d)

Chromosome 4A

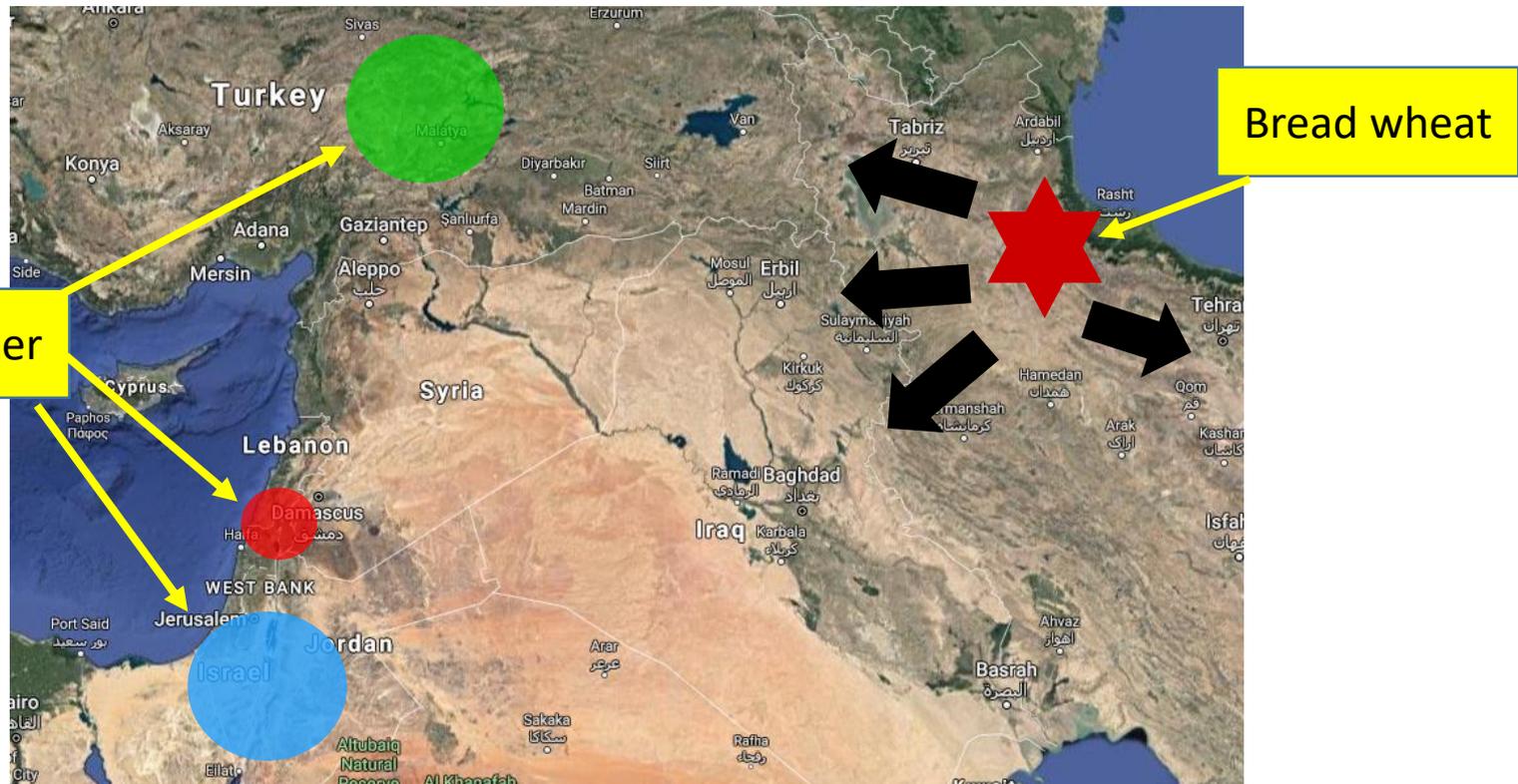
Wheat landrace population:

