Enhancing Genetic Insights with Polyploid Tools and User-friendly Interfaces

Cris Taniguti
cht47@cornell.edu

Genomics Coordinator
Breeding Insight, Cornell University

Plant Science Symposium – University of Florida

January 21, 2025

Outline

- About Breeding Insight
- Tools for Polyploids Project
- Ploidy Estimation
- Updates on Linkage Map Packages
- Microhaplotypes
- User-Friendly Interfaces

Breeding Insight

Mission

Transform breeding
by enabling the implementation
of **genomic insight and selection**
as part of **routine specialty crop and livestock breeding** across **all of USDA-ARS**.

About Breeding Insight

Who we are:

- We connect and build resources for specialty crop and livestock breeders.
- We are a USDA-ARS-funded initiative located at Cornell University.
- We are a service and technology transfer model providing additional support to USDA specialty breeding programs.
- We are a team made up of scientists and software developers.

About Breeding Insight

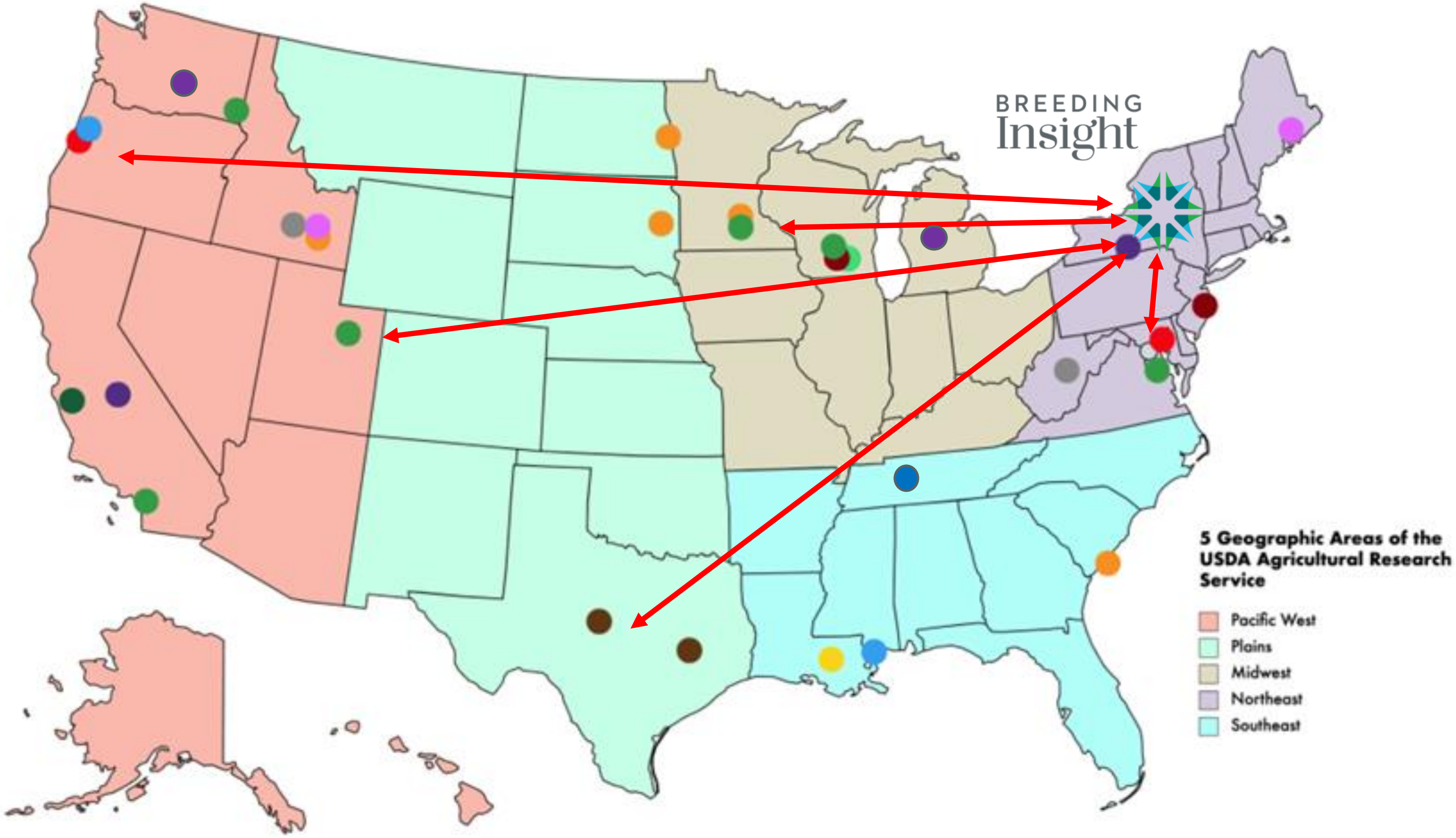
What we are not:

- We are not a private company.
- We are not interested in duplicating established technology, techniques, or patents.
- We are not selling and we are not receiving commission for the technology we provide or recommend.

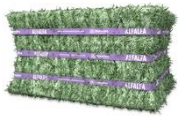
Breeding Insight Tools and Capacity

- Phenomics
- Genomics
- Bioinformatics
- Statistics
- Software Developers
- Image Analysis
- Data QC
- Pedigree
- Marker Validation

27 species
59 breeding programs



Exploring Resistance Traits



- ➢ Salt resistance
- ➢ Drought Stress tolerance
- ➢ pH tolerance
- ➢ Aphanomyces resistance
- ➢ Anthracnose resistance

- ➢ Potato Cys Nematore (PCN) resistance

- ➢ Nematode resistance
- ➢ Fusarium resistance
- ➢ Wireworm, Diabrotica, and Systena (WDS) resistance
- ➢ Insect resistance

- ➢ Mite resistance

- ➢ Powdery mildew resistance

- ➢ Powdery mildew resistance

- ➢ Impatiens necrotic spot virus resistance

- ➢ Powdery mildew resistance
- ➢ Downy mildew resistance
- ➢ Cold resistance

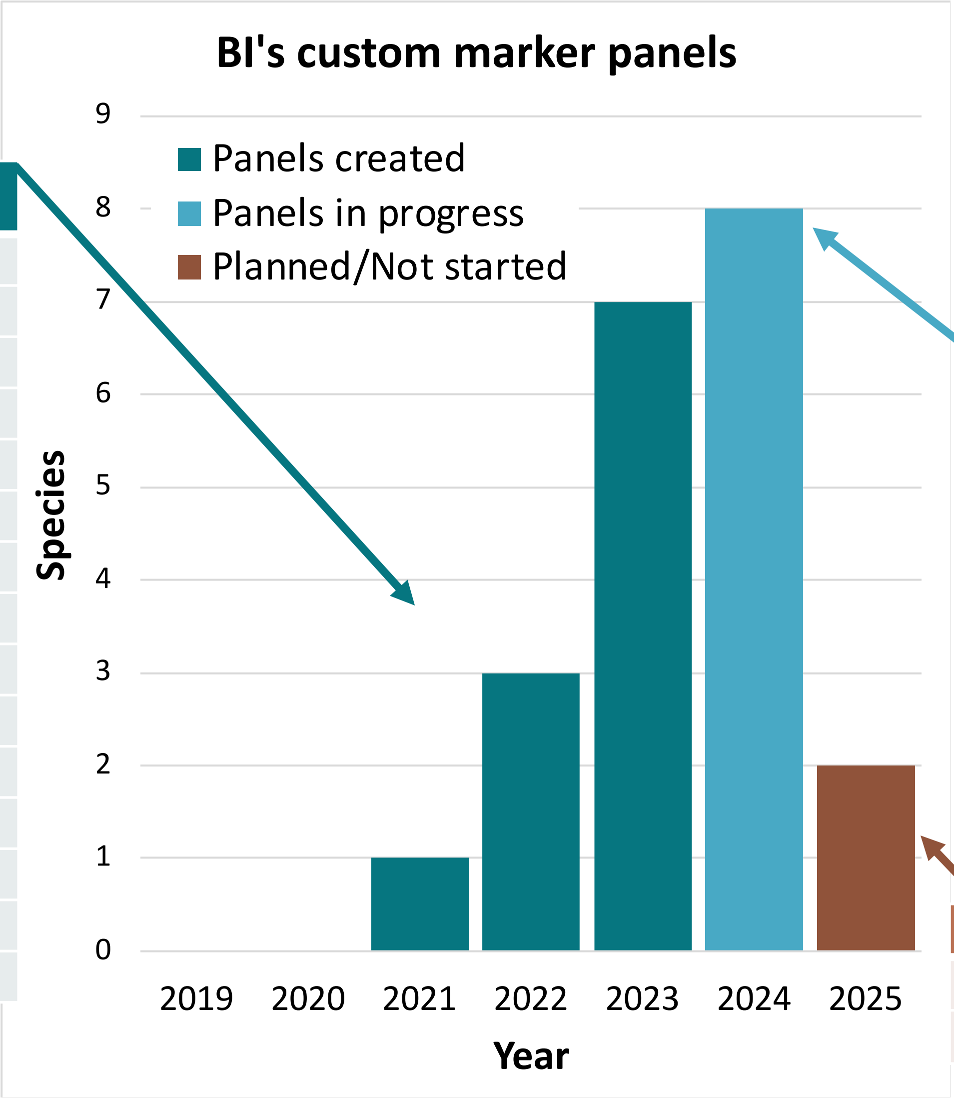
- ➢ Shock virus resistance

- ➢ Resilience to changes in water quality and pathogen load

Genomic Resources

*Purchased by collaborators
**Designed and purchased by collaborators
***Co-designed with collaborators

Species	Available Panels
Grape	2K rhAmpSeq**
Alfalfa	3K DArTag
Blueberry	3K DArTag
NA Atlantic Salmon	50K Infinium array*
NA Atlantic Salmon	3K DArTag
Lettuce	3K DArTag
Cucumber	3K DArTag
Pecan	3K DArTag
Cranberry	3K DArTag
Honeybee	81 DArTmp
Sweetpotato weevil	101 DArTmp
Sweetpotato	3K DArTag
Strawberry	3K DArTag**
Strawberry	5K DArTag**
Cranberry v2.0	3K DArTag



- ☑ Accessible
- ☑ Reliable
- ☑ Fast
- ☑ Affordable
- ☑ Repeatable

Species	In Progress Panels
Grape & Muscadine	3K DArTag***
Hop	3K DArTag
Hemp	3K DArTag
Blackberry	3K DArTag
Red clover	3K DArTag
Lettuce (parentage ver.)	100 KASP
Honeybee	105 AgriSeq
Trout	GTSeq
Sugar beet	5K AgriSeq**

Species	Planned Panels
Raspberry	3K DArTag
Hydrangea	3K DArTag

Genomic Challenges for Specialty Species

Haplodiploid
1x & 2x

Heteromorphic sex chromosomes

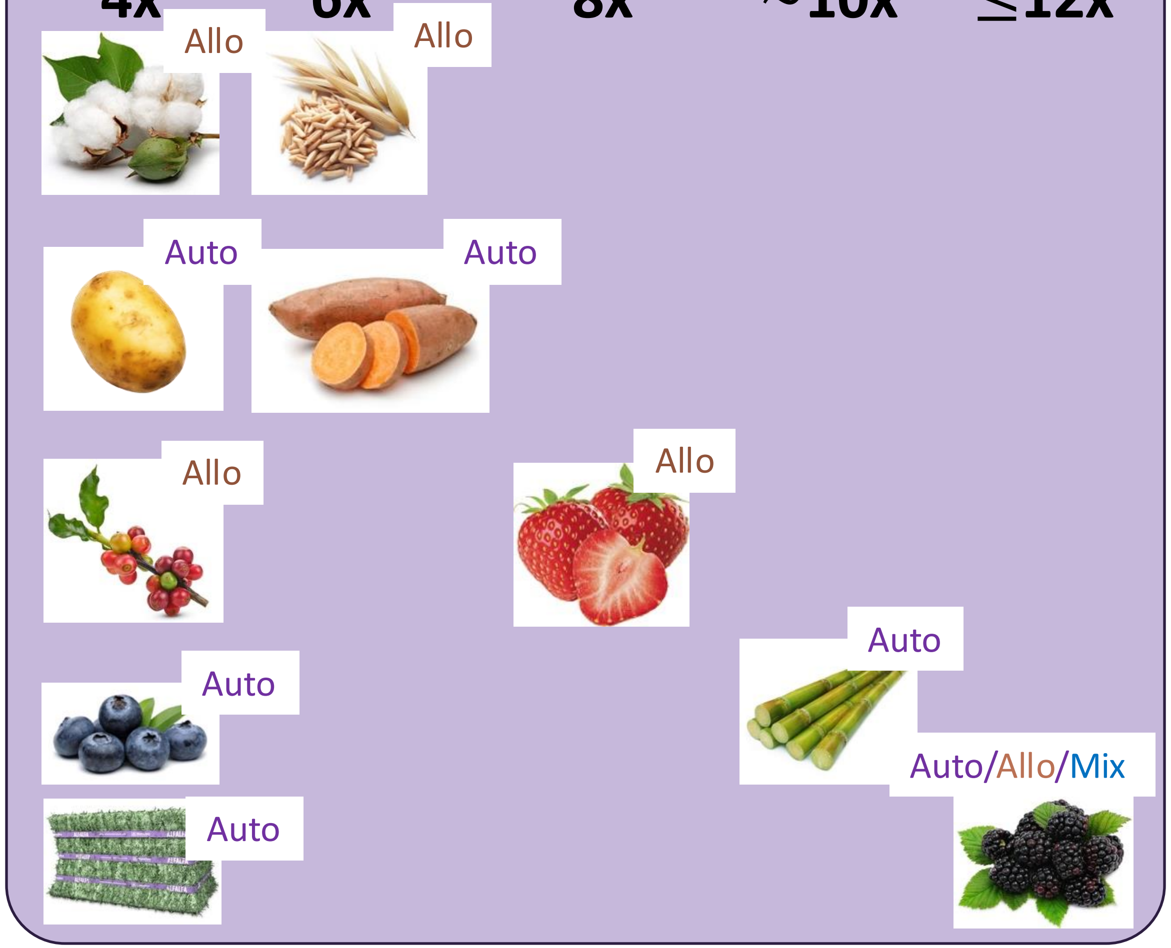
Non-binary sexes

Allopolyploid
Autopolyploid
Mixed (some level of preferential paring)

Diploids
2x

Polyploids

4x **6x** **8x** **~10x** **≤12x**



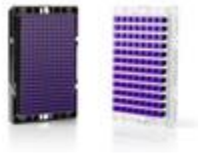
Tools for Polyploids Project

- Computational Tools Development and Improvement
- Collaboration between developers and breeders
- Website: **www.polyploids.org**
- Workshops 2021 – 2024 (records available)

