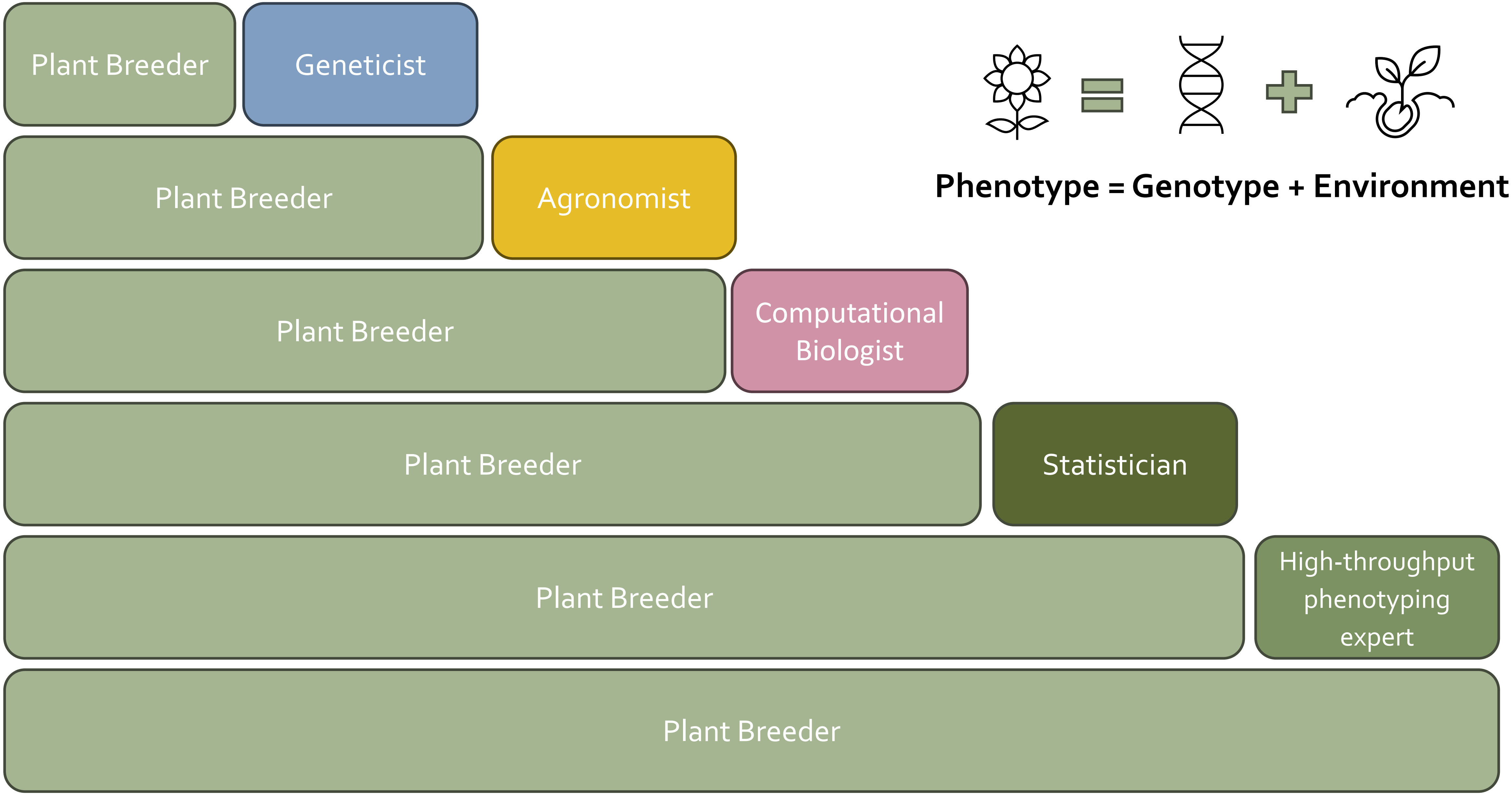
IDENTIFYING A CORE SET OF MARKERS AND HAPLOTYPE STRUCTURES TO TRACK SEGREGATION IN DOMESTICATED SUNFLOWER