- Asia/India
- Western Europe
- China
- Middle East
- Eastern Europe
- Turkey

Distribution of wild emmer introgression across the wheat genome in landraces

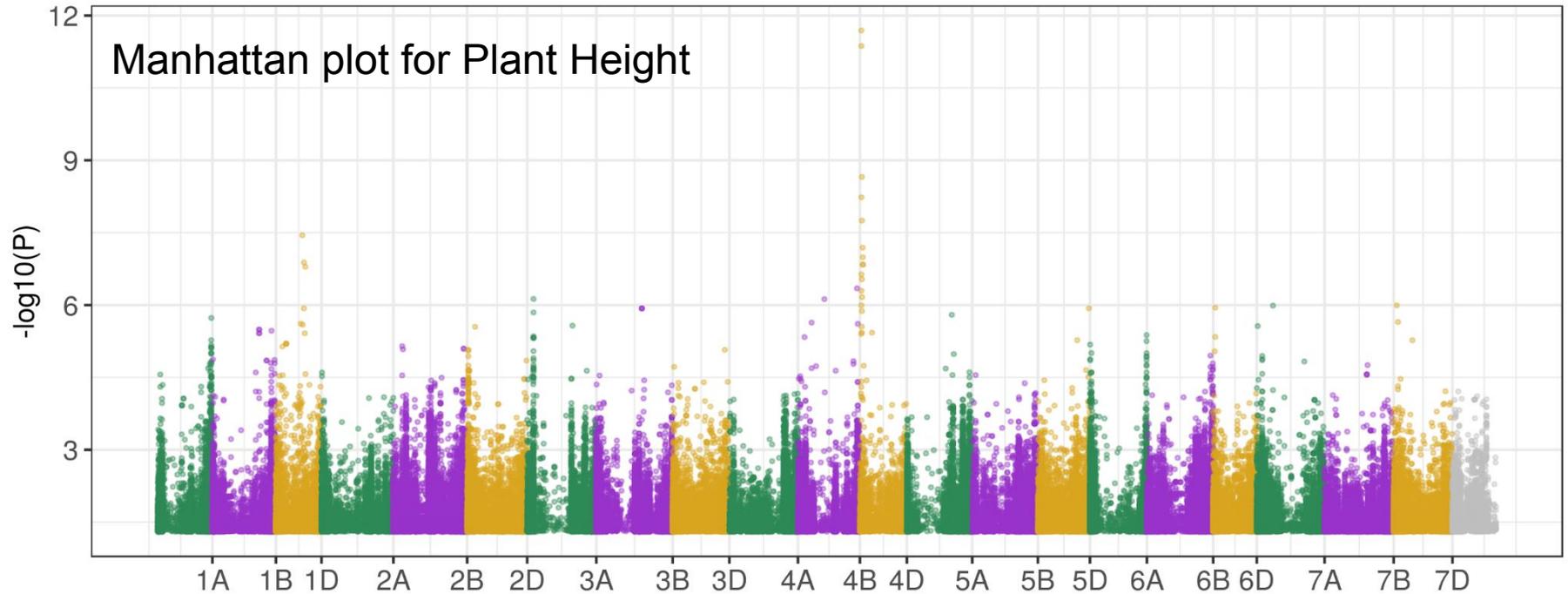


What is the role of wild emmer introgression in shaping bread wheat phenotypes?