Genetic Analysis Tools for Polyploid

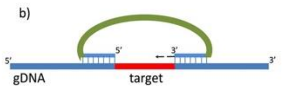
Genotyping Array

Axiom Array
Illumina



Target Genotyping

DArTag
GTseq
AgriSeq

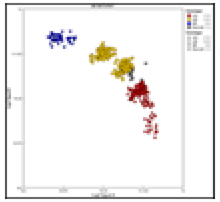


Genome-wide

GBS
WGS

Genotype calling

fitPoly
polyRAD
updog



Micro-haplotypes

Mchap
MicroHapDB

Reads processing

SNP calling
Genotype calling
comparison

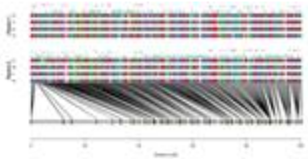
Reads2Map

Ploidy Estimation

Qploidy

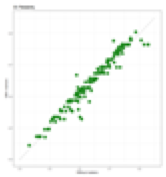
Linkage mapping/haplotyping

MAPpoly PolyOrigin
polymapR MAPpoly2



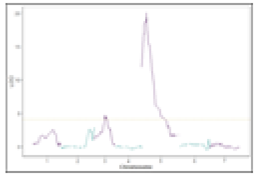
Genomic prediction

StageWise



QTL analysis

QTLpoly
polyqtlR
diaQTL
GWASpoly

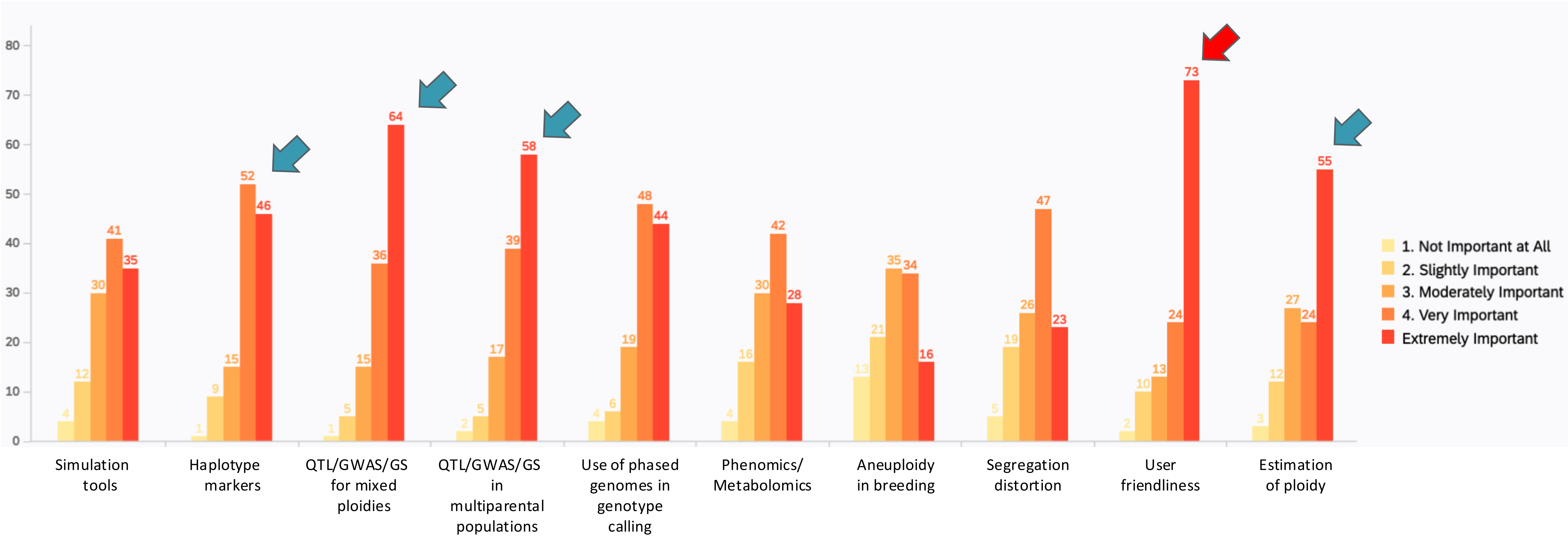


User-friendly interfaces

VIEWpoly
BIG app
HaploSearch

Tools for Polyploids Project Survey

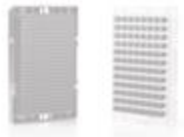
David Byrne
Texas A&M University



Genetic Analysis Tools for Polyploid

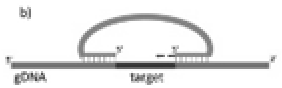
Genotyping Array

Axiom Array
Illumina



Target Genotyping

DArTag
GTseq
AgriSeq

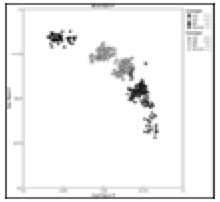


Genome-wide

GBS
WGS

Genotype calling

fitPoly
polyRAD
updog



Micro-haplotypes

Mchap
MicroHapDB

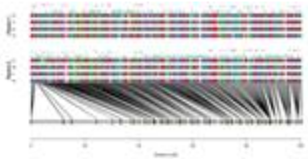
Reads processing

SNP calling
Genotype calling
comparison

Ploidy Estimation

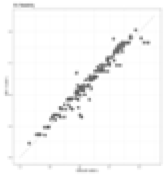
Linkage mapping/haplotyping

MAPpoly PolyOrigin
polymapR **MAPpoly2**



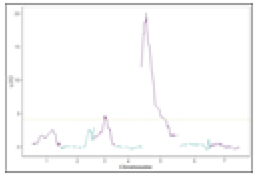
Genomic prediction

StageWise



QTL analysis

QTLpoly
polyqtlR
diaQTL
GWASpoly



User-friendly interfaces

HaploSearch

Ploidy Estimation

- Requisite for all downstream analysis
- Texas A&M postdoc
  - Roses
  - Axiom Array data
  - 89% agreement
  - 0.83 Cohen's Kappa coefficient
  - MS submitted!
  - Package and tutorial already available!

Cristianetaniguti/Qploidy


Oscar Riera-Lizarazu


Tessa Hocchaus


Jeekin Lau

		Based on past reports		
Based on Qploidy	Ploidy level	2x	3x	4x
	2x	78	1	0
	3x	4	27	0
	4x	2	15	90
	5x	0	0	1

Input File Formats

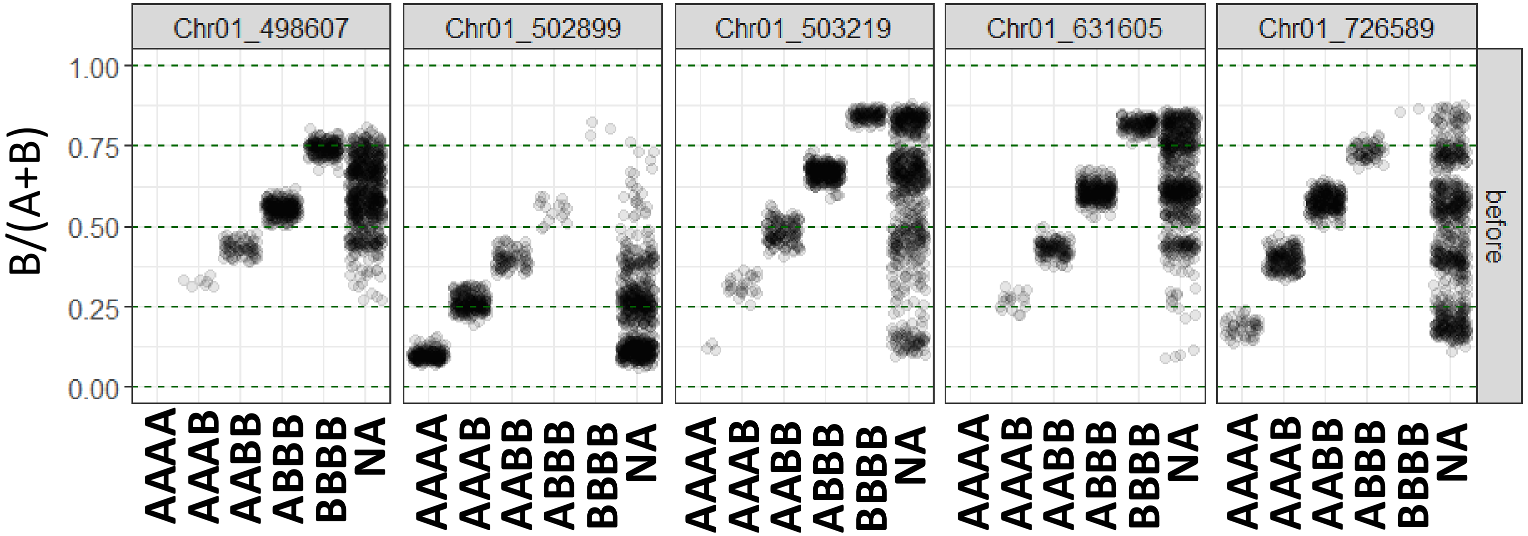
Sequencing	Array
VCF (w/allele depth)	Illumina Axiom

Qploidy Method

- B allele frequency= B/(A+B)

1 marker: Total depth/intensity = 100; ideal ratios:

AAAA	**AAAB**	**AABB**	**ABBB**	**BBBB**
A = 100	A = 75	A = 50	A = 25	A = 0
B = 0	B = 25	B = 50	B = 75	B = 100
Ratio = 0	Ratio = 0.25	Ratio = 0.5	Ratio = 0.75	Ratio = 1

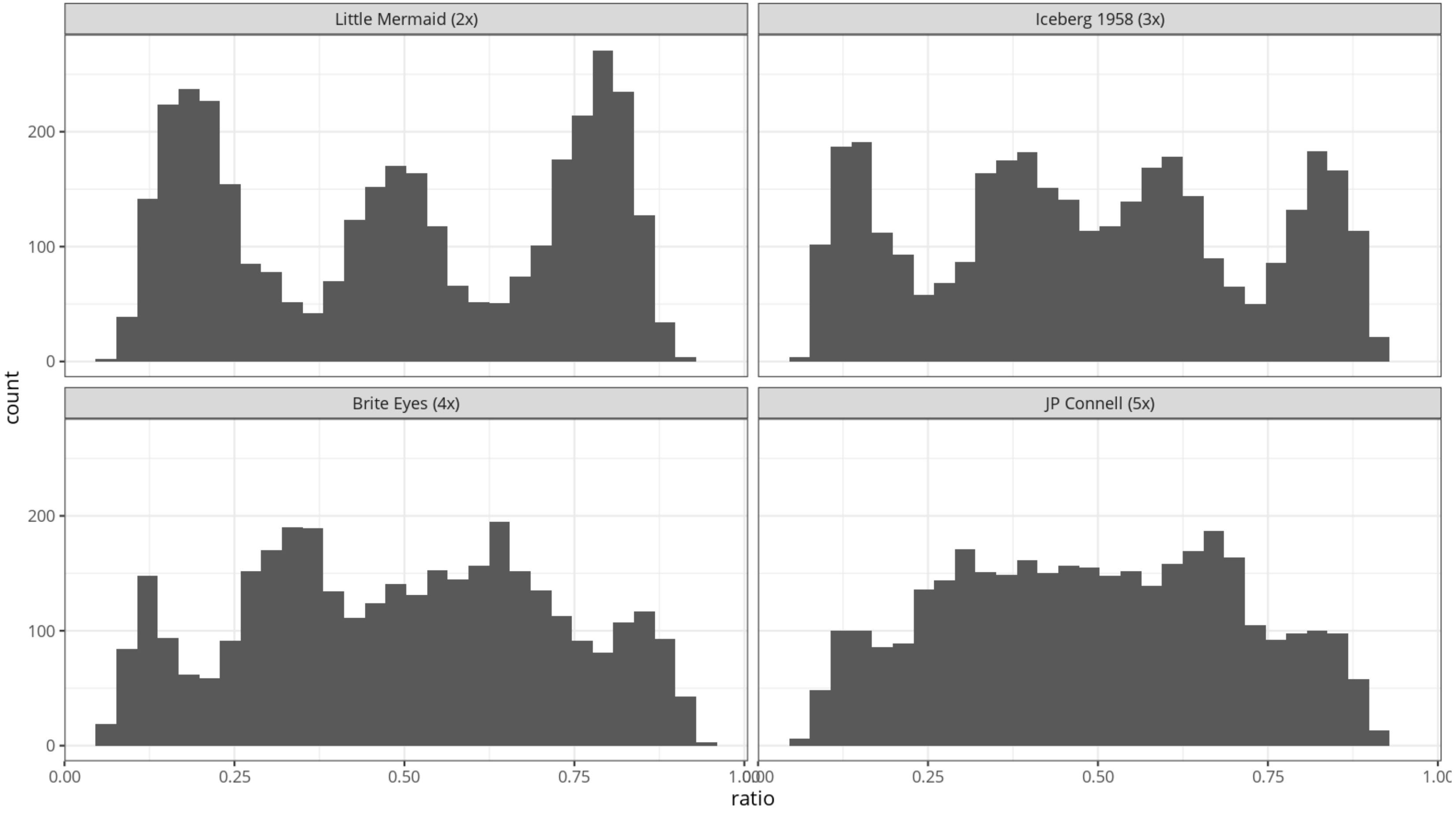


Reality for Arrays and Target Sequencing:

- High variation in marker-to-marker ratio
- Low variation in sample-to-sample ratio

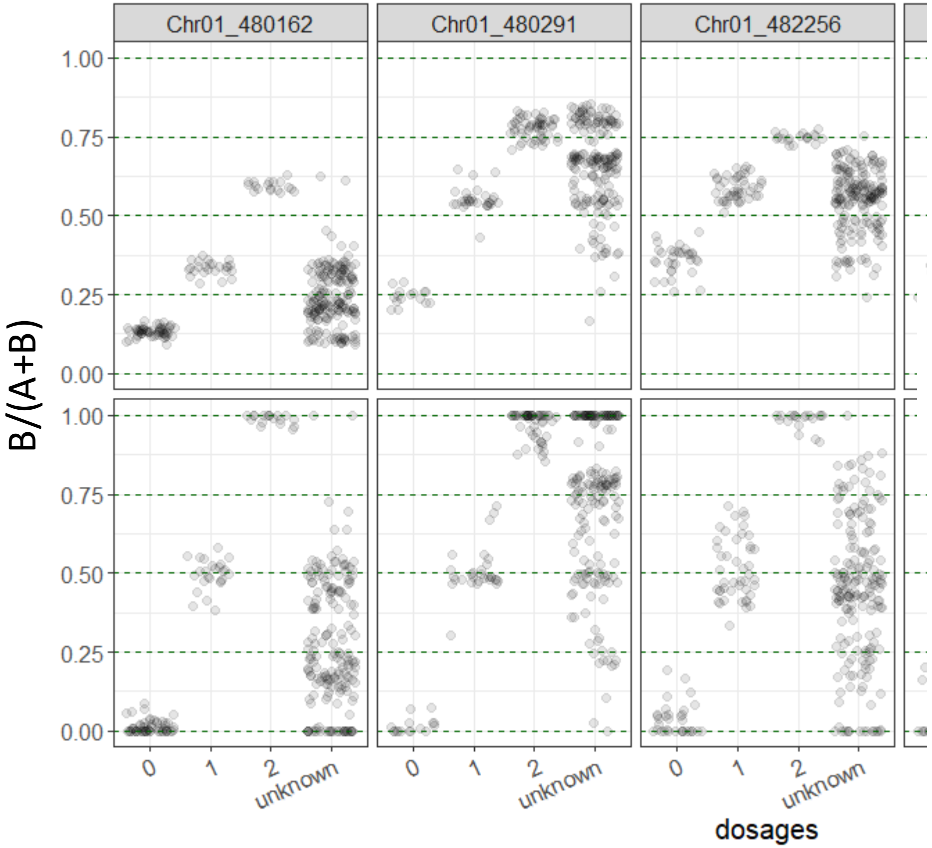
Sample-level ploidy determination

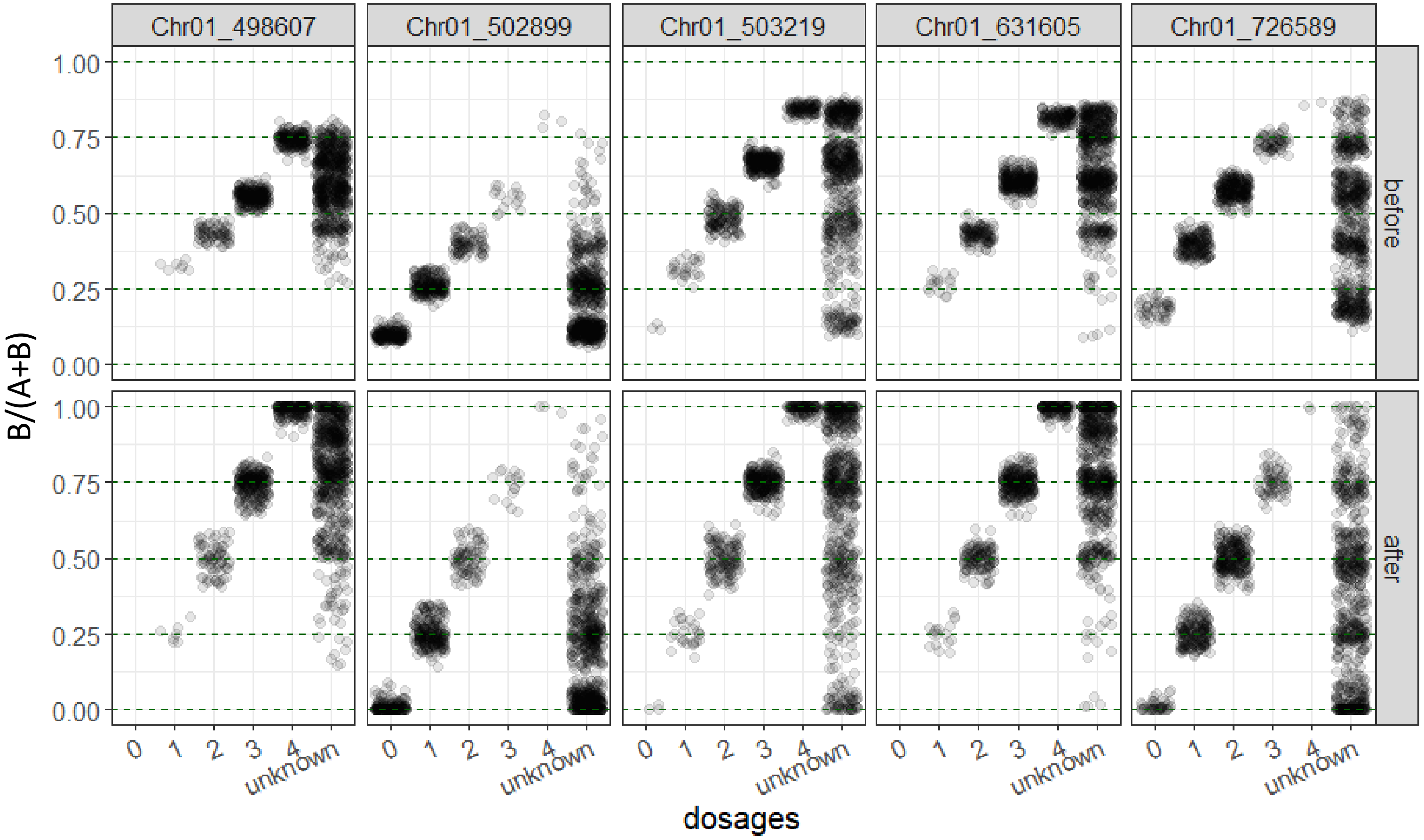
4 samples and all markers in chromosome



Human Genetics

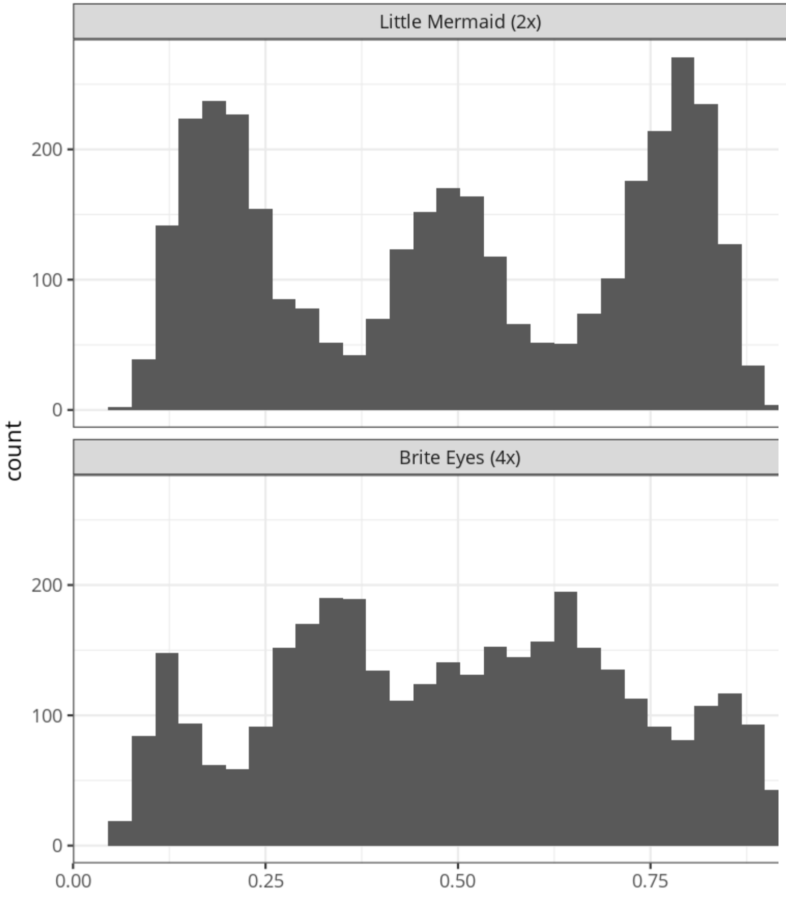
- CNV studies in humans: PennCNV, QuantiSNP, Birdsuite, …
- They always have diploid samples as reference
- Qploidy is the expansion of the PennCNV method (Wang et al., 2007)
- Using higher ploidies as reference

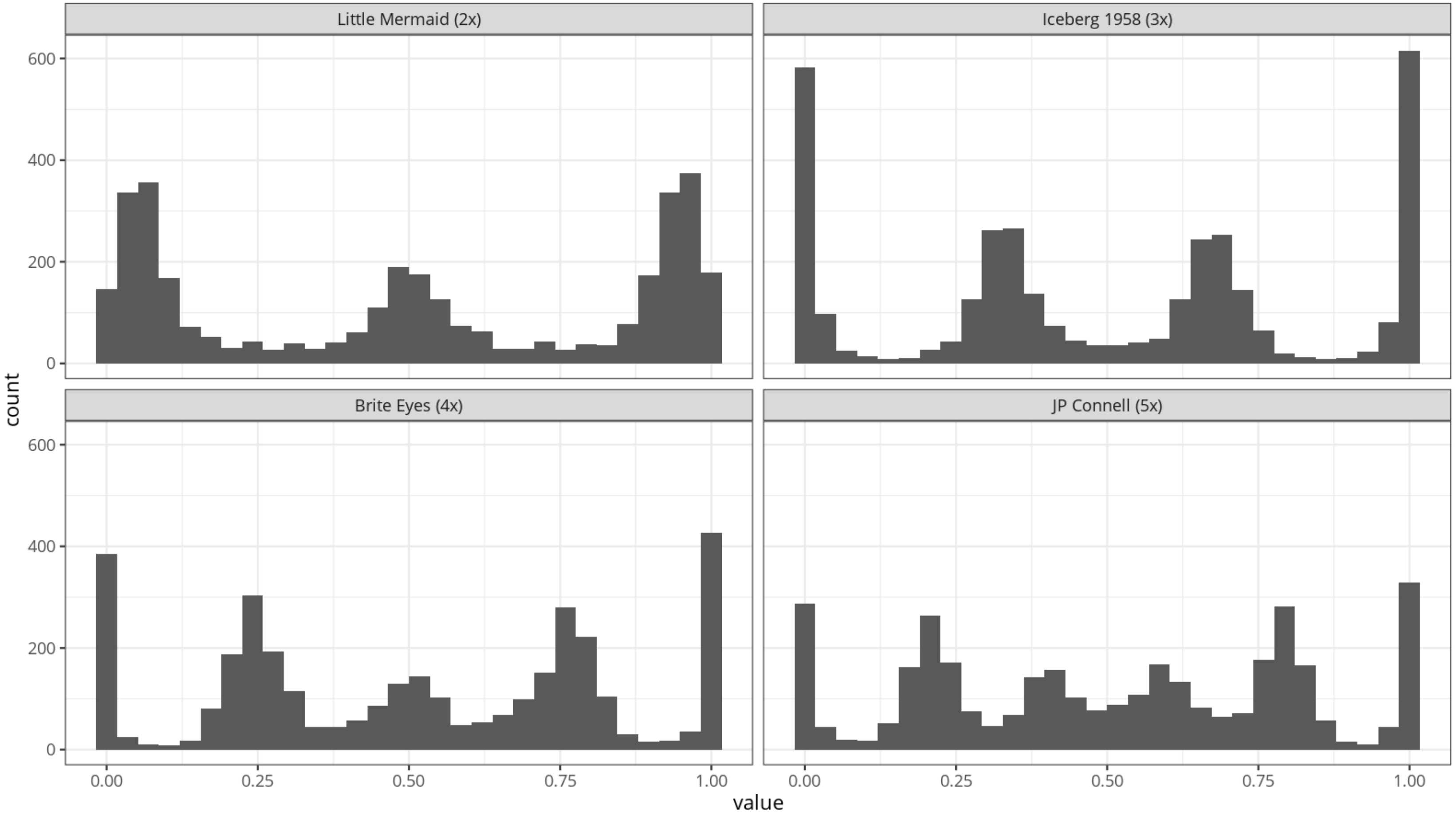




Sample-level ploidy determination

4 samples and all markers

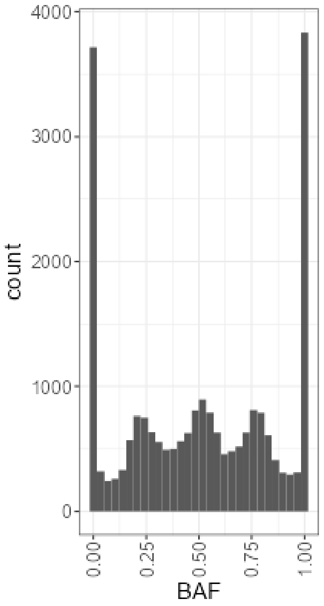




The improvement – chromosomal-level estimations

90_MBxBE (4x)

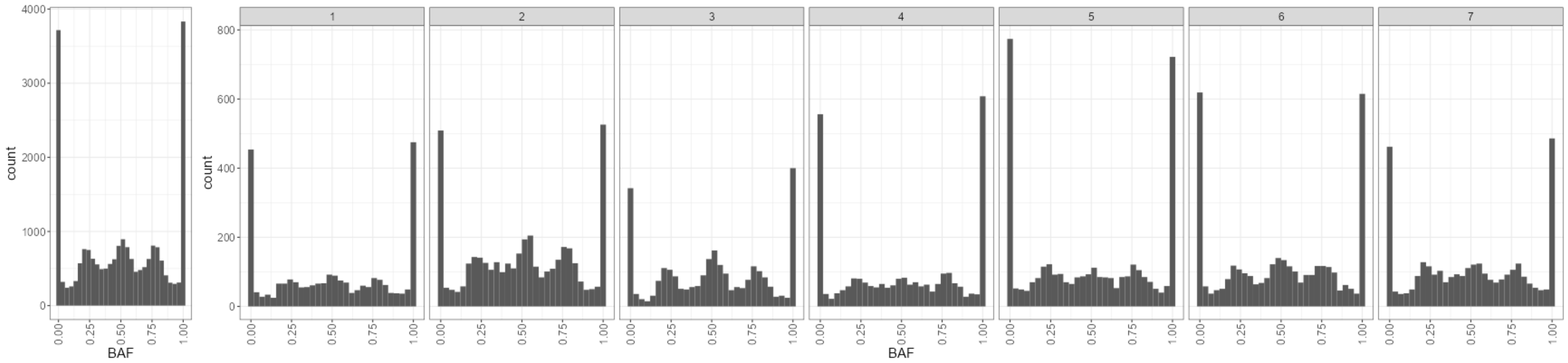
Diploids as reference



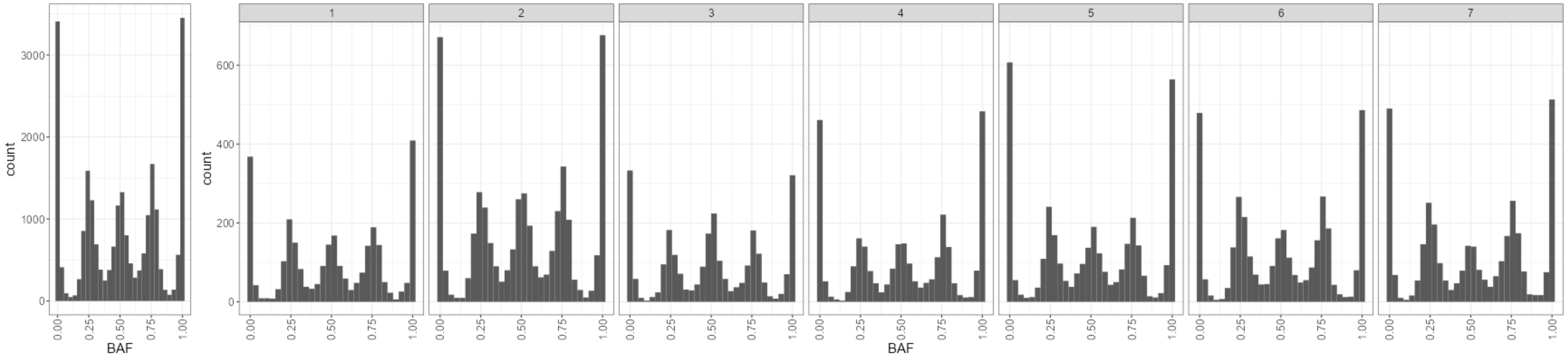
The improvement – chromosomal-level estimations

Diploids as reference

90_MBxBE (4x)



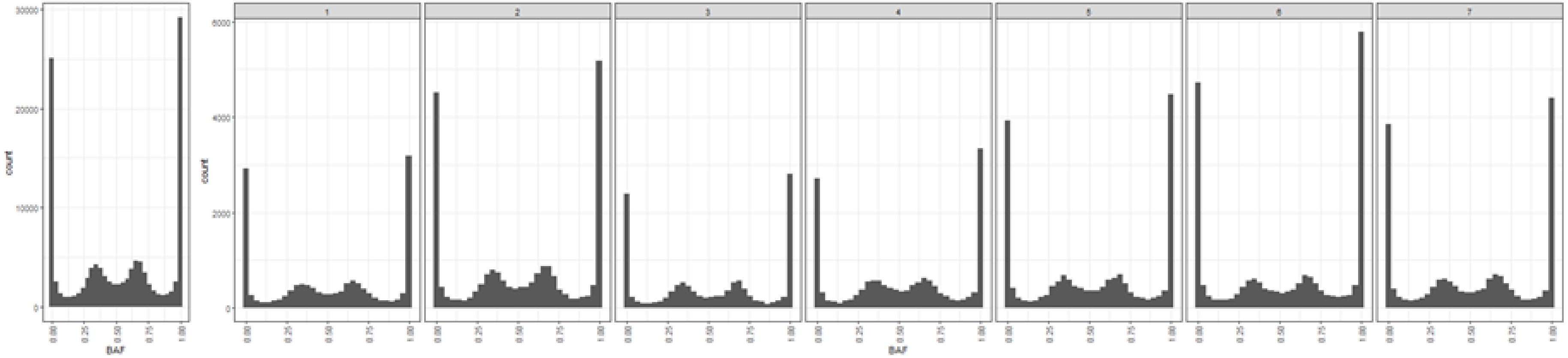
Tetraploids as reference



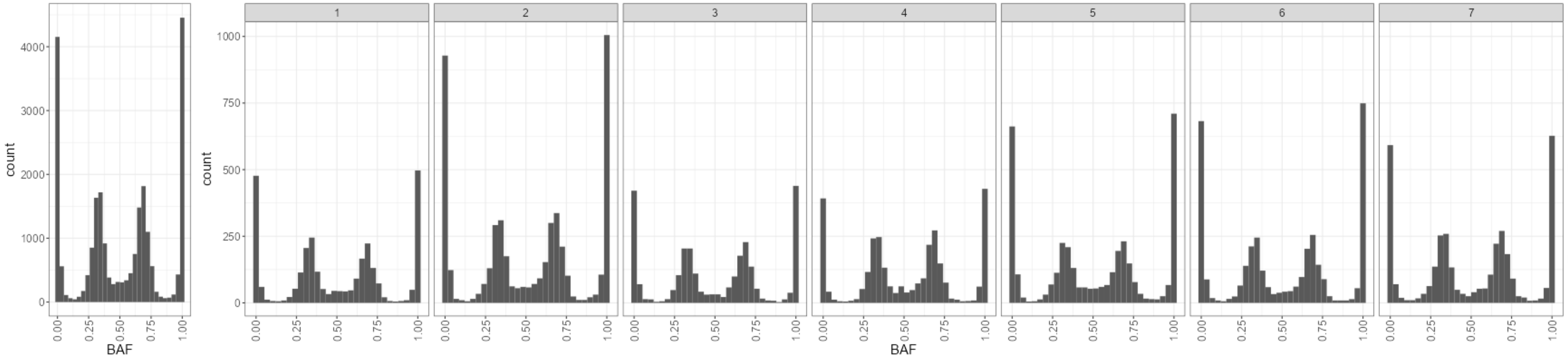
The improvement – chromosomal-level estimations

A triploid sample using diploids as reference

Sally Holmes (3x)



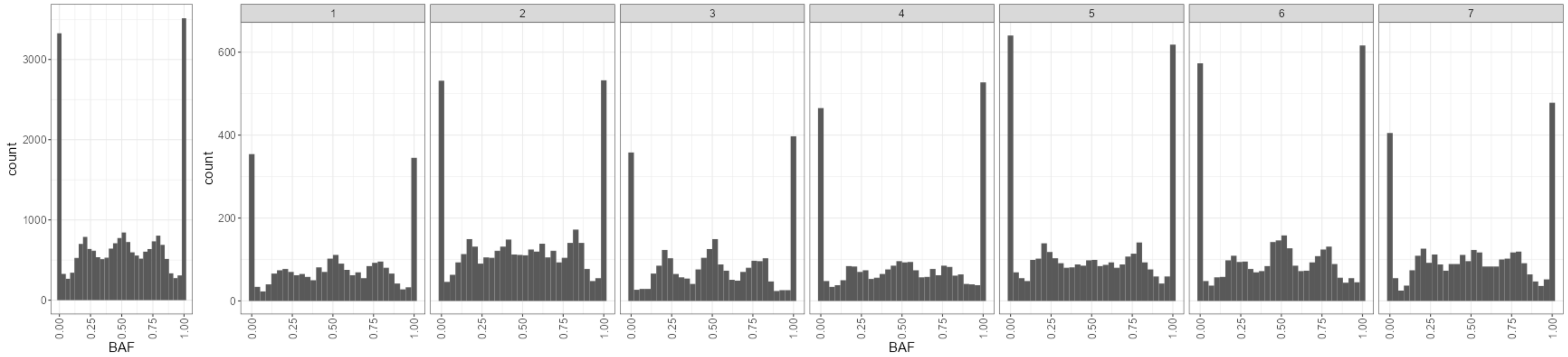
Tetraploids as reference



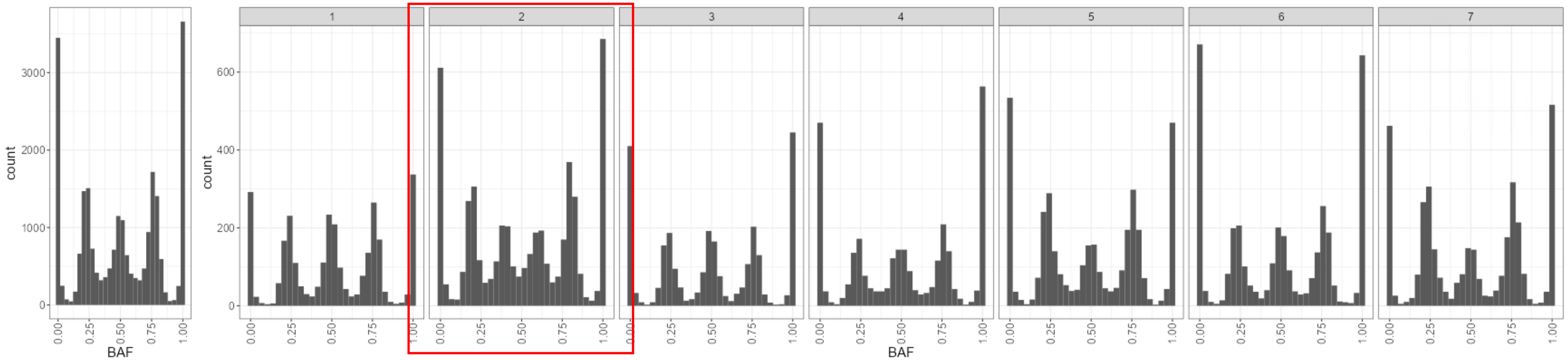
The improvement – chromosomal-level estimations

Diploids as reference

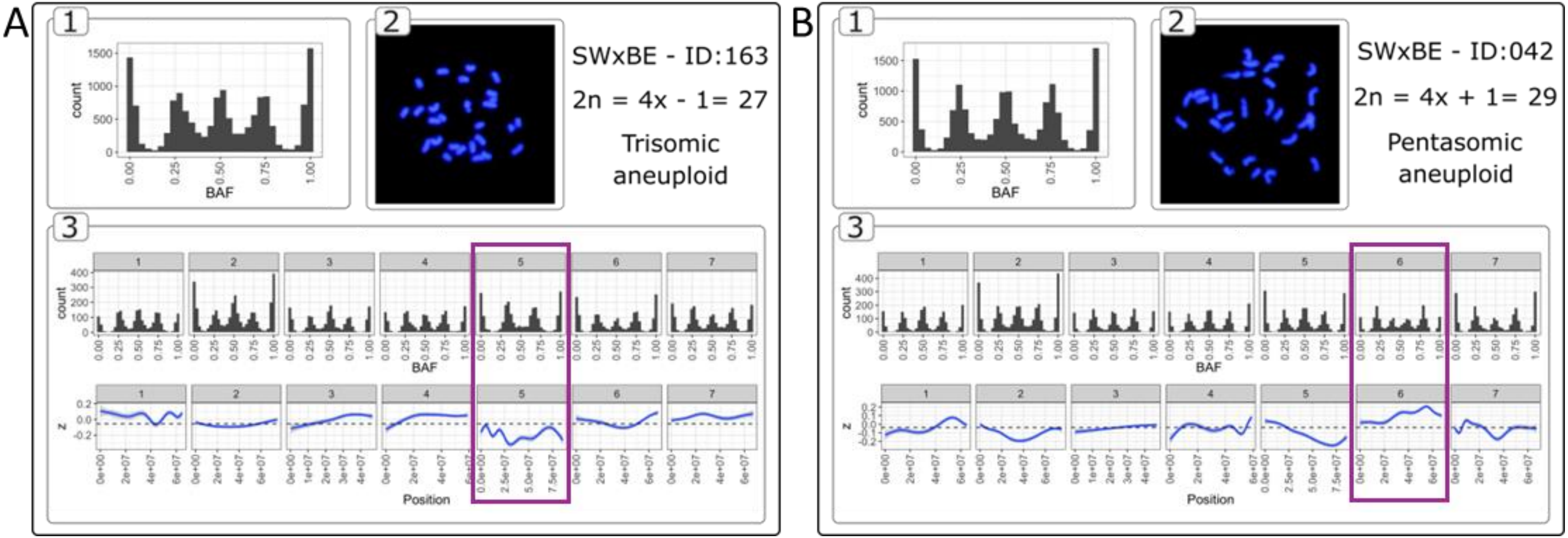
16401_N025 (4x + 1)



Tetraploids as reference



Z Score of the Total Allele Intensity/Depth

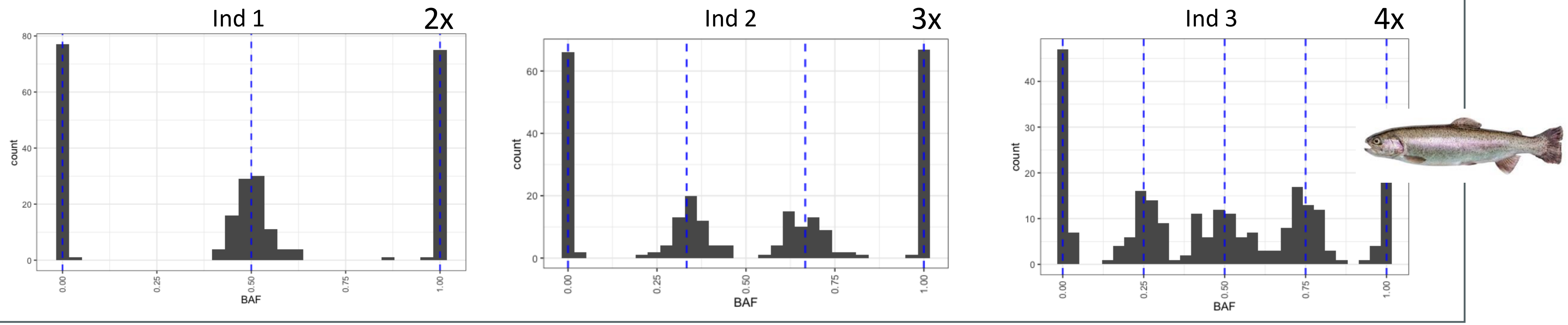


Other Species and Technologies

- - - Expected peak position

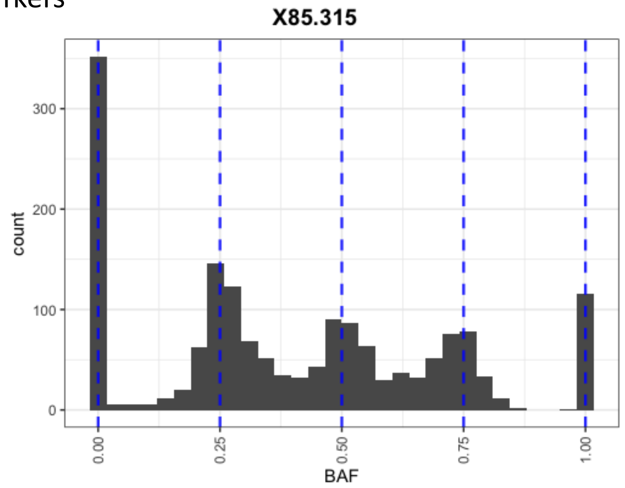
254 markers

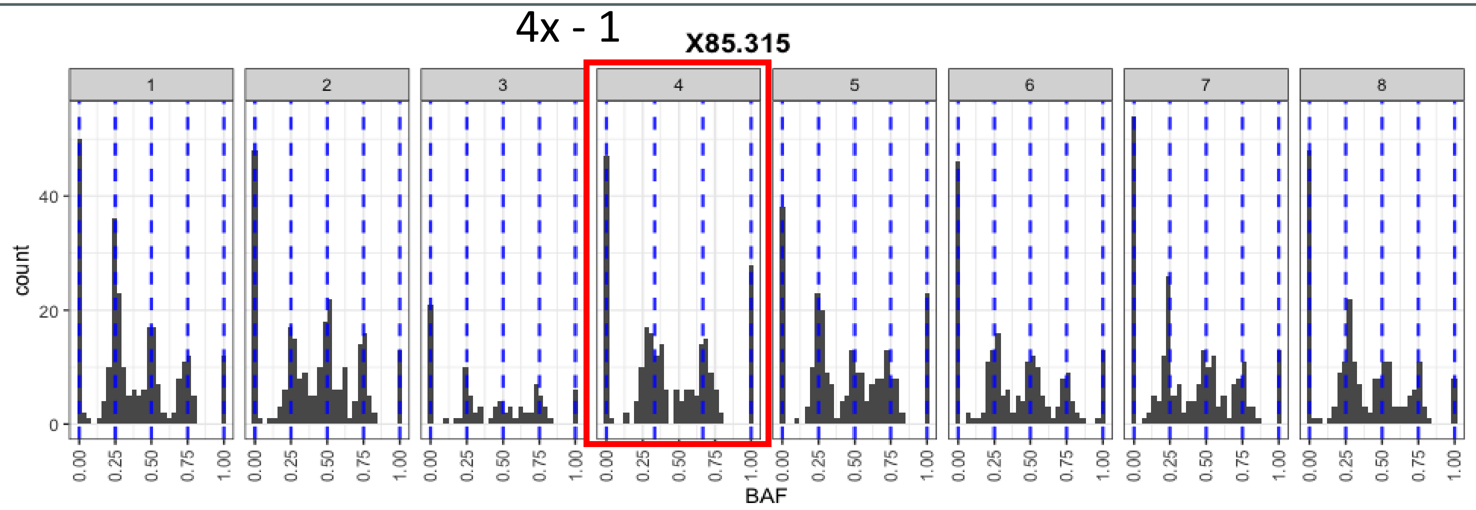
Trout – GTseq

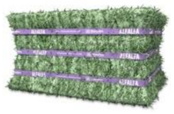


1677 markers

Alfalfa – DArTag







When to use

When does Qploidy work?

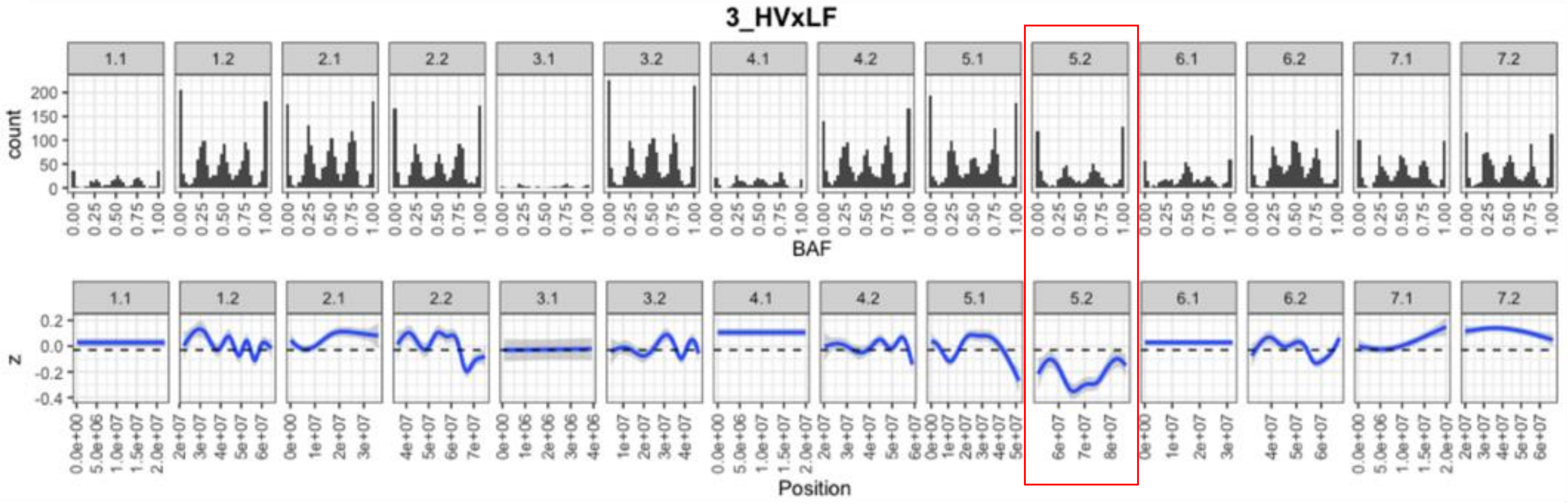
- Marker data derived from **Array** or **Target sequencing** platforms
- All DNA samples prepared following **same library preparation**
- **Known ploidy** of at least 60 samples or **known most common ploidy**
- Heterozygous
- Sample collected **represents whole individual**

When Qploidy does NOT work?

- RADseq or GBS libraries
- Combined datasets from genotyping in **different batches**
- **Do not have a subset of samples with known ploidy** or **lack a predominant ploidy**
- Samples consist of **inbred lines**
- Sample is possible from **chimeric or mosaic** tissue

What is next for Qploidy?

- Integrate Z score with BAF (HMM, deep learning?)
- Locus-level resolution



Linkage Mapping


Augusto Garcia
University of São Paulo


Marcelo Mollinari
NCState University


Gabriel Gesteira
NCState University

- Outcrossing populations
- Autopolyploids
- Interconnected full-sib populations
- Combination of ploidies
- Speed optimization
- Documentation
- Prepared to be implemented in an interface

Currently available at:

Cristianetaniguti/onemap and CRAN

OneMap

Since 2007

mmollina/MAPpoly and CRAN

MAPpoly

Since 2018

mmollina/mappoly2

MAPpoly2

Since 2024

Linkage Mapping


Augusto Garcia
University of São Paulo


Marcelo Mollinari
NCState University


Gabriel Gesteira
NCState University

- OneMap: no more new features since 3.2.0
  - Still the best option for:
    - Recombinant Inbred Lines
    - Dominant and multiallelic markers
- OneMap and MAPpoly are stable
- Development efforts are now in MAPpoly2

Currently available at :

Cristianetaniguti/onemap and CRAN

mmollina/MAPpoly and CRAN

mmollina/mappoly2

OneMap

Since 2007

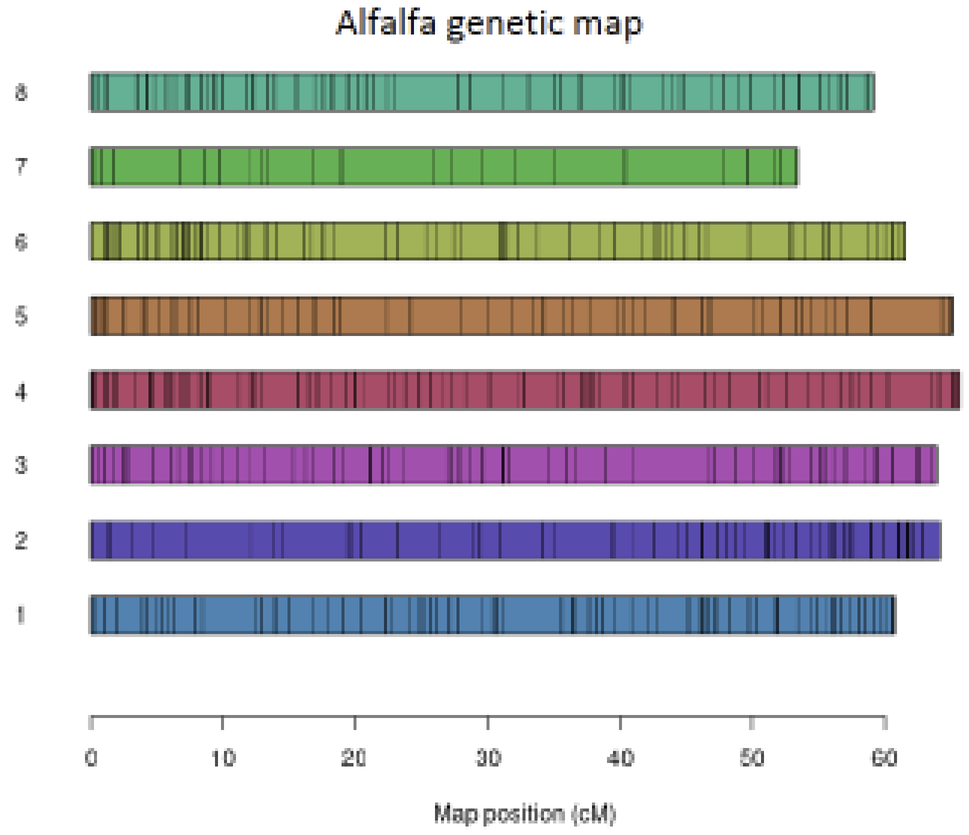
MAPpoly

Since 2018

MAPpoly2

Since 2024

Linkage Mapping - Alfalfa



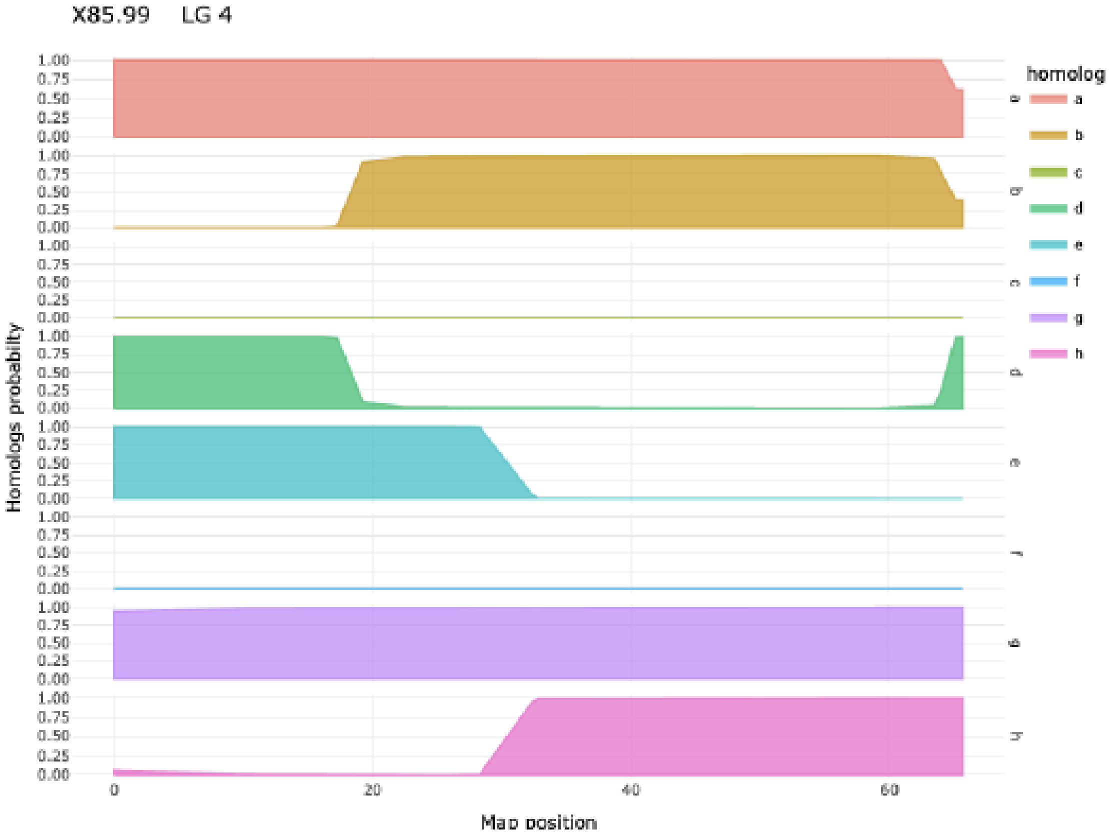
Summary

LG	Map length (cM)	Markers/cM	Simplex	Double-simplex	Multiplex	Total	Max gap
1	60.69	2.39	37	27	81	145	4.53
2	64.12	1.89	39	24	58	121	5.01
3	63.89	1.77	35	23	55	113	5.81
4	65.66	2.15	51	22	68	141	3.23
5	65.14	1.58	40	17	46	103	5.32
6	61.58	2.24	47	32	59	138	3.86
7	53.39	0.75	15	7	18	40	7.49
8	59.18	2.23	46	19	67	132	4.86
Total	493.65	1.88	310	171	452	933	5.01

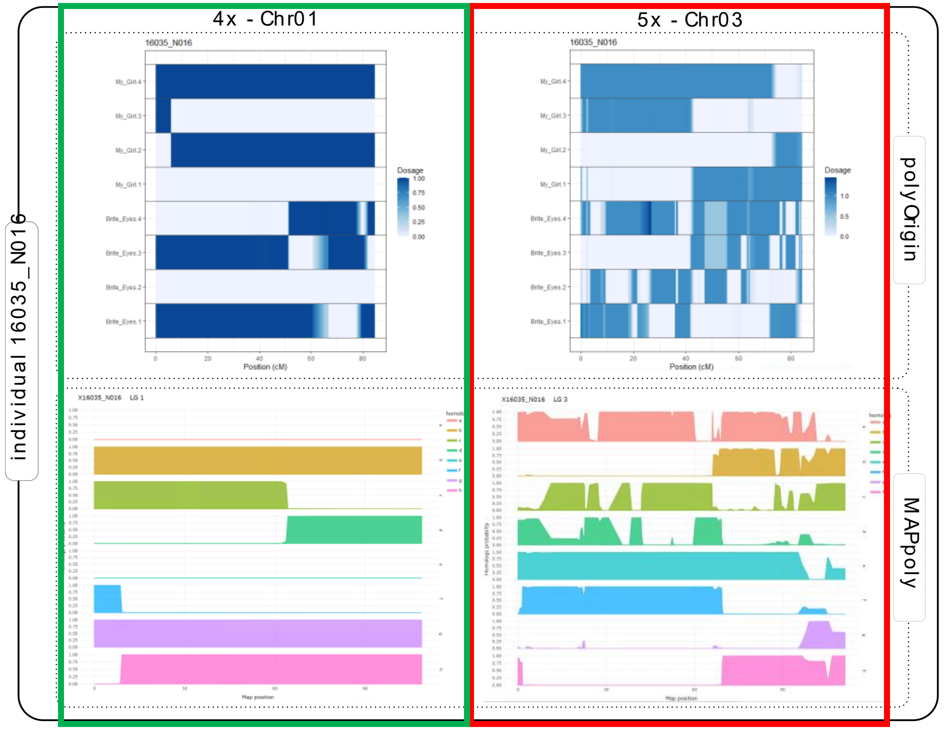
Based on bi-parental F1 population of parents susceptible (I195) and resistant (J432) to *Aphanomyces euteiches*

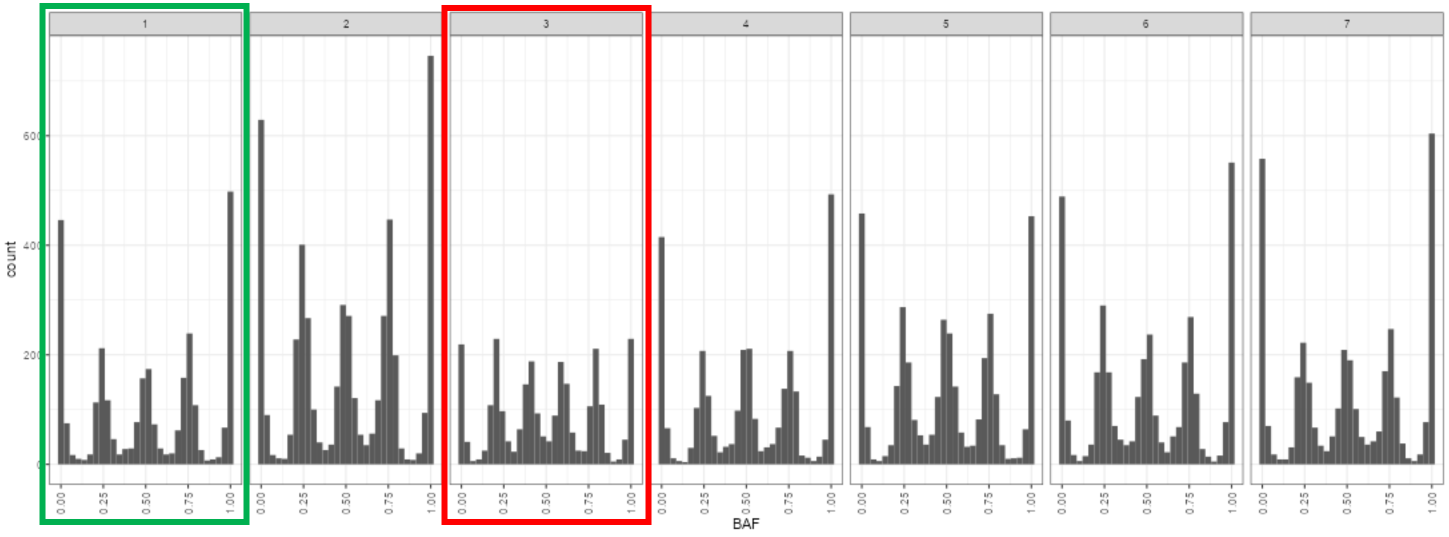
Haplotype Reconstruction

Example of haplotype reconstruction in individual 85_99, LG4



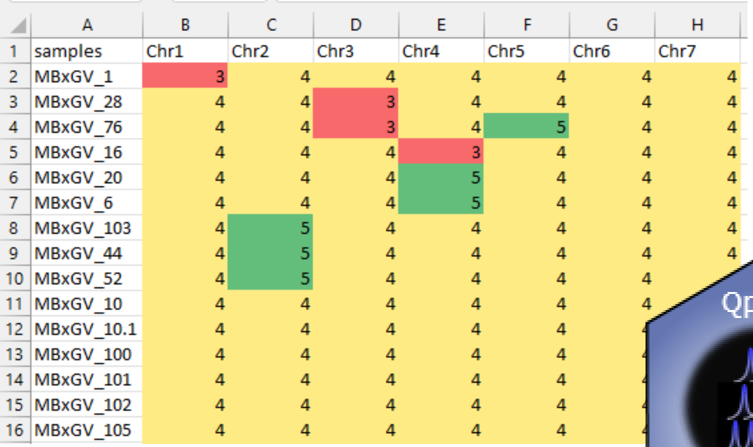
Aneuploids and Linkage Maps





Impact on Linkage Maps

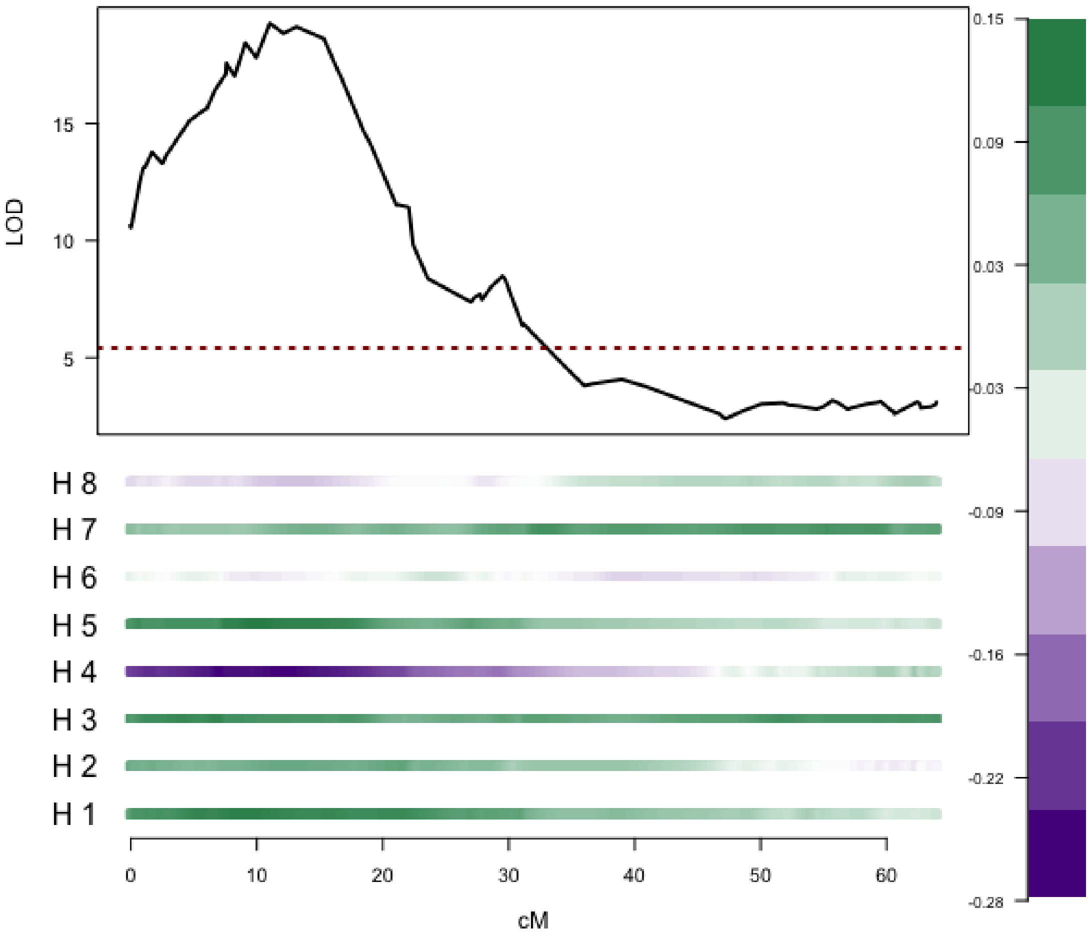
```
filter_aneuploid(
    mappoly.obj,
    Qploidy.out.csv)
```



	A	B	C	D	E	F	G	H
1	samples	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7
2	MBxGV_1	3	4	4	4	4	4	4
3	MBxGV_28	4	4	3	4	4	4	4
4	MBxGV_76	4	4	3	4	5	4	4
5	MBxGV_16	4	4	4	3	4	4	4
6	MBxGV_20	4	4	4	5	4	4	4
7	MBxGV_6	4	4	4	5	4	4	4
8	MBxGV_103	4	5	4	4	4	4	4
9	MBxGV_44	4	5	4	4	4	4	4
10	MBxGV_52	4	5	4	4	4	4	
11	MBxGV_10	4	4	4	4	4	4	
12	MBxGV_10.1	4	4	4	4	4		
13	MBxGV_100	4	4	4	4	4		
14	MBxGV_101	4	4	4	4	4		
15	MBxGV_102	4	4	4	4	4		
16	MBxGV_105	4	4	4	4	4		

Population ID	# of samples	Female parent	Male parent	Aneuploid individuals	Aneuploid chromosomes
MBxBE	95	Morden Blush	Brite Eyes	18 (18.94 %)	2.3 %
HVxLF	91	High Voltage	Lemon Fizz	13 (14.28 %)	1.9 %
MBxRR	94	Morden Blush	Ramblin Red	14 (14.89 %)	2.0 %
MBxGV	94	Morden Blush	George Vancouver	9 (9.57 %)	1.3 %
BExMG	121	Brite Eyes	My Girl	20 (16.53 %)	2.6 %
SWxBE	208	Stormy Weather	Brite Eyes	51 (24.52 %)	3.3 %
BExMG	66	My Girl	Brite Eyes	7 (10.6 %)	1.5 %

Alfalfa - Aphanomyces Resistance QTL Analysis



LG	Position (cM)	LOD	PVE (BLUE*)	PVE (SMA**)
3	11.04	19.28	38.61	16.13

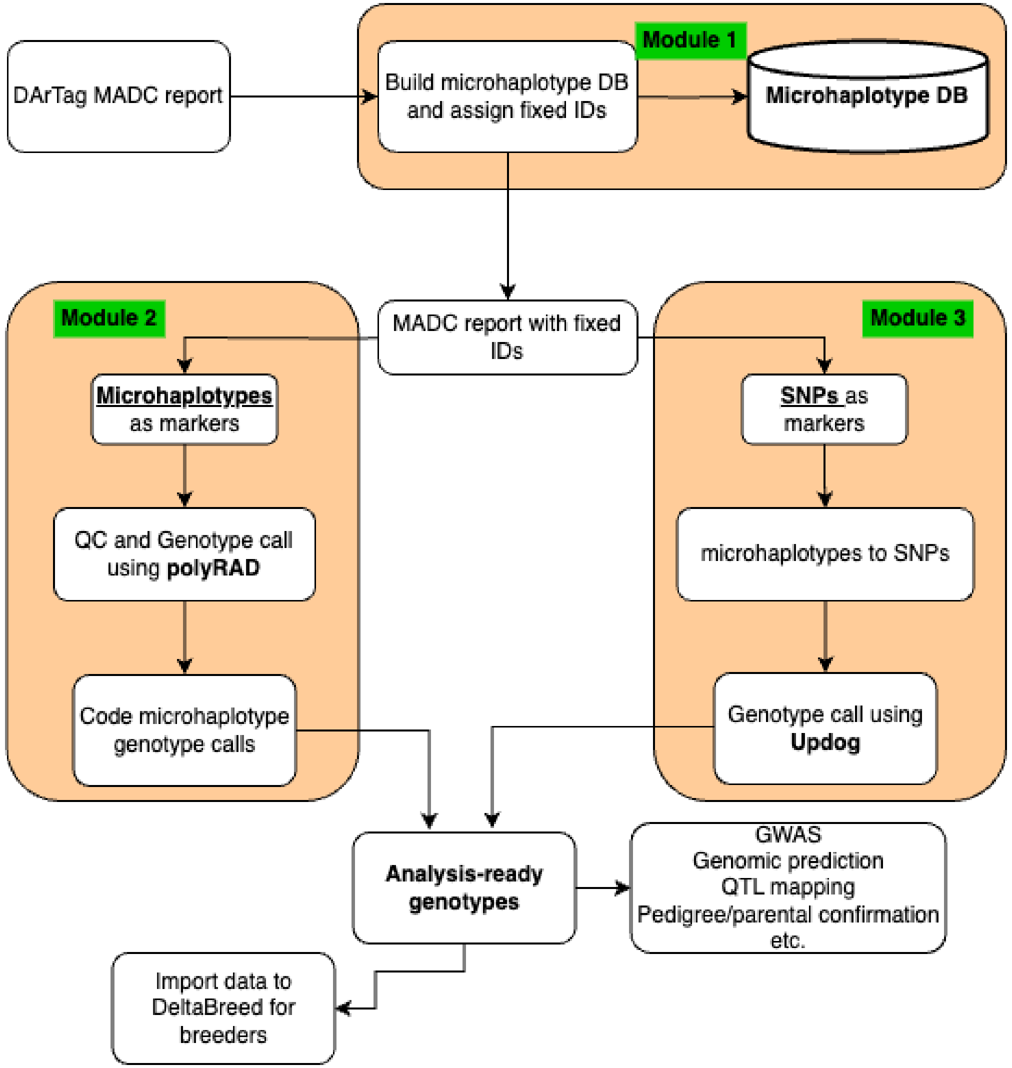
PVE - Percentage of variation explained
*BLUE: Best linear unbiased estimates
** Single marker analysis

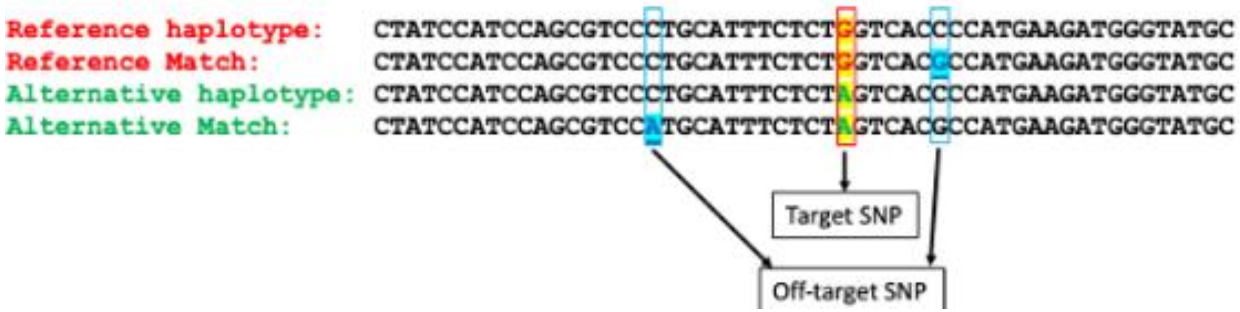
- The alleles originating from homologues 1, 2, 3 of parent 1 and homologues 5 and 7 of parent 2 contributes positively to the trait.
- The most appropriate model is oooQ x oQQQ.

Limited to Mapping Populations Designs

Development of Microhaplotype Database

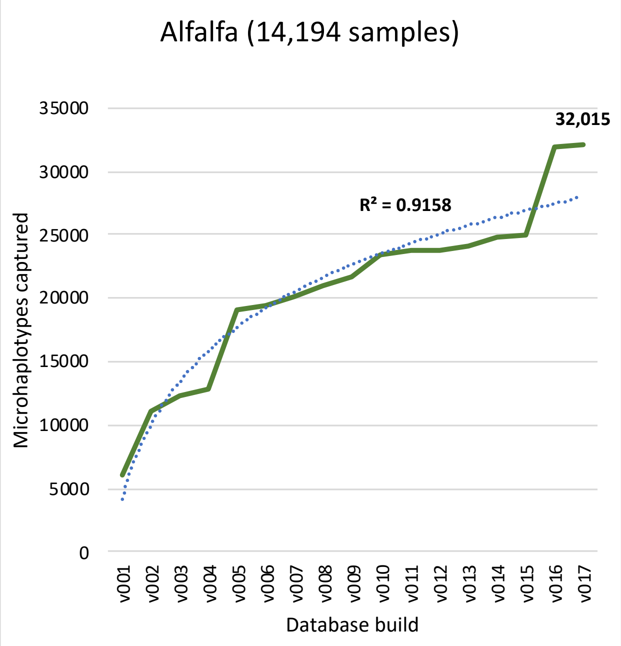
DONGYAN ZHAO
Genomics Lead

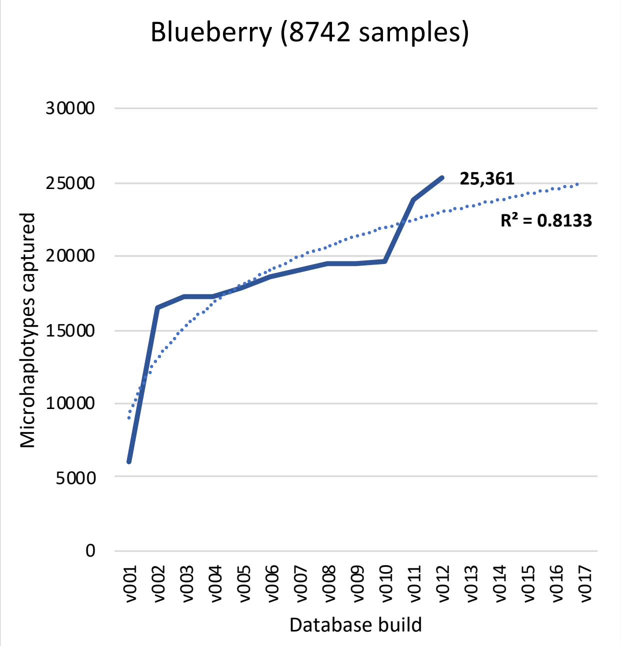


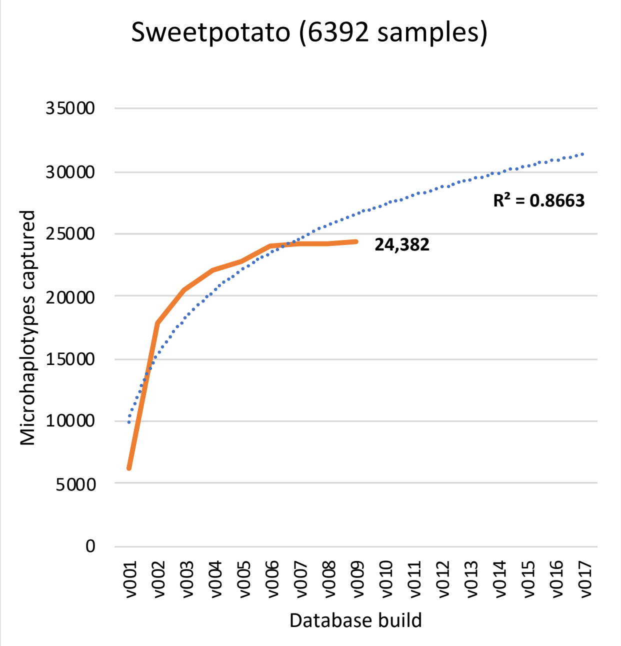


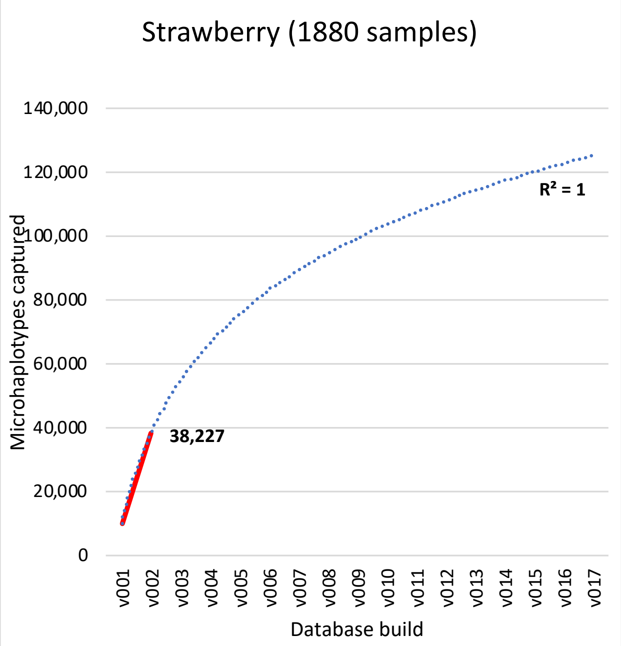
- Establish best workflow for filtering and processing genotype data
- Create file formats that work with other applications for downstream analysis

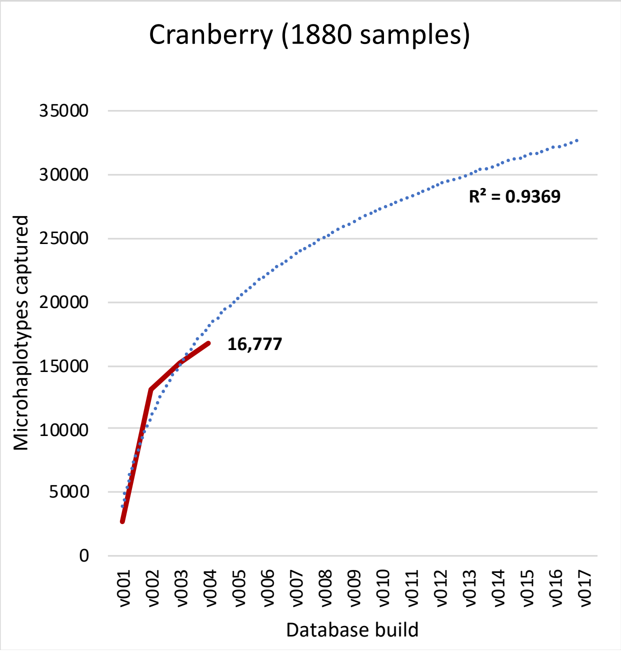
Microhaplotype Database allele capture curves

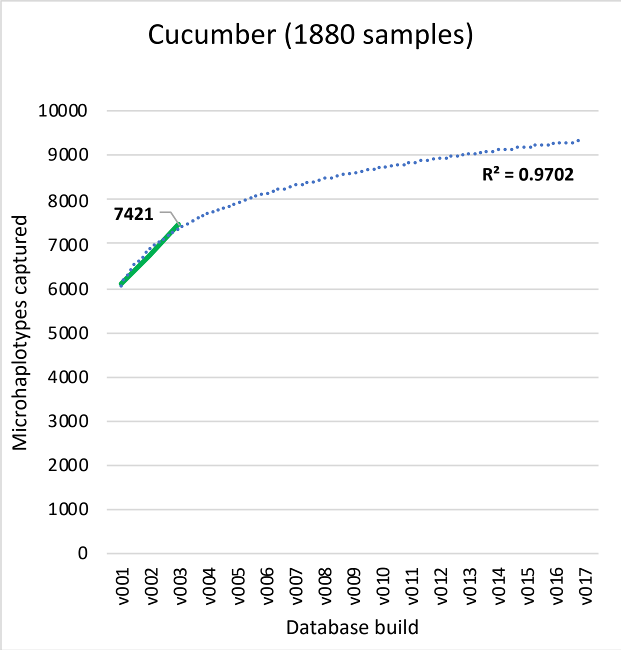

Alfalfa (14,194 samples)

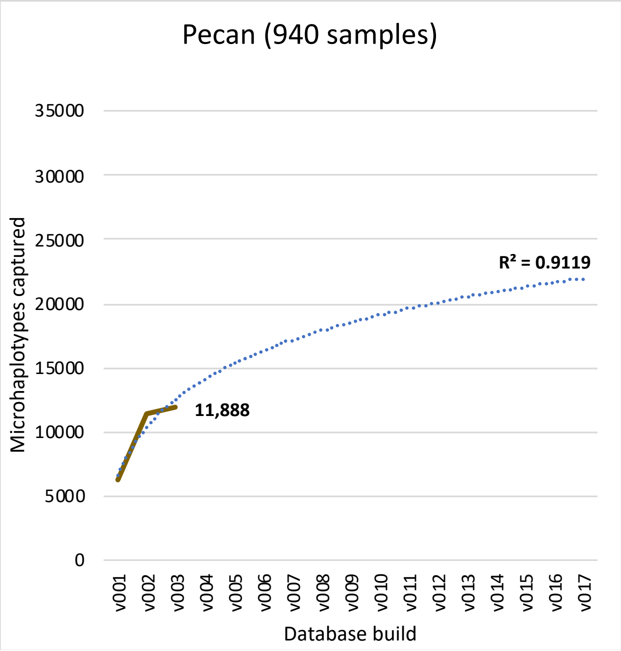

Blueberry (8742 samples)


Sweetpotato (6392 samples)


Strawberry (1880 samples)


Cranberry (1880 samples)

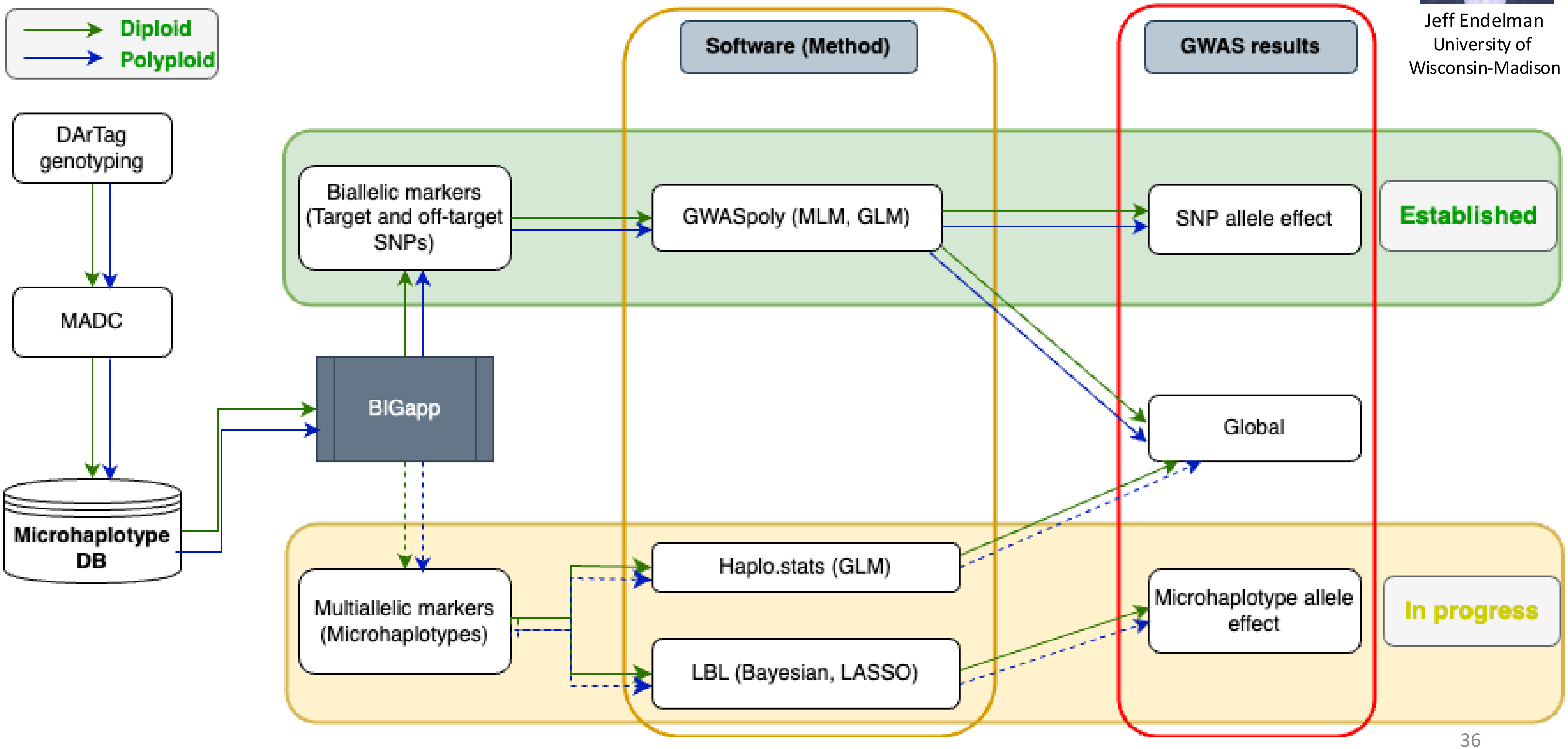

Cucumber (1880 samples)


Pecan (940 samples)

- 156,076 alleles across 7 species
- Allele capture curves are expected to follow a diminishing returns pattern. The slope is very steep in early builds but begins to asymptote in later builds as fewer new alleles are discovered.
- Databases are considered mature when they asymptote.

Multiallelic Genotypes for GWAS

Jeff Endelman
University of
Wisconsin-Madison



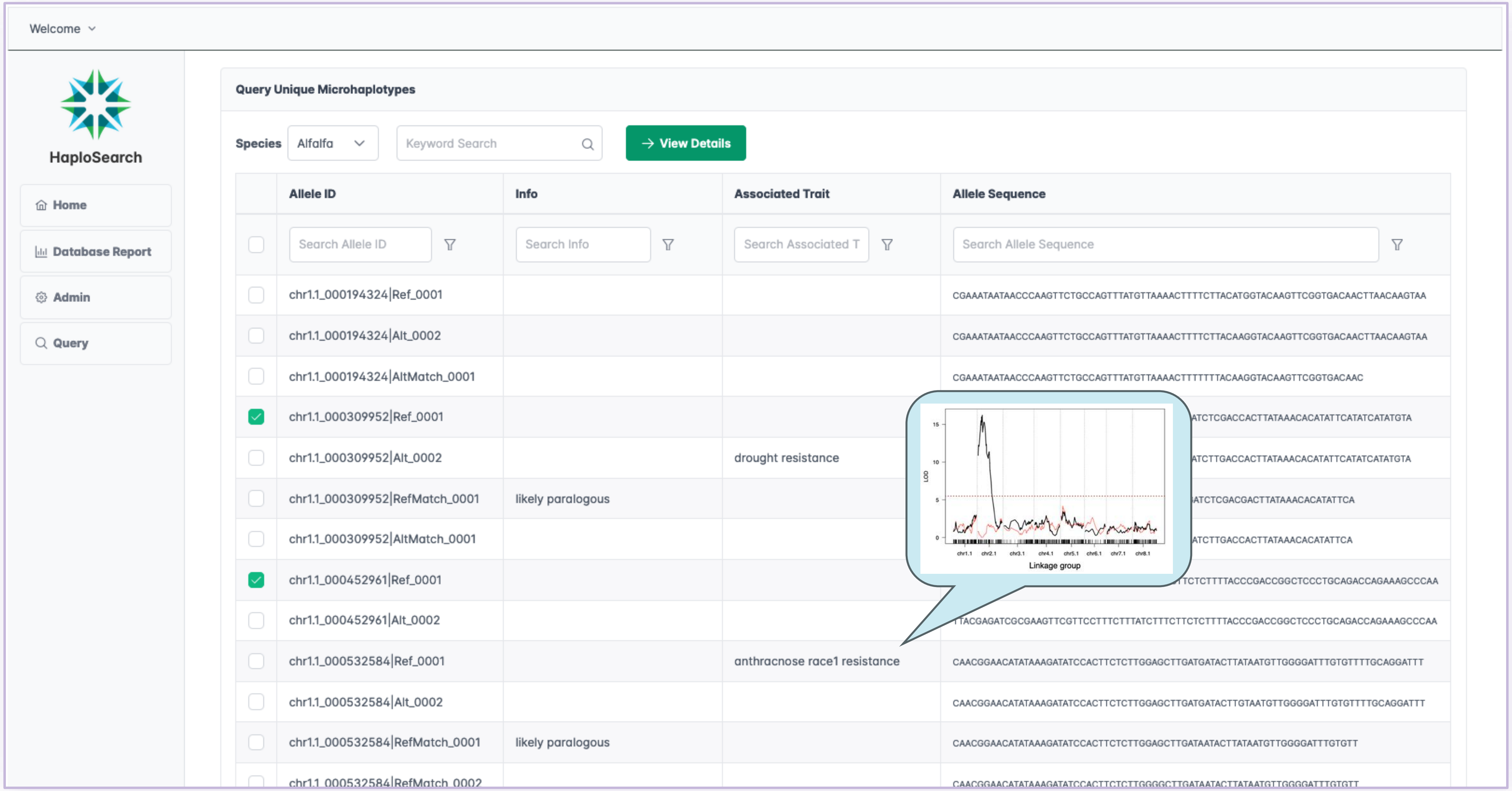
HaploSearch: user interface to search for microhaplotypes

TYLER SLONECKI

Biosoftware Coordinator

MENG LIN

Bioinformatics
Coordinator



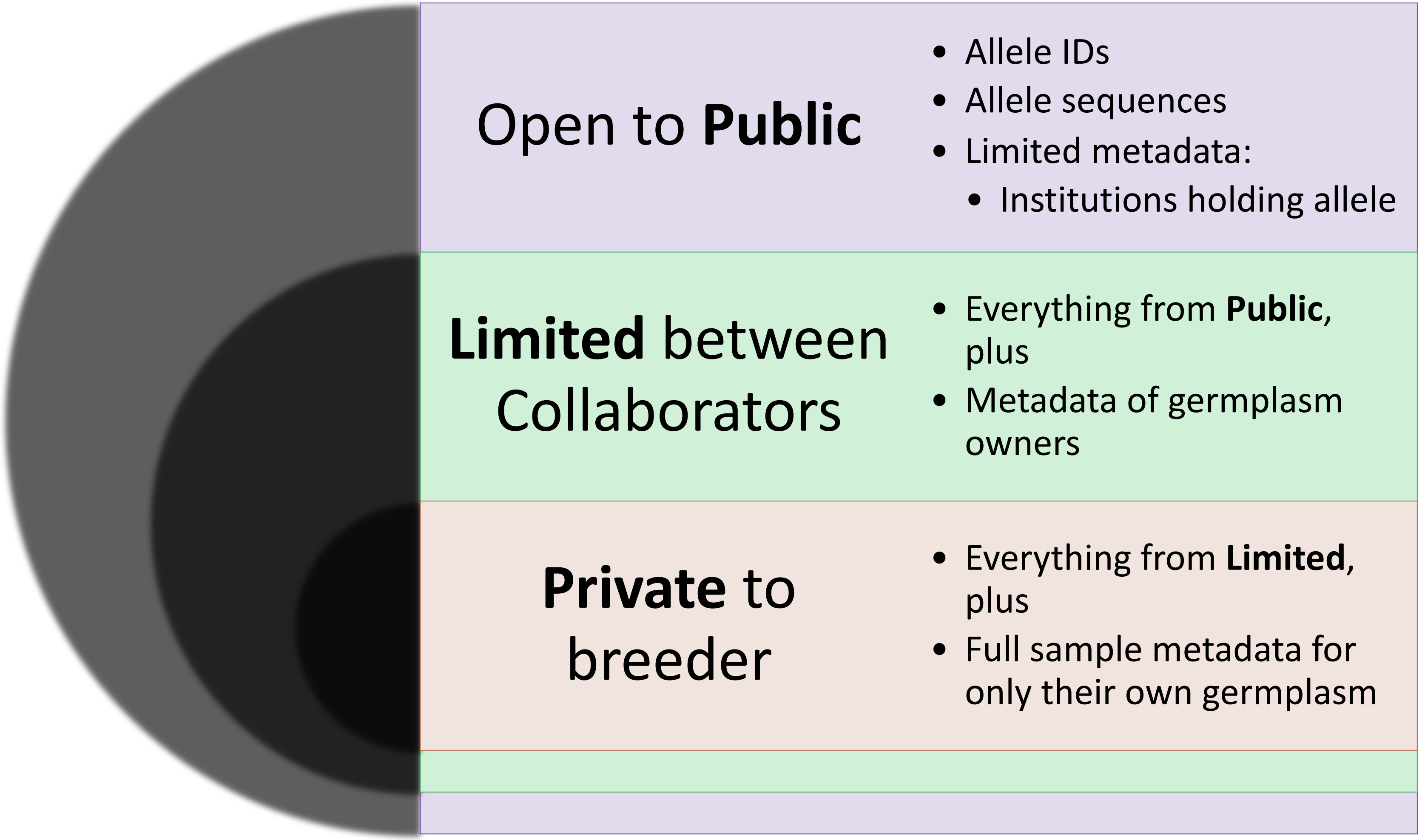
Query Unique Microhaplotypes

Species: Alfalfa

Keyword Search

	Allele ID	Info	Associated Trait	Allele Sequence
☐	Search Allele ID	Search Info	Search Associated T	Search Allele Sequence
☐	chr1.1_000194324|Ref_0001			CGAAATAATAACCCAAGTTCTGCCAGTTTATGTTAAAACTTTTCTTACATGGTACAAGTTCGGTGACAACTTAACAAGTAA
☐	chr1.1_000194324|Alt_0002			CGAAATAATAACCCAAGTTCTGCCAGTTTATGTTAAAACTTTTCTTACAAGGTACAAGTTCGGTGACAACTTAACAAGTAA
☐	chr1.1_000194324|AltMatch_0001			CGAAATAATAACCCAAGTTCTGCCAGTTTATGTTAAAACTTTTTTTACAAGGTACAAGTTCGGTGACAAC
☑	chr1.1_000309952|Ref_0001			ATCTCGACCACTTATAAACACATATTCATATCATATGTA
☐	chr1.1_000309952|Alt_0002		drought resistance	ATCTTGACCACTTATAAACACATATTCATATCATATGTA
☐	chr1.1_000309952|RefMatch_0001	likely paralogous		ATCTCGACGACTTATAAACACATATTCA
☐	chr1.1_000309952|AltMatch_0001			ATCTTGACCACTTATAAACACATATTCA
☑	chr1.1_000452961|Ref_0001			TTCTCTTTTACCCGACCGGCTCCCTGCAGACCAGAAAGCCCAA
☐	chr1.1_000452961|Alt_0002			TTACGAGATCGCGAAGTTCGTTCCTTTCTTTATCTTTCTTCTCTTTTACCCGACCGGCTCCCTGCAGACCAGAAAGCCCAA
☐	chr1.1_000532584|Ref_0001		anthracnose race1 resistance	CAACGGAACATATAAAGATATCCACTTCTCTTGGAGCTTGATGATACTTATAATGTTGGGGATTTGTGTTTTGCAGGATTT
☐	chr1.1_000532584|Alt_0002			CAACGGAACATATAAAGATATCCACTTCTCTTGGAGCTTGATGATACTTGTAATGTTGGGGATTTGTGTTTTGCAGGATTT
☐	chr1.1_000532584|RefMatch_0001	likely paralogous		CAACGGAACATATAAAGATATCCACTTCTCTTGGAGCTTGATAATACTTATAATGTTGGGGATTTGTGTT
☐	chr1.1_000532584|RefMatch_0002			CAACGGAACATATAAAGATATCCACTTCTCTTGGGGCTTGATAATACTTATAATGTTGGGGATTTGTGTT

Data sharing model



Open to **Public**

- Allele IDs
- Allele sequences
- Limited metadata:
  - Institutions holding allele

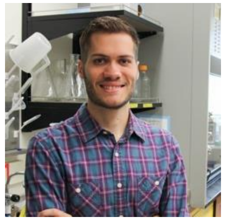
Limited between Collaborators

- Everything from **Public**, plus
- Metadata of germplasm owners

Private to breeder

- Everything from **Limited**, plus
- Full sample metadata for only their own germplasm

Introducing BIGapp

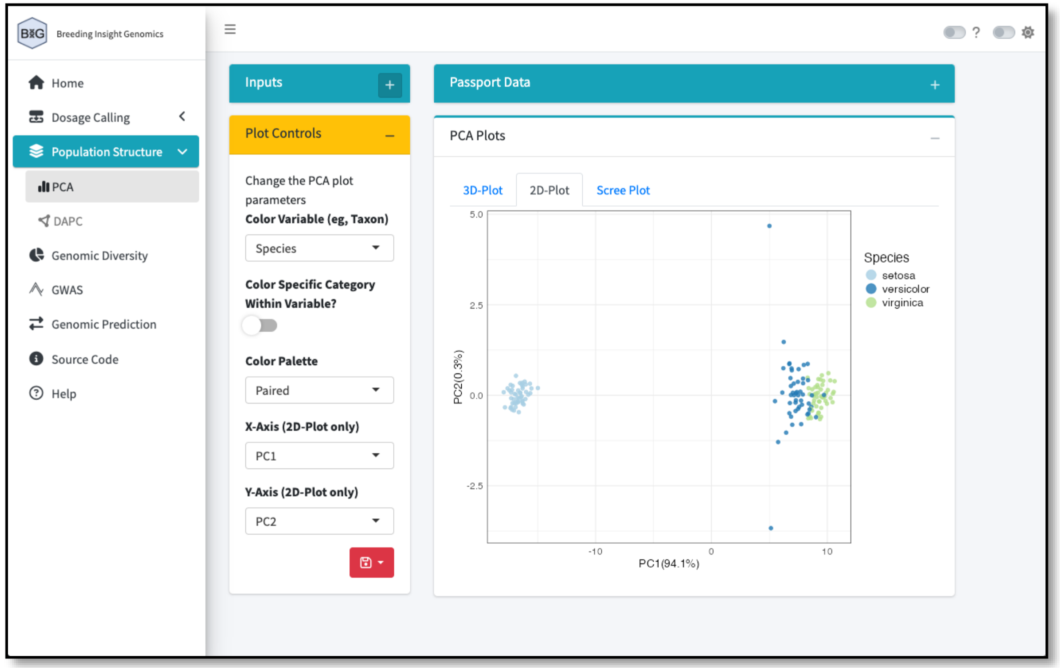


ALEX SANDERCOCK
Genomics Postdoc

BIGapp is a web-based application to make BI genomics and bioinformatics analyses

BIGapp's Key Design Principles:

1. Accessibility:
    - Intuitive interface; no coding experience needed.
2. Integrated Help:
    - Tutorials and help pages for guidance and result interpretation.
3. Broad Support:
    - Designed for all species and ploidy levels.



Change the PCA plot parameters

Color Variable (eg, Taxon): Species

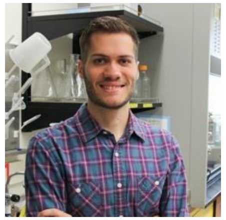
Color Specific Category Within Variable?

Color Palette: Paired

X-Axis (2D-Plot only): PC1

Y-Axis (2D-Plot only): PC2

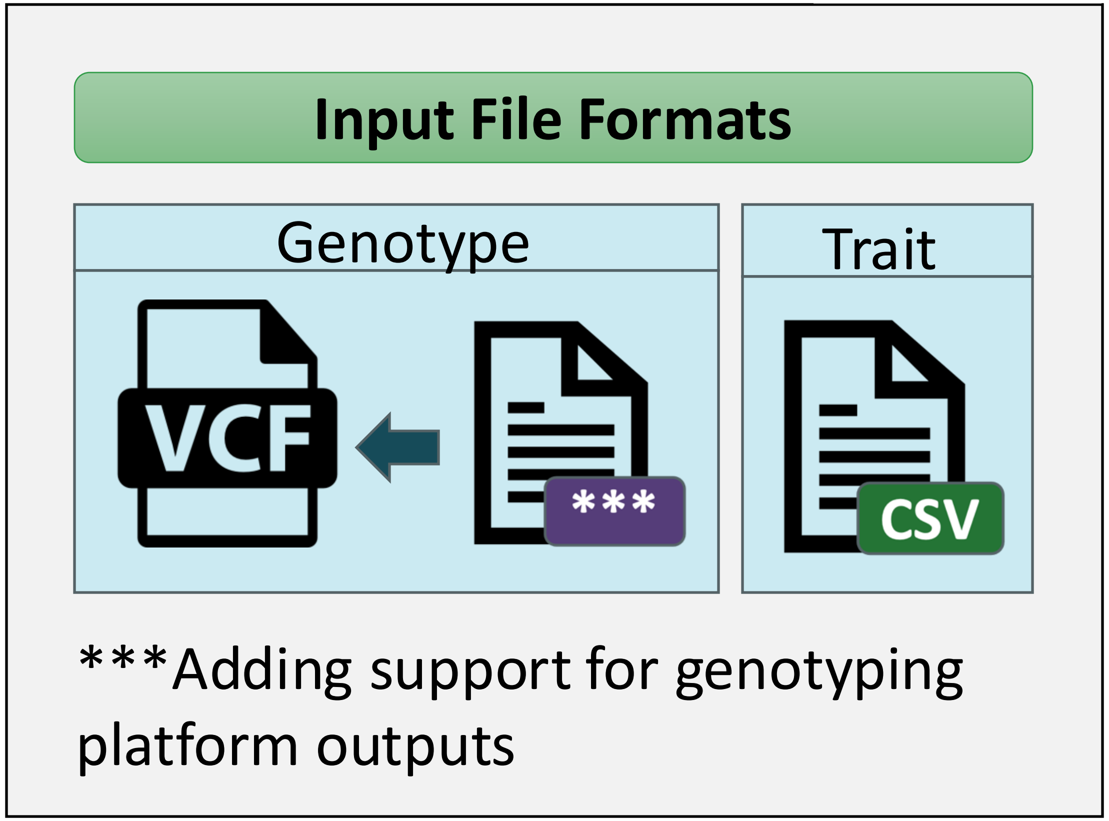
https://github.com/Breeding-Insight/BIGapp
https://github.com/Breeding-Insight/BIGr



ALEX SANDERCOCK
Genomics Postdoc

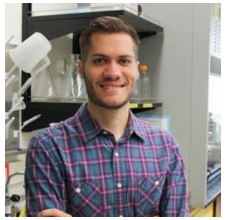
BIGapp v1.0 Features

- Genotype processing
  - Dosage call from read counts
  - SNP filtering
  - Sample filtering
- Summary metrics
  - SNP minor allele frequency
  - Sample observed heterozygosity
  - Population dosage ratios
- Population Structure
  - PCA
  - DAPC
- Genome-wide association studies (GWAS)
- Genomic Selection (GS)



***Adding support for genotyping platform outputs

Availability and Future Plans



ALEX SANDERCOCK
Genomics Postdoc

How to get BIGapp?

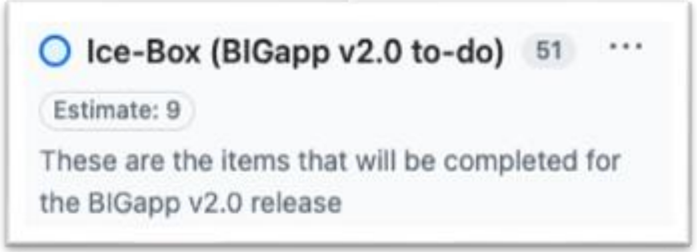
- Visit https://github.com/Breeding-Insight and install it like an R package.

Future Plans:

1. Integrate ploidy estimation support (Qploidy)
2. Integrate linkage mapping support (MAPpoly2).
3. Integrate support for GWAS analysis with microhaplotypes

(B)reeding (I)nsight (G)enomics app (BIGapp)

BIGapp is a user-friendly web application built with R and Shiny, designed to simplify the processing of low to mid-density genotyping data for both diploid and polyploid species. It provides a powerful and intuitive interface for researchers and breeders to analyze genomic data without requiring command-line expertise.



○ Ice-Box (BIGapp v2.0 to-do) 51 ···

Estimate: 9

These are the items that will be completed for the BIGapp v2.0 release

Highlights

- New Tools available:
    - Qploidy
    - MAPpoly2
    - BIGapp
- Keep and eye out for updates and releases:
    - Qploidy
    - MAPpoly2
    - BIGapp
    - MicroHaplotypeDB
    - HaploSearch
    - GWASpoly
- Shiny Power: Implementing Genomic Tools in user-friendly interfaces

BI is supported by

BI OnRamp Staff:
Amanda Hulse-Kemp
Keo Corak
Heather Manching
Brian Scheffler

Breeding Insight Staff:
AJ Ackerman
Bhoja Basnet
Craig Beil
Sam Bouabane
Alex Casa
Shufen Chen
Josue Chinchilla
Rebecca Cubitt
Meng Lin
Kristen Lind
Matthew Mandyck
David Meidlinger
Shahana Mustafa
Nick Palladino
Kristina Petrilose
Kirsten Richardson
Edwin Reidel
Alex Sandercock
Manoj Sapkota
Tyler Slonecki
Dexter Stigger
Madhav Subedi
Heather Sweeney
Xuemei Tang
Cris Taniguti
Tyr Weisner-Hanks
Meseret Wondifraw
Shawn Yarnes
Dongyan Zhao

Scientific Advisory Board:
Pete Cyr
Gina Brown-Guedira
Scott Jackson (Vice Chair)
Steve Kresovich
Steve Larson
Tim Parsons
Trevor Rife
Steve Rounsley
Pete Selby
Philipp Simon (Chair)
Allen Van Deynze
Margaret Worthington
Ed Buckler
Jean-Luc Jannink

ARS ONP:
Jose Costa
Kevin Hackett
Joe Munyaneza
Jack Okamuro
Peter Vadas
Caird Rexroad III
Tim Rinehart (POC)

GitHub: https://github.com/Breeding-Insight

Former BI members:
Katherine Mejia-Guerra, Liz Woods,
Tim Parsons, Chris Tucker,
Vanessa Greenlee, & Chris Strock

Funding and Support

U.S. Department of Agriculture – Agriculture Research Service
through Cornell University