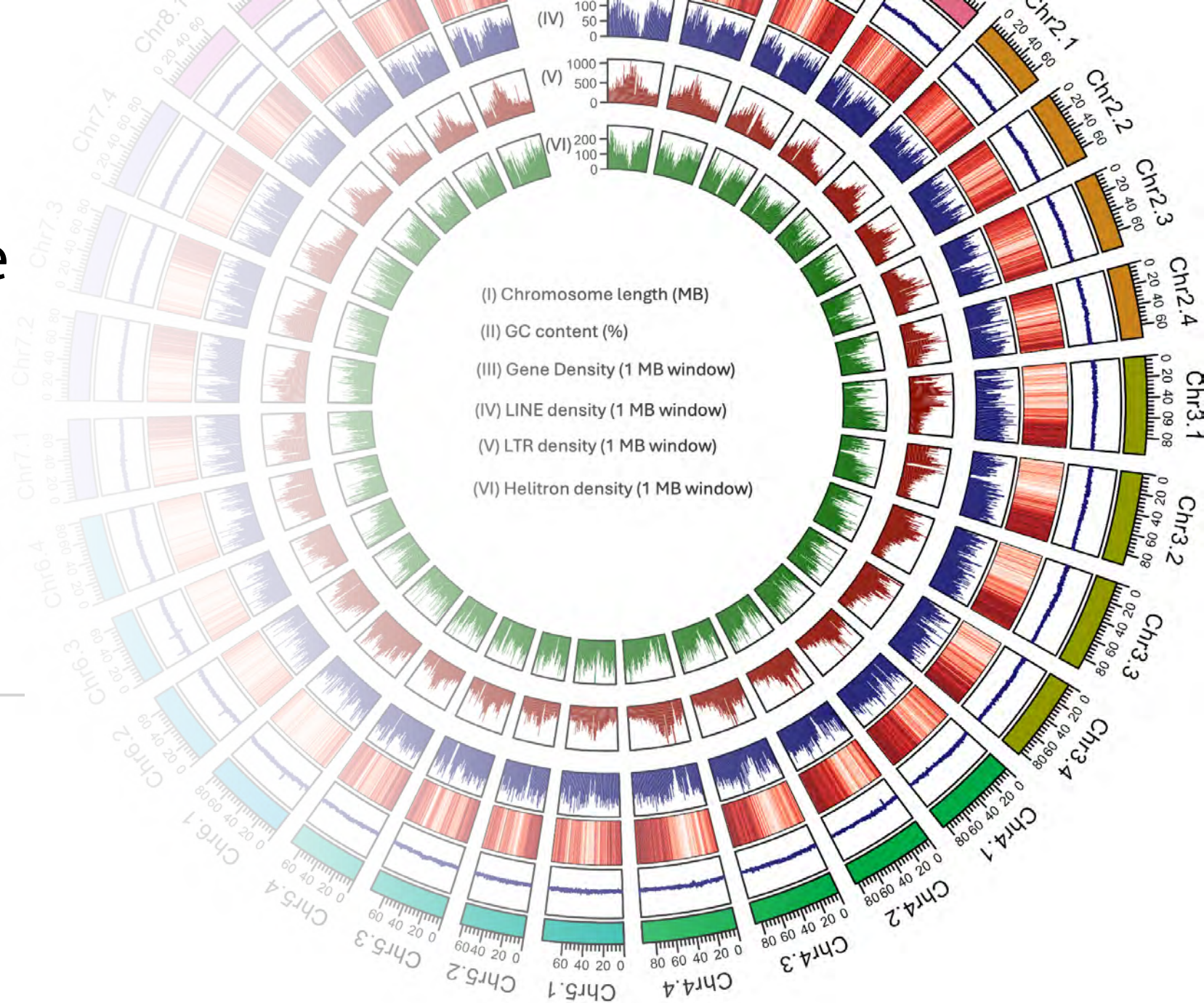


Four Haplotypes, One Genome: Uncovering hidden structural diversity in alfalfa RegenSY27x genotype