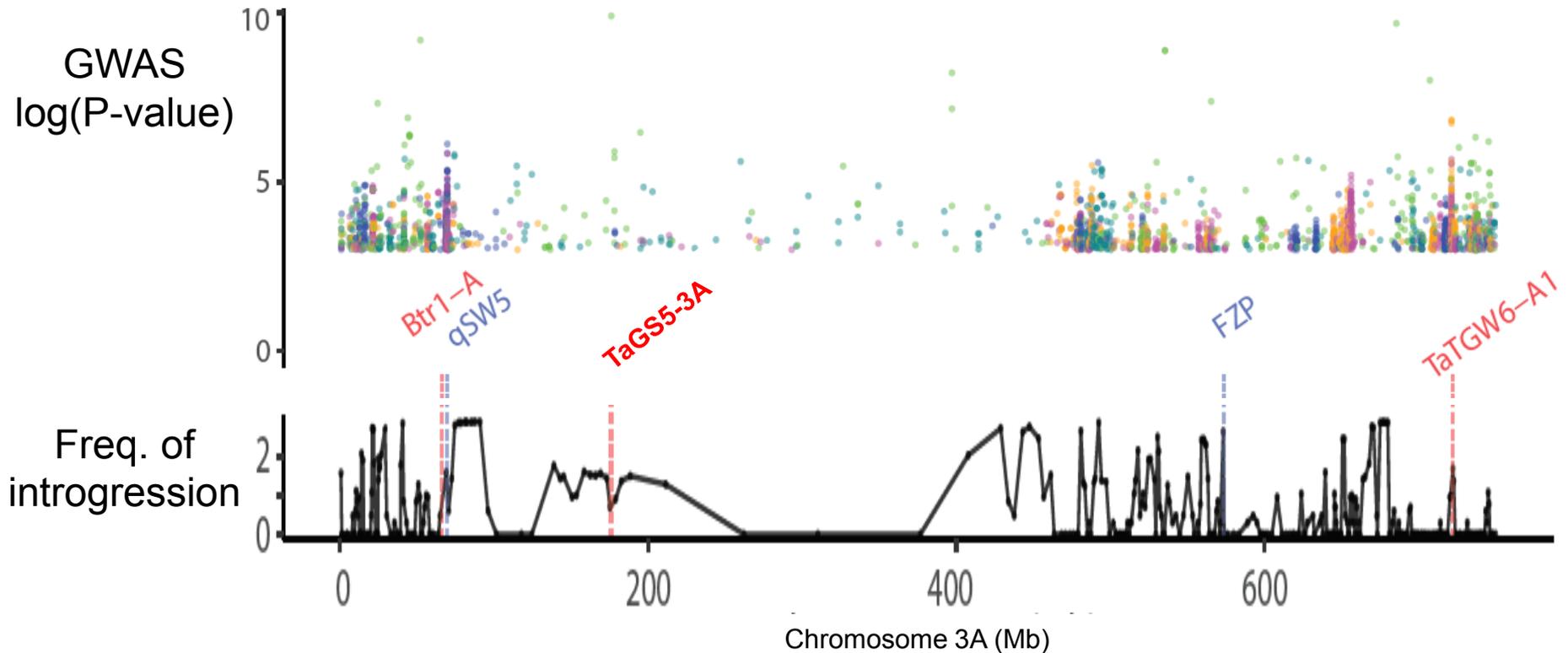
Those three wild emmer population are adapted to environments which are different to southwestern Caspian Sea. Maybe the wild emmer alleles helped wheat to adapt to new environment?