Ashley C. Barstow
North Dakota State University
Dr. Brent Hulke

NDSU

Phenotype = Genotype + Environment



Plant Breeder | Geneticist

Plant Breeder | Agronomist

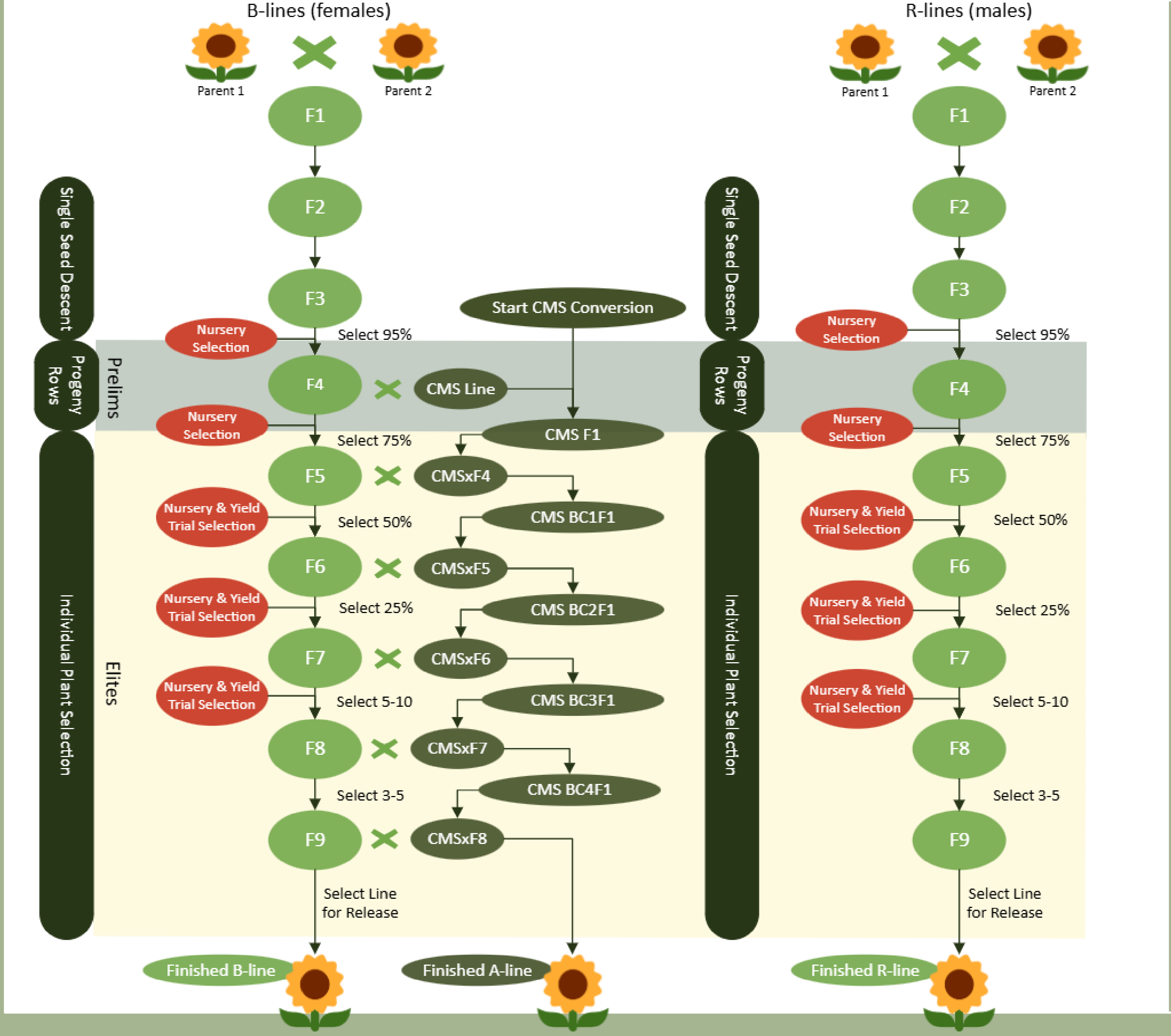
Plant Breeder | Computational Biologist

Plant Breeder | Statistician

Plant Breeder | High-throughput phenotyping expert

Plant Breeder

AN OVERVIEW OF THE SUNFLOWER BREEDING PROGRAM



B-lines (females)

Parent 1 × Parent 2

F1

F2

F3

Nursery Selection

Select 95%

F4 × CMS Line

Start CMS Conversion

Nursery Selection

Select 75%

CMS F1

F5 × CMSxF4

CMS BC1F1

Nursery & Yield Trial Selection

Select 50%

F6 × CMSxF5

CMS BC2F1

Nursery & Yield Trial Selection

Select 25%

F7 × CMSxF6

CMS BC3F1

Nursery & Yield Trial Selection

Select 5-10

F8 × CMSxF7

CMS BC4F1

Select 3-5

F9 × CMSxF8

Select Line for Release

Finished B-line

Finished A-line

Single Seed Descent

Progeny Rows

Prelims

Individual Plant Selection

Elites

R-lines (males)

Parent 1 × Parent 2

F1

F2

F3

Nursery Selection

Select 95%

F4

Nursery Selection

Select 75%

F5

Nursery & Yield Trial Selection

Select 50%

F6

Nursery & Yield Trial Selection

Select 25%

F7

Nursery & Yield Trial Selection

Select 5-10

F8

Select 3-5

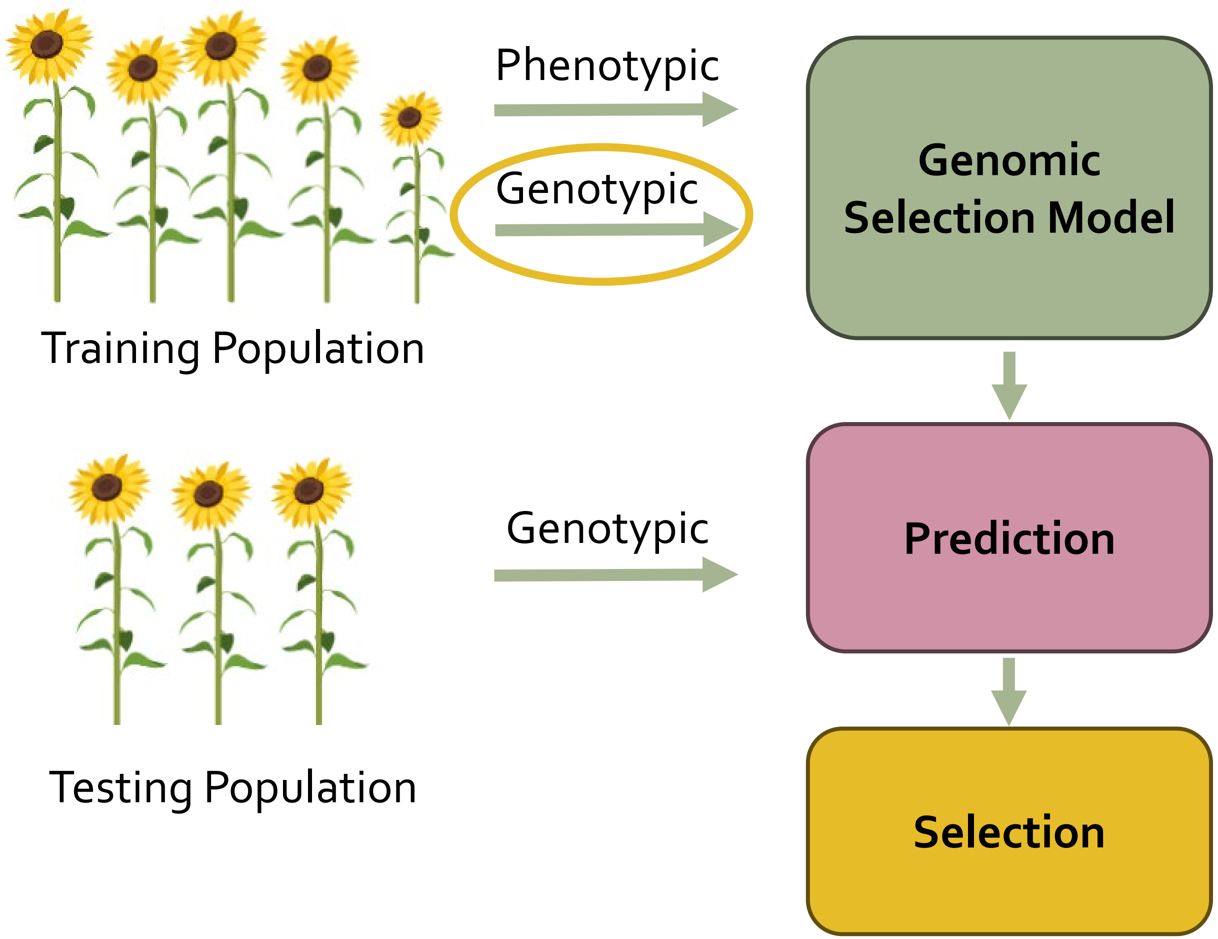
F9

Select Line for Release

Finished R-line

Modified from Brady Koehler

A FASTER STRATEGY FOR PLANT BREEDING: GENOMIC SELECTION



Training Population

Phenotypic

Genotypic

Genomic Selection Model

Testing Population

Genotypic

Prediction

Selection

Objectives:

1. Develop a set of genomic markers in sunflower suitable for future genomic prediction efforts.
2. Identify haplotype structures within the sunflower genome using multiple different populations from the USDA sunflower breeding program.