Presenter: Harpreet Kaur

Postdoctoral Associate

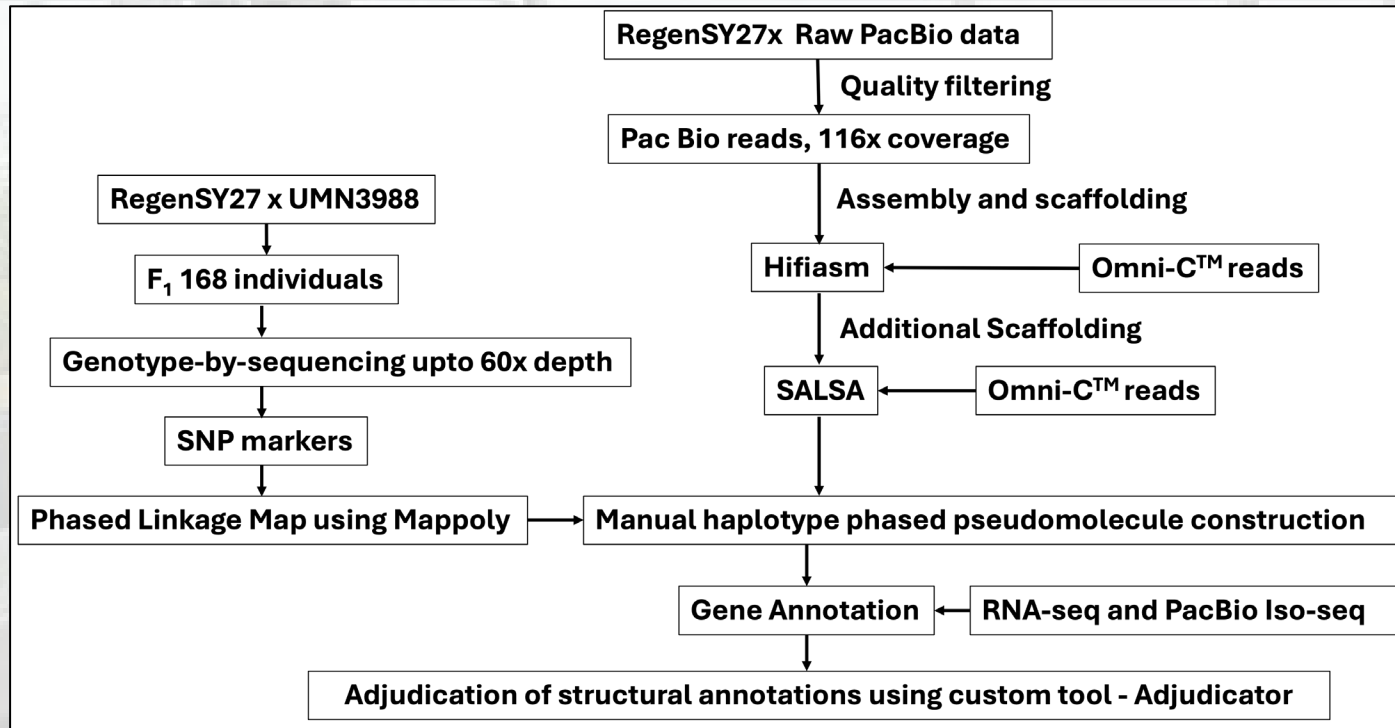
University of Minnesota and USDA-ARS, Saint Paul, MN



Introduction

- Alfalfa is a perennial, autotetraploid ($2n=4x=32$) forage legume species.
- Recently, three autotetraploid alfalfa genome assemblies have been published –
 - Zhongmu No. 1 (consensus genome assembly) Shen et al. 2020
 - Xin Jiang Da Ye (phased haplotype-aware assembly) Chen et al. 2020
 - Bolivia (phased haplotype-aware assembly) Zhang et al. 2025
 - Zhongmu No. 4 (phased haplotype-aware assembly, improved version 2) Sod et al. 2026
- RegenSY27x genotype widely used for its high regenerative capacity under tissue culture