GWAS provides estimate of genome-wide QTL for traits



Evaluate the impact of introgression using GWAS results

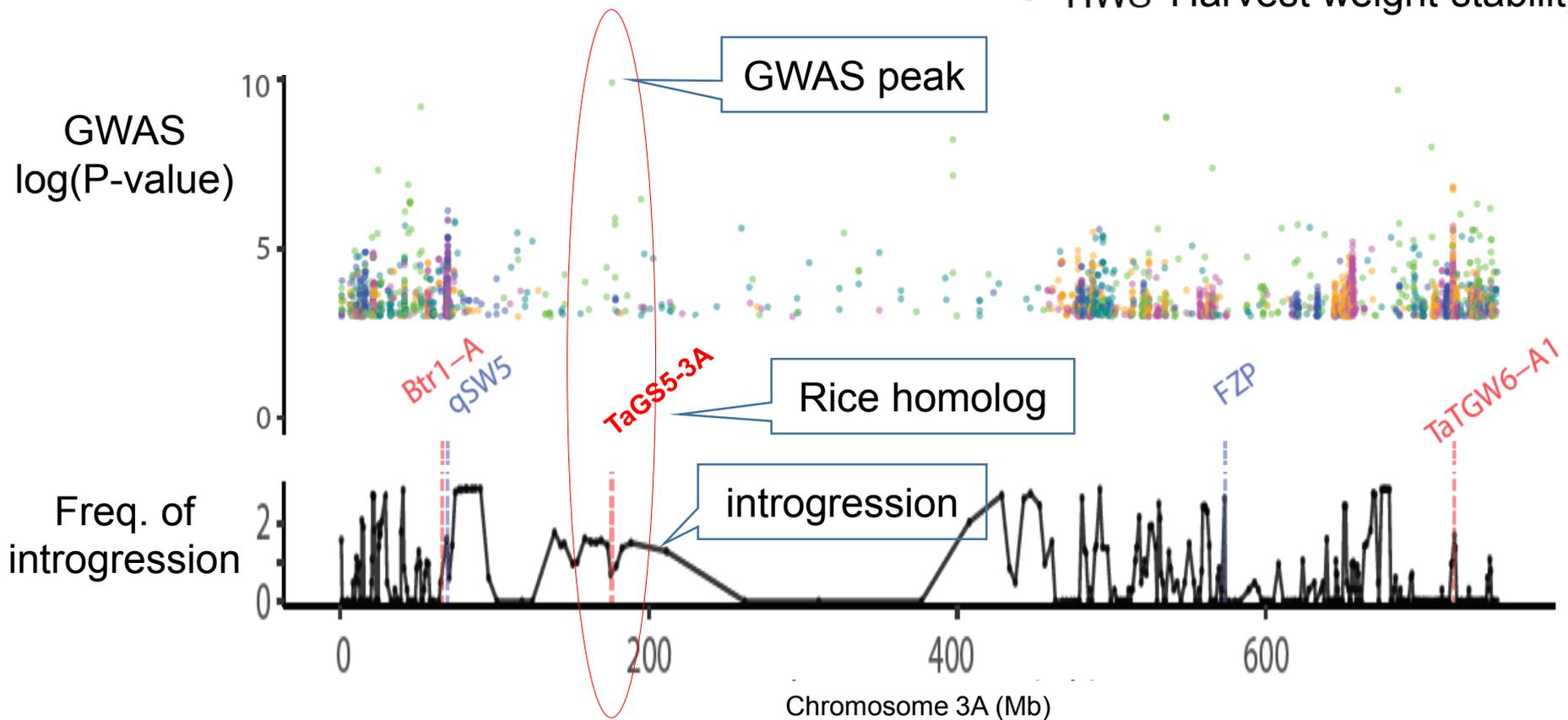
- GFP Grain filling period
- HD Heading date
- PHT Plant height
- HW Harvest weight
- HWS Harvest weight stability



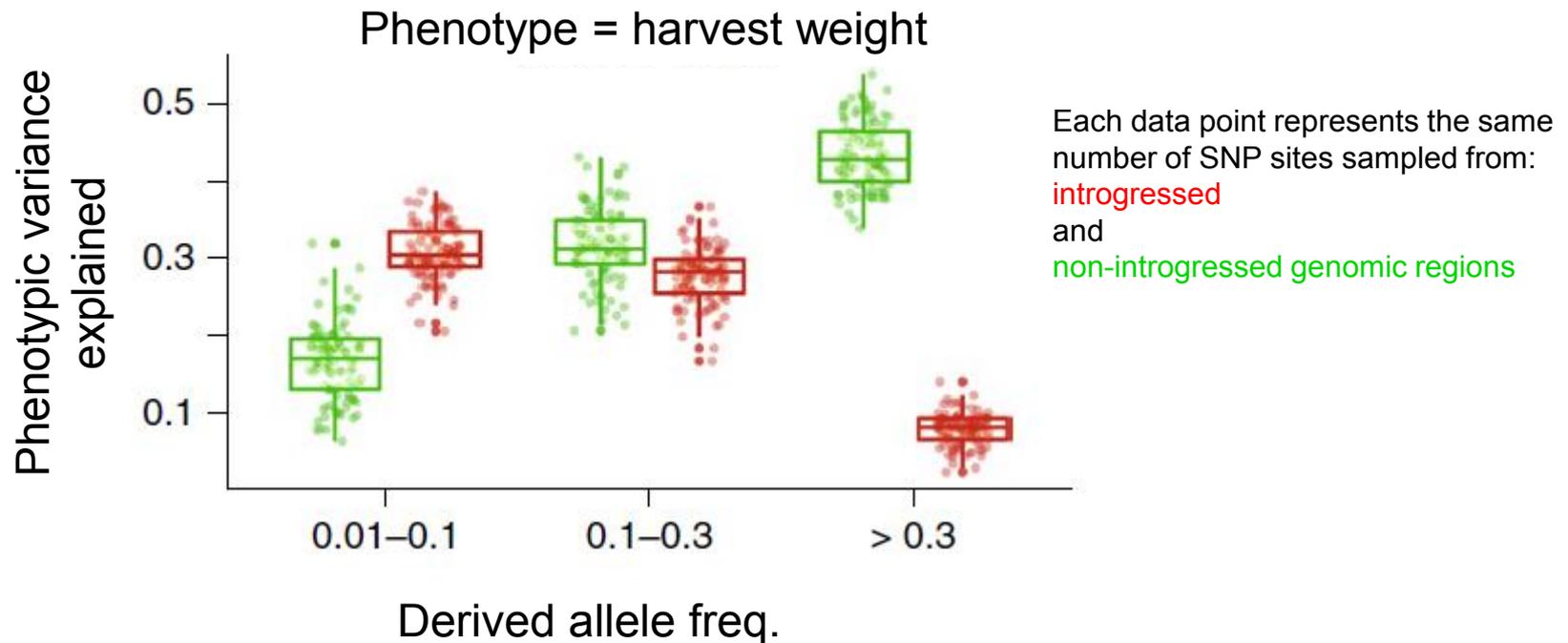
Introgression may contribute allele for agronomical traits