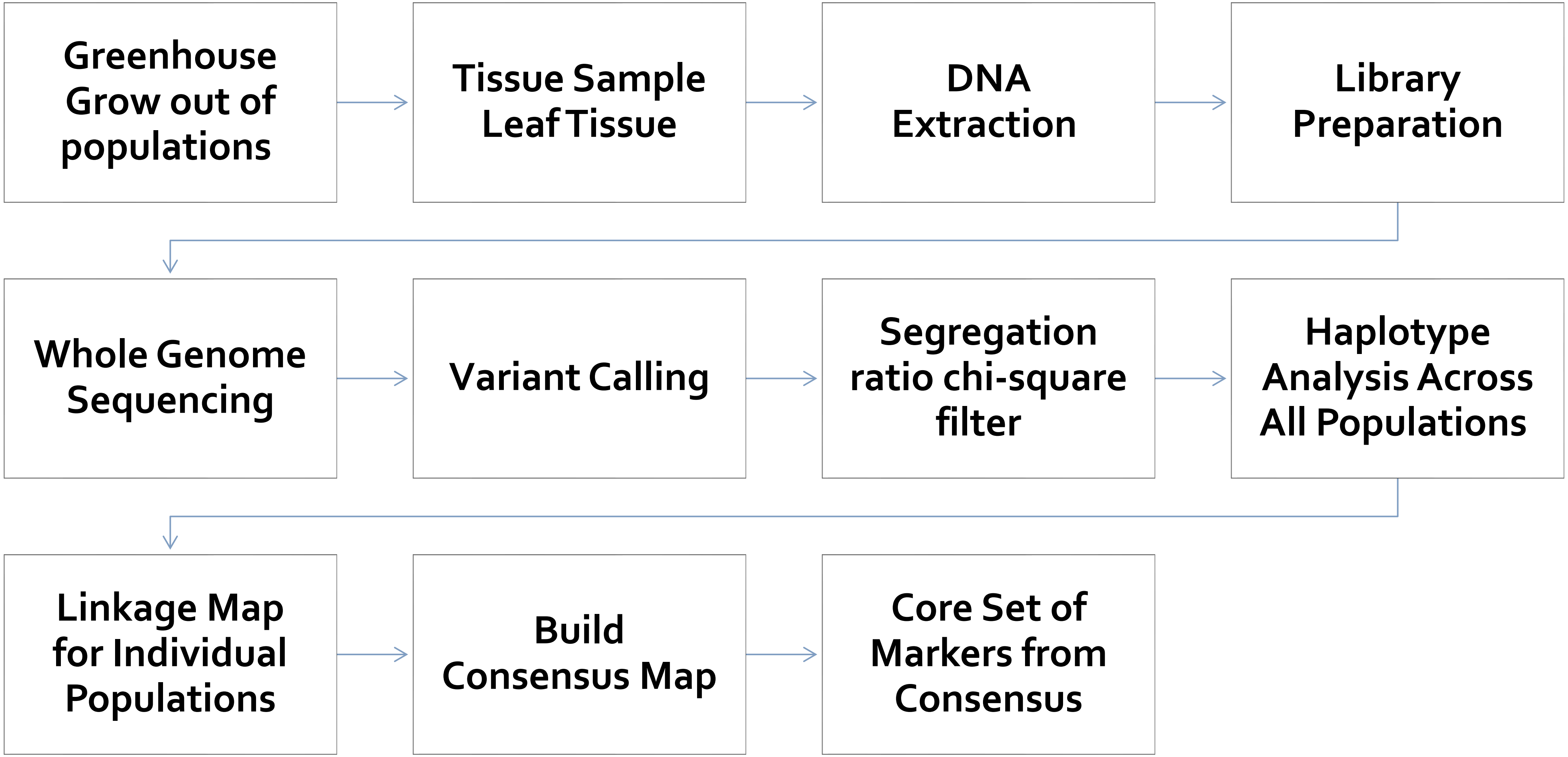
Objectives:

1. Develop a set of genomic markers in sunflower suitable for future genomic prediction efforts.
2. Identify haplotype structures within the sunflower genome using multiple different populations from the USDA sunflower breeding program.

Agronomically important traits and utility for future genomic studies

Pedigree	Population Type	B-line/ R-line	Oil	Disease Resistance	Other	Entries
16_1852/16_1410	F4	B-line	HO	Sclerotinia, Phomopsis	Drought, Herbicide Resistance (IMI)	76
HA 482/HA 487	F4	B-line	HO	Sclerotinia, Phomopsis	Herbicide Resistance (IMI)	33
RHA 474/R101dmr/6/RHA 344/NID HO//NS-H-924/4/RHA 418/RHA 419/3/RO12-13//RHA 274/PRS5/5/RHA 475/RHA 464	F4	R-line	HO	Sclerotinia, Downy Mildew, Rust	Herbicide Resistance (IMI)	45
RHA 476/IS 3401	F4	R-line	HO		Early Maturity	61
RHA 478/16_1570	F4	R-line	HO	Sclerotinia, Phomopsis	Herbicide Resistance (IMI)	70
RHA 483/RHA 479	F4	R-line	HO	Sclerotinia, Phomopsis	Herbicide Resistance (IMI)	28
RHA 483/RHA 480	F4	R-line	HO	Sclerotinia, Phomopsis	Herbicide Resistance (IMI)	47
RHA 485/4/RHA 473/3/RHA 462//PET-1gammarays/6116	F4	R-line	HO	Sclerotinia		33
RHA 484//RHA 472/R101dmr	F4	R-line	HO	Sclerotinia, Downy Mildew, Rust	Herbicide Resistance (IMI)	56
RHA 485/RHA 476	F4	R-line	HO	Phomopsis	High Oil, Early Maturity	101
RHA 486/RHA 476	F4	R-line	HO	Sclerotinia, Phomopsis, Downy Mildew	Early Maturity	50
RHA 476/687	F4	R-line	HO		Early Maturity	72
HA 456/HA 434	F6 Nectar Mapping	B-line				191
HA 467/PI170415	F7-9 pericarp Mapping	B-line				159
Total						1022

Methodology



Greenhouse Grow out of populations

Tissue Sample Leaf Tissue

DNA Extraction

Library Preparation

Whole Genome Sequencing

Variant Calling

Segregation ratio chi-square filter

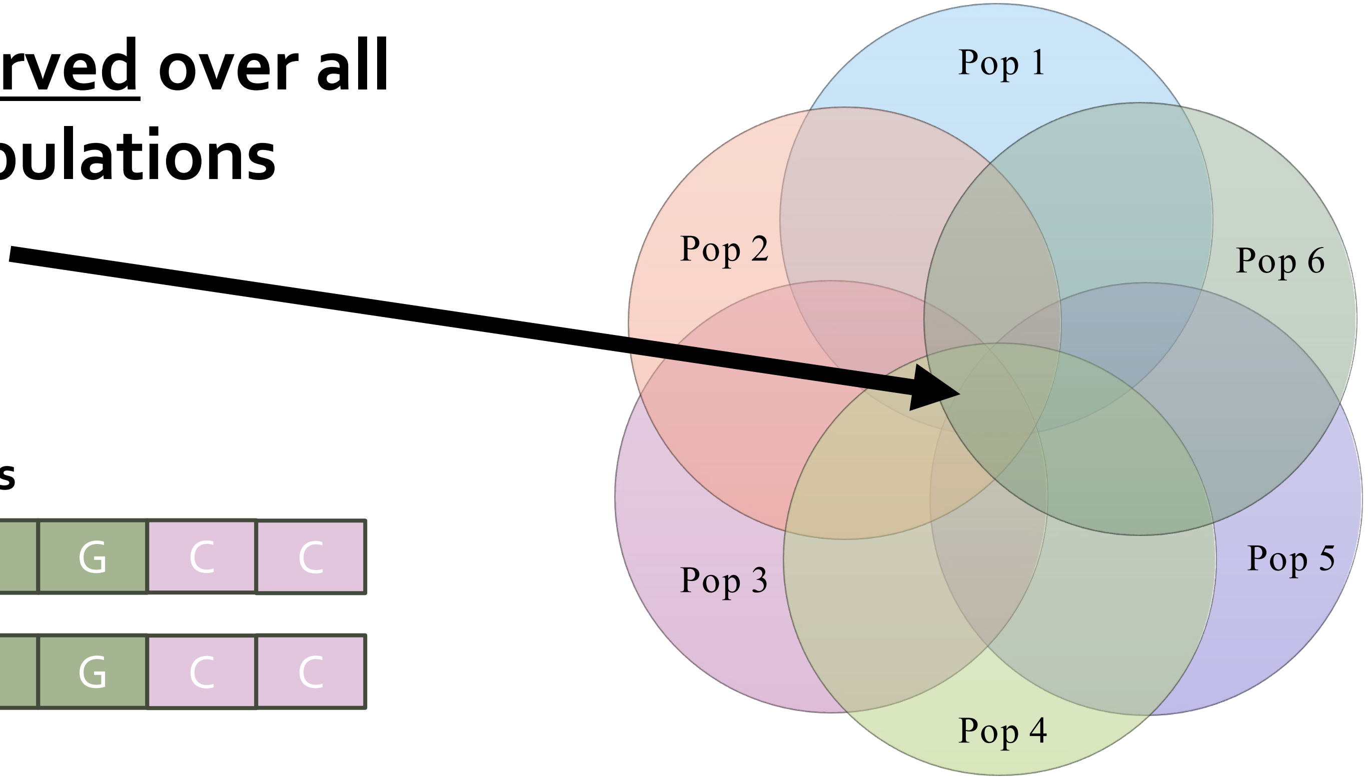
Haplotype Analysis Across All Populations

Linkage Map for Individual Populations

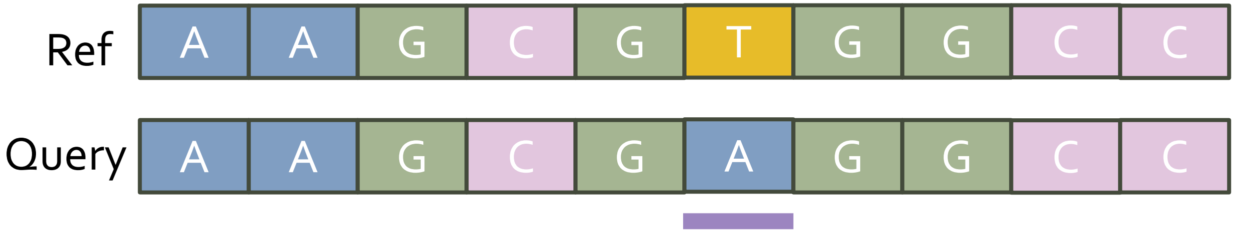
Build Consensus Map

Core Set of Markers from Consensus

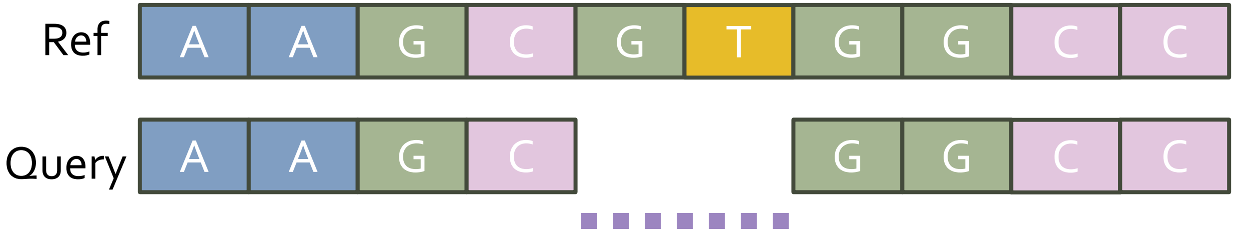
Variants need to be conserved over all breeding/mapping populations



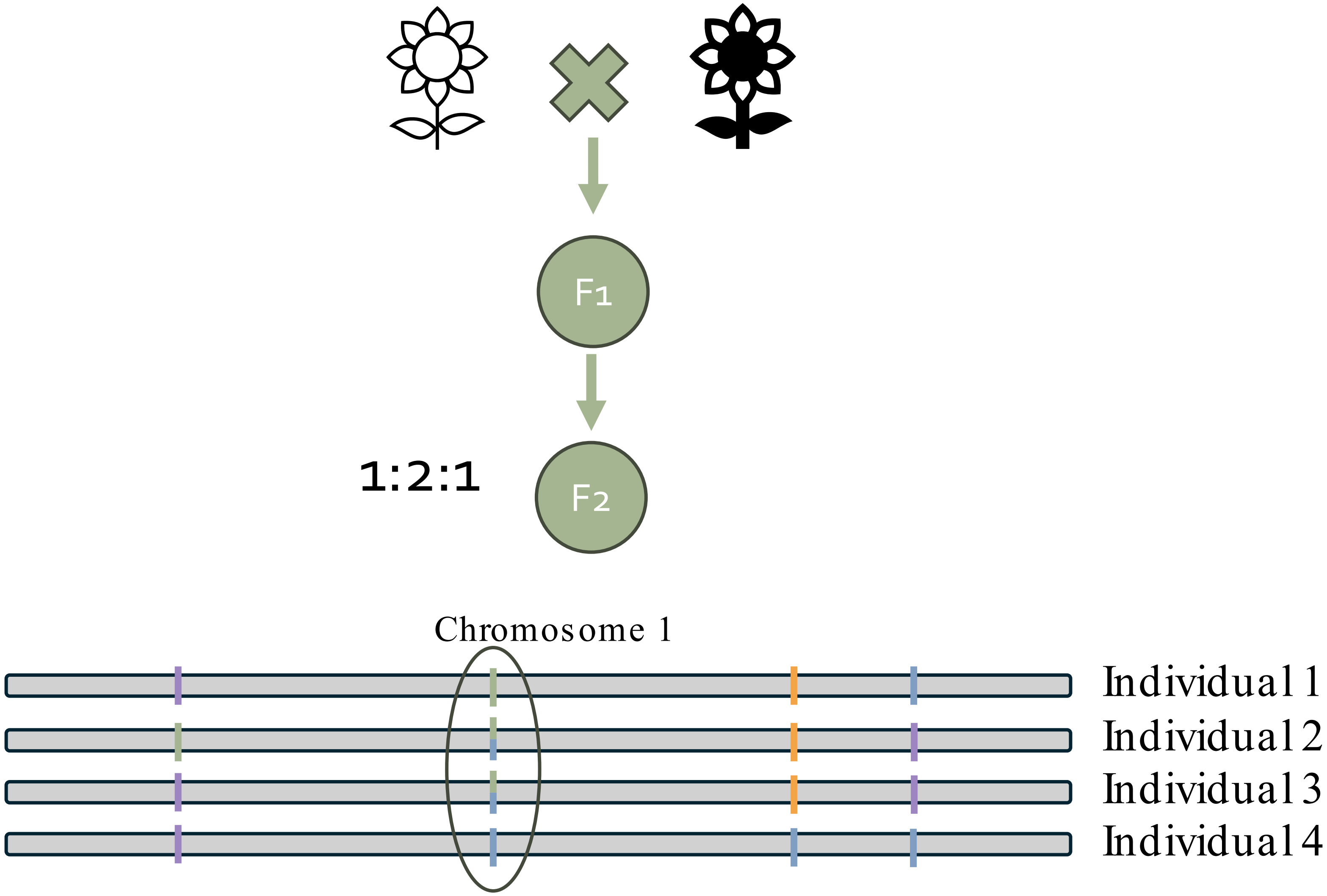
Sequence variants



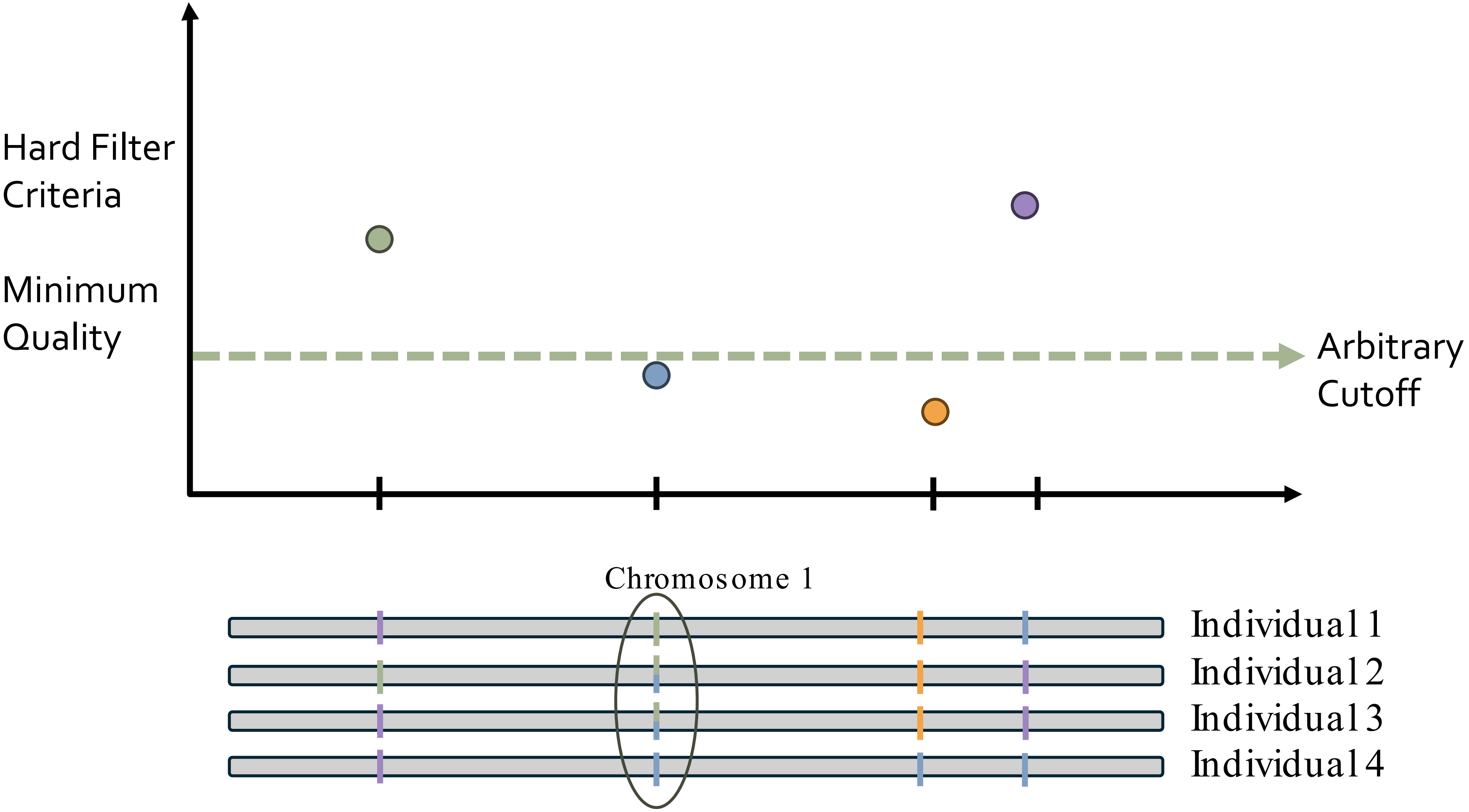
INDEL (Insertion or Deletion)



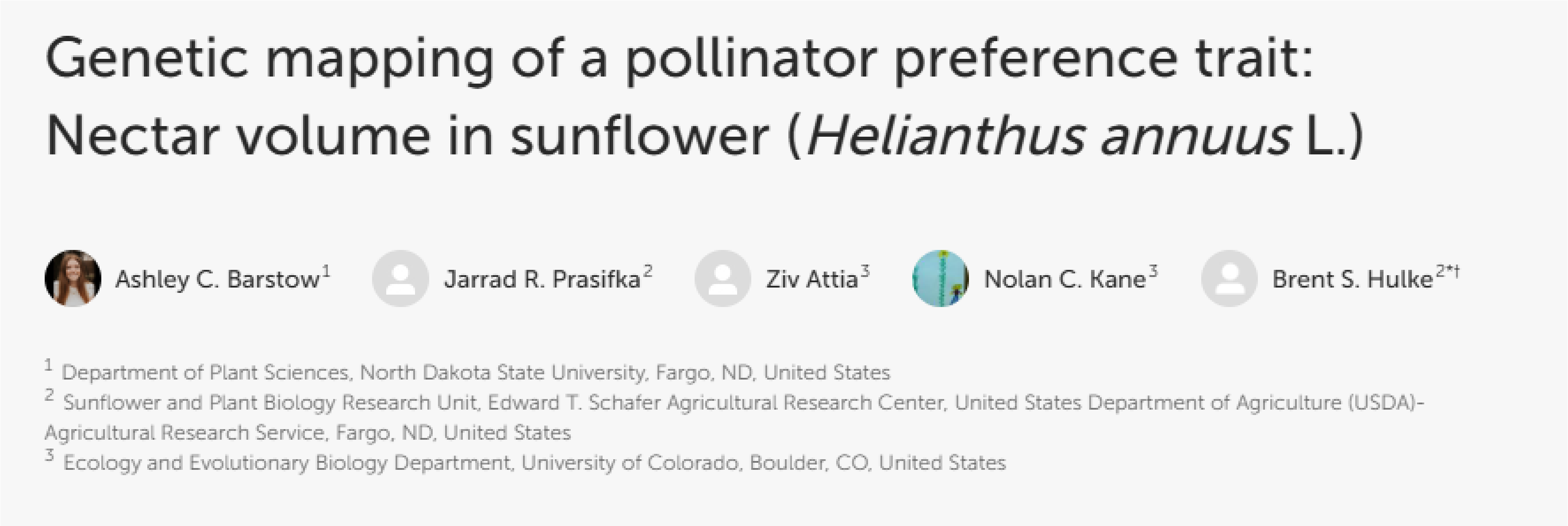
Fundamental Mendelian Ratios can be used to find Reproducible and Reliable Markers



Hard Filters for Variant Calling could remove valuable markers



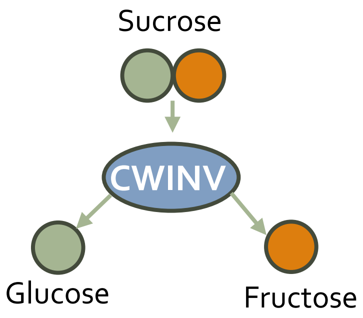
Case Study: Redo mapping without hard filters



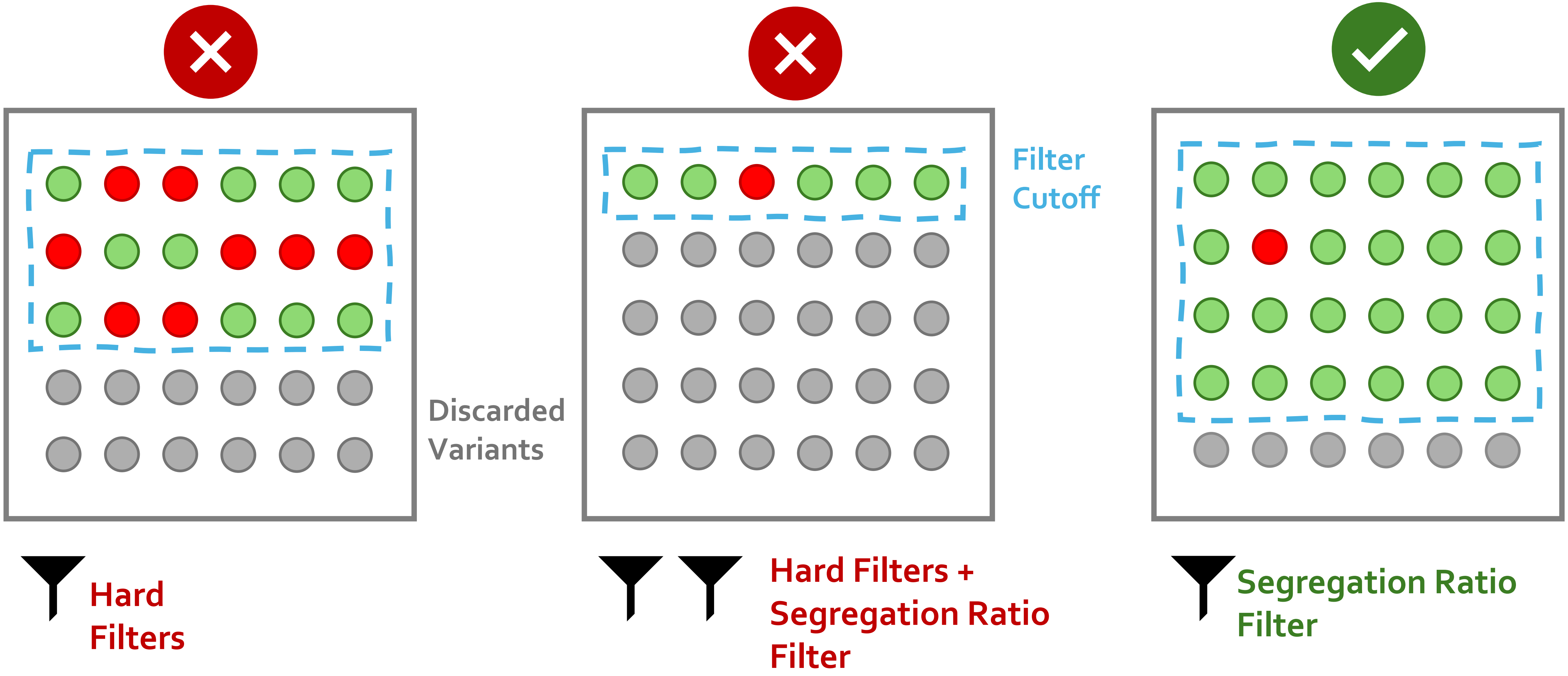
Genetic mapping of a pollinator preference trait: Nectar volume in sunflower (*Helianthus annuus* L.)

Ashley C. Barstow[1] Jarrad R. Prasifka[2] Ziv Attia[3] Nolan C. Kane[3] Brent S. Hulke[2*†]

[1] Department of Plant Sciences, North Dakota State University, Fargo, ND, United States
[2] Sunflower and Plant Biology Research Unit, Edward T. Schafer Agricultural Research Center, United States Department of Agriculture (USDA)-Agricultural Research Service, Fargo, ND, United States
[3] Ecology and Evolutionary Biology Department, University of Colorado, Boulder, CO, United States



A chi-square filtering strategy based on segregation ratio retained a **larger** number of valid markers with better genome coverage.



Discarded Variants

Filter Cutoff

Hard Filters

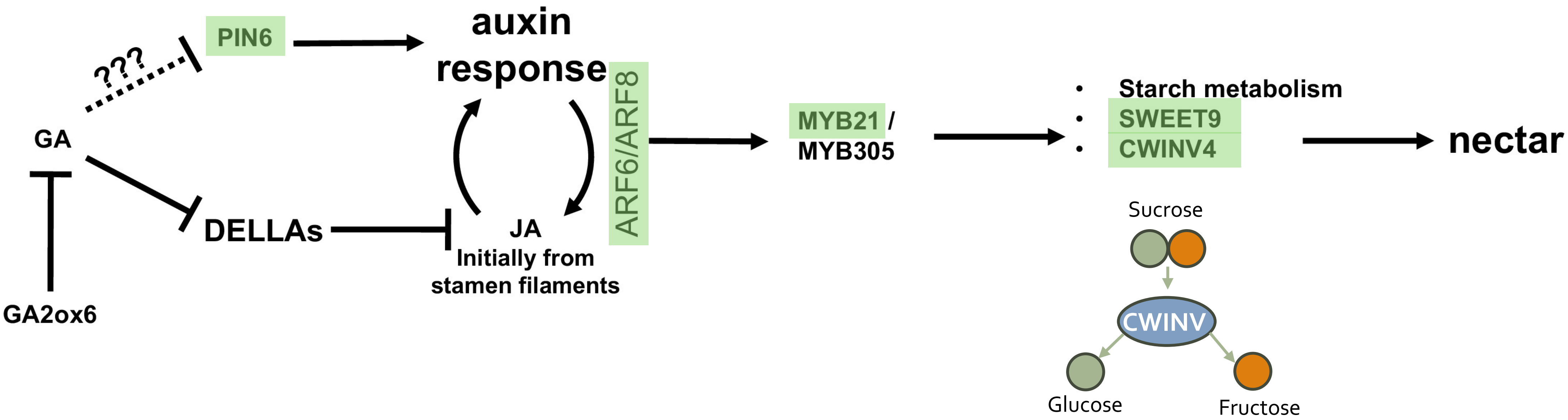
Hard Filters + Segregation Ratio Filter

Segregation Ratio Filter

More candidate regions with nectar-related genes were discovered because of genetic map **improvements**.

Model of Floral Nectary Regulation

Arabidopsis thaliana



PIN6

???

auxin response

ARF6/ARF8

GA

GA2ox6

DELLAs

JA

Initially from stamen filaments

MYB21 / MYB305

- Starch metabolism
- SWEET9
- CWINV4

nectar

Sucrose

CWINV

Glucose

Fructose

CRABS CLAW	Transcription factor needed for nectary development (Bowman & Smyth, 1999)
TIR1	Auxin receptor that mediates auxin-regulated transcription (Schmitt et al., 2018)

(Roy et al., 2017)

Methodology

Greenhouse Grow out of populations

Tissue Sample Leaf Tissue

DNA Extraction

Library Preparation

Whole Genome Sequencing

Variant Calling

Segregation ratio chi-square filter

Haplotype Analysis Across All Populations

Linkage Map for Individual Populations

Build Consensus Map

Core Set of Markers from Consensus

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

New Results

Variant filters using segregation information improve mapping of nectar-production genes in sunflower (*Helianthus annuus* L.)

Ashley C. Barstow, James P. McNellie, Brian C. Smart, Kyle G. Keepers, Jarrad R. Prasifka, Nolan C. Kane, Brent S. Hulke

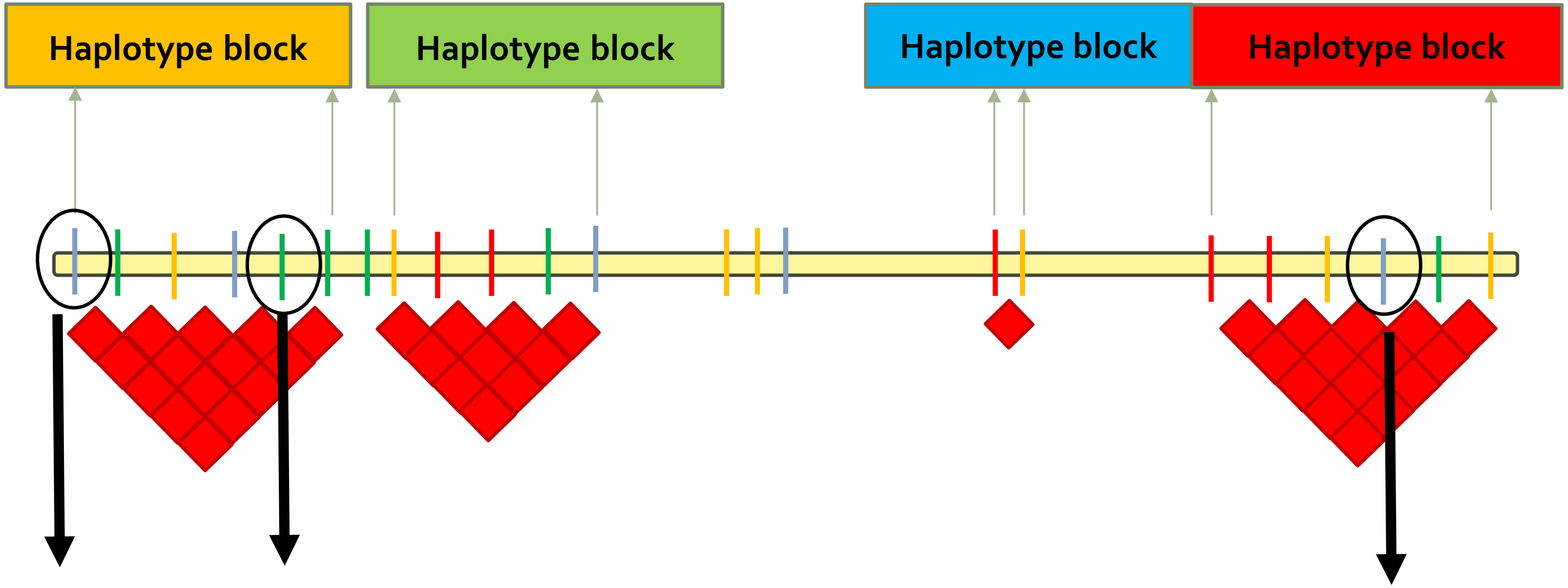
doi: https://doi.org/10.1101/2024.12.03.626666

This article is a preprint and has not been certified by peer review [what does this mean?].

Objectives:

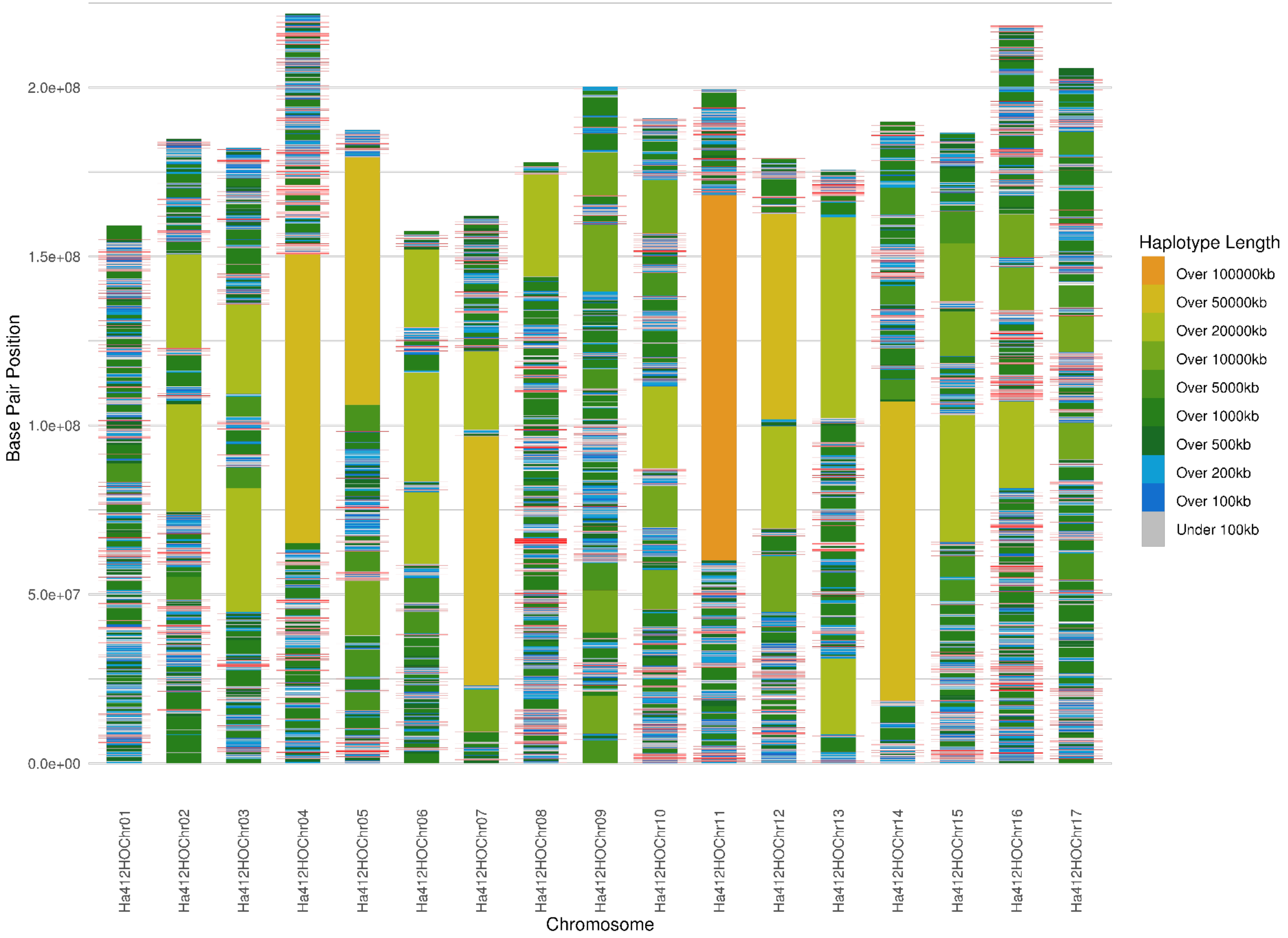
1. Develop a set of genomic markers in sunflower suitable for future genomic prediction efforts.
2. Identify haplotype structures within the sunflower genome using multiple different populations from the USDA sunflower breeding program.

Haplotype Block discovery to reduce marker numbers and improve genomic selection accuracy



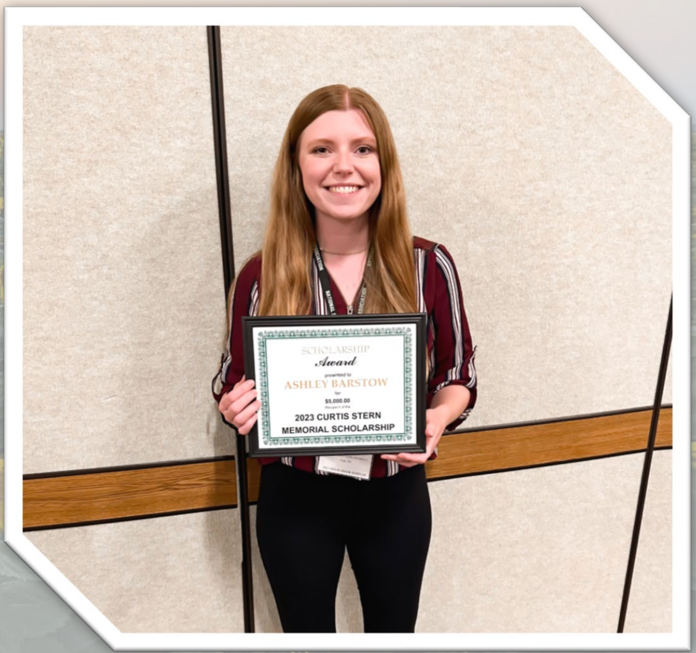
These variants are in <u>strong LD</u> with the rest of the markers in the block and can be used to characterize the genetic diversity of the entire block to reduce the marker set needed for future genomic work.

Haplotype Blocks across all Chromosomes and Populations



Acknowledgments

- Dr. Brent Hulke, USDA-ARS, Fargo, ND
- Jarrad Prasifka, USDA-ARS, Fargo, ND
- Nolan Kane, University of Colorado – Boulder
- Brian Smart, NDSU, Fargo, ND
- James McNellie, USDA-ARS, Fargo, ND
- Brady Koehler, USDA-ARS Fargo, ND
- Kyle Keepers, University of Colorado – Boulder
- Zachary Marcus, University of Colorado – Boulder
- All other graduate students and interns



QUESTIONS?

NDSU NORTH DAKOTA STATE UNIVERSITY

USDA ars Agricultural Research Service

University of Colorado **Boulder**

National Sunflower ASSOCIATION