Assembly Pipeline

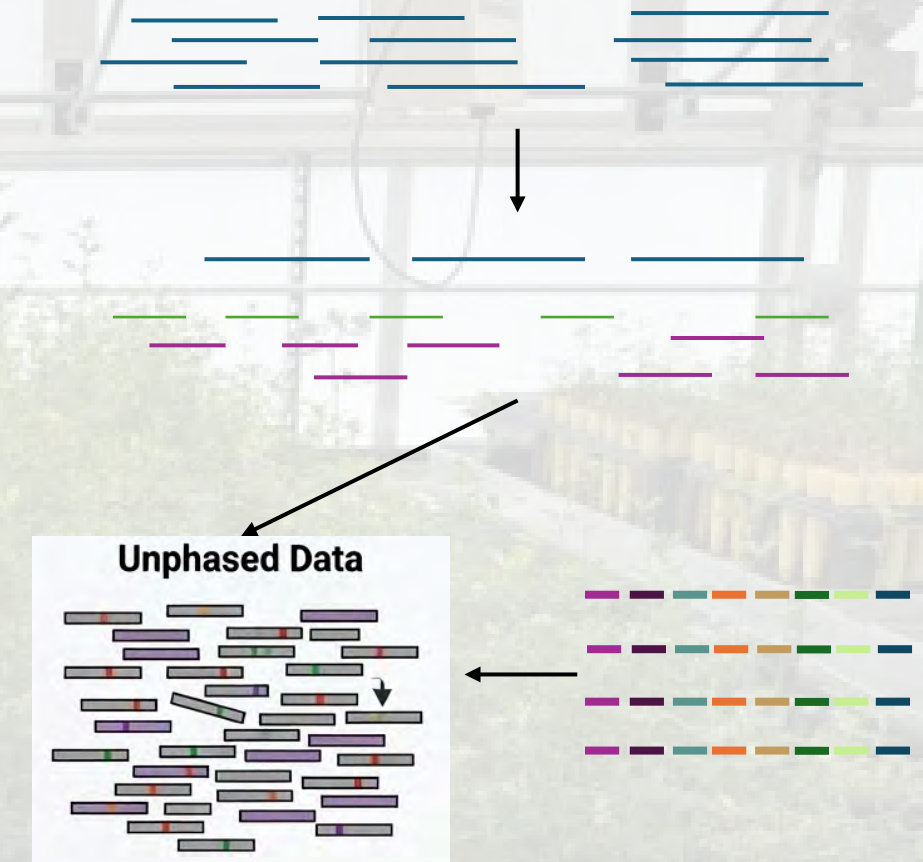
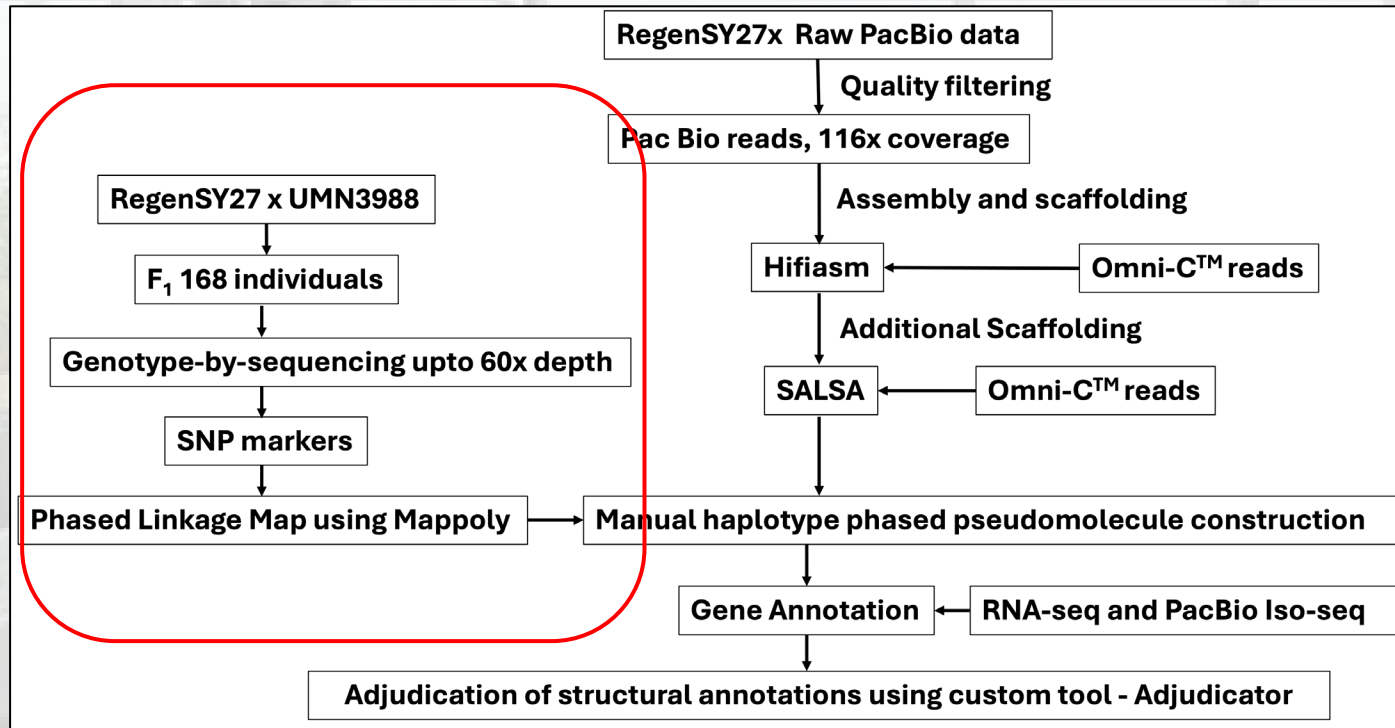


Autotetraploid Genome



8 Chromosomes per genome set
32 total chromosomes
Genome size = 800 MB
Total size = 800 x 4 = 3200 MB

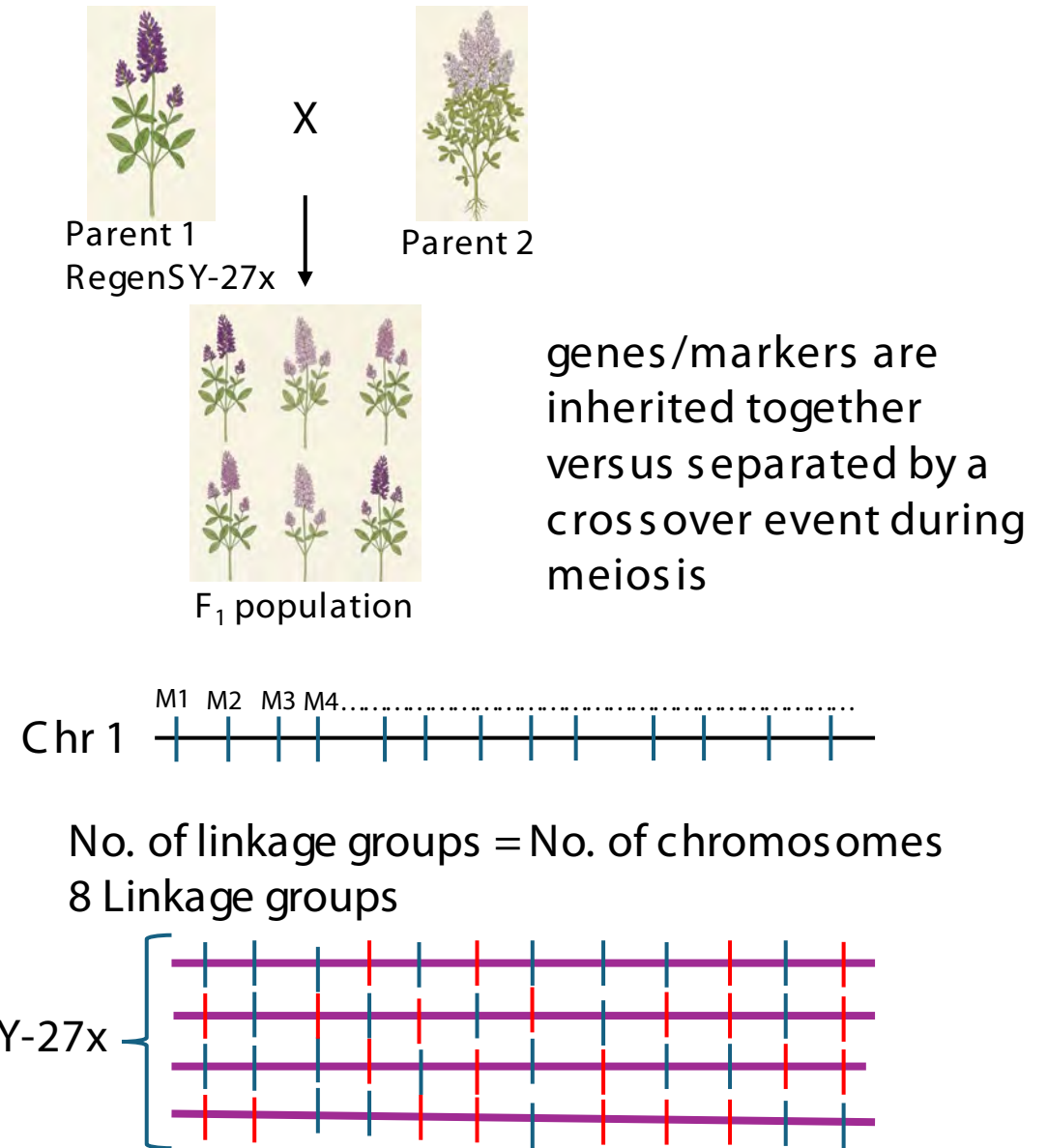
Assembly Pipeline



8 Chromosomes per genome set
32 total chromosomes

Linkage Maps were used for Haplotype Phasing

- Autotetraploid dosage aware Mappoly software (Mollinari & Garcia, 2019) was used to develop a phased RegenSY-27x genotype phased linkage map
- High read depth Genotype-by-sequencing (upto 60x) data to confidently differentiate three heterozygous classes using biallelic SNPs
 - ABBB
 - AABB
 - AAAB
- Align the linkage based phased haplotype data to our assembled sequences to separate the four haplotypes