Evaluate the impact of introgression using GWAS results

- GFP Grain filling period
- HD Heading date
- PHT Plant height
- HW Harvest weight
- HWS Harvest weight stability



Introgressed genomic segments explain substantial proportion of phenotypic variance



- Selection against introgression
- Adaptive only in specific but geographically constrained habitats

Interpreting the functional role of introgression through comparative genomics and population genetics

GWAS hits ($p < 10^{-4}$)

- Grain filling period
- Heading date
- Plant height
- Harvest weight
- Harvest weight stability

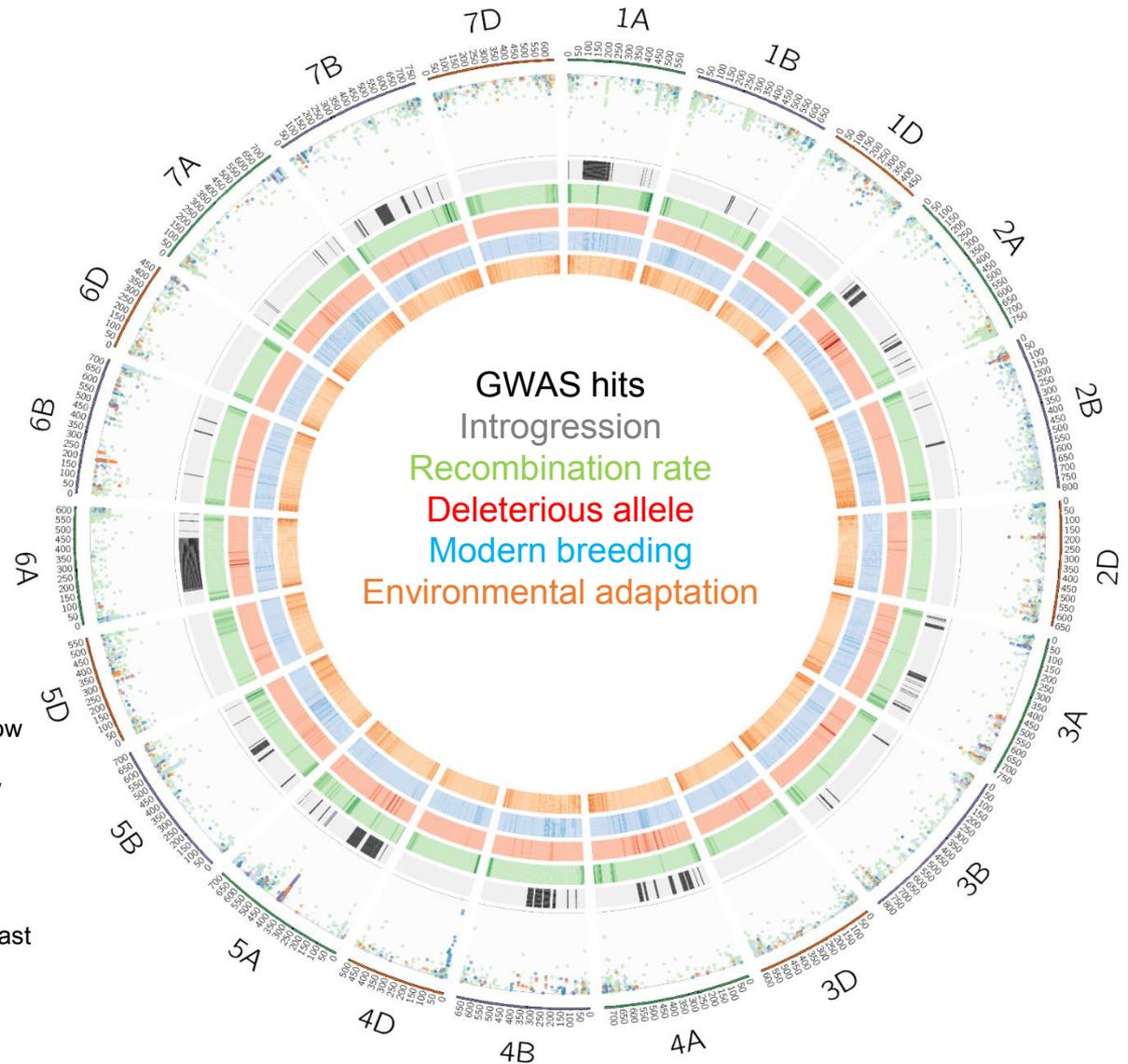
Introgression: Introgressed Genomic regions

Recombination rate: mean rate in 10Mb window

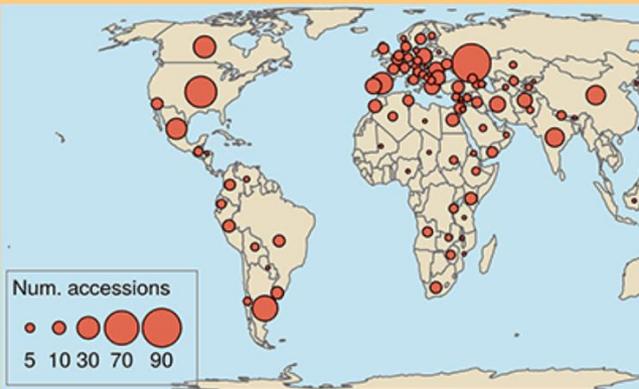
Deleterious allele: mean load in 10Mb window

Modern breeding: 1Mb window with top1% XPCLR outlier

Environmental adaptation: Windows with at least 4 bayenv top1% outlier SNP



Exome sequencing of a diverse collection of wheat landraces, cultivars and breeding lines



About 1000 wheat exomes project

Genome-level DNA sequence variation map is required to establish links between causal variants and phenotypes as well as to understand the role of environmental, demographic and human-driven factors in shaping the genomic diversity of modern wheat. Here, we used a reference wheat genome IWGSC RefSeq v1.0 to generate a haplotype map based on the targeted re-sequencing of more than 1,000 diverse wheat landraces and cultivars, and tetraploid wild and domesticated relatives.

Geographic distribution of the analyzed wheat lines

<http://wheatgenomics.plantpath.ksu.edu/1000EC/>

Acknowledgements

Kansas State University

Eduard Akhunov

Allan Fritz

**K-State Integrated Genomics
Facility (IGF)**

Alina Akhunova

**Agriculture Victoria, and,
La Trobe University**

Matthew Hayden

Raj Pasam

Fan Shi

Surya Kant

Gabriel Keeble-Gagnere

Pippa Kay

Kerrie Forrest

German Spangenberg

Ben Hayes

Hans Daetwyler

Josquin Tibbits

University of Saskatchewan

Curtis Pozniak

Pierre Hucl

Krystalee Wiebe

University of Minnesota

Peter L. Morrell

Corteva Agriscience

John P. Davies

Steve R. Webb

**Swift Current Research and
Development Centre**

Ron Knox

Richard Cuthbert

Agriculture Victoria Services

K-State Integrated Genomics Facility (IGF)

K-State **Beocat** High-Performance Computing cluster

Canadian *Triticum* Applied Genomics grant (CTAG)



AGRICULTURE VICTORIA

BILL &
MELINDA
GATES
foundation



United States Department of Agriculture
National Institute of Food and Agriculture

And many thanks to Akhunov Lab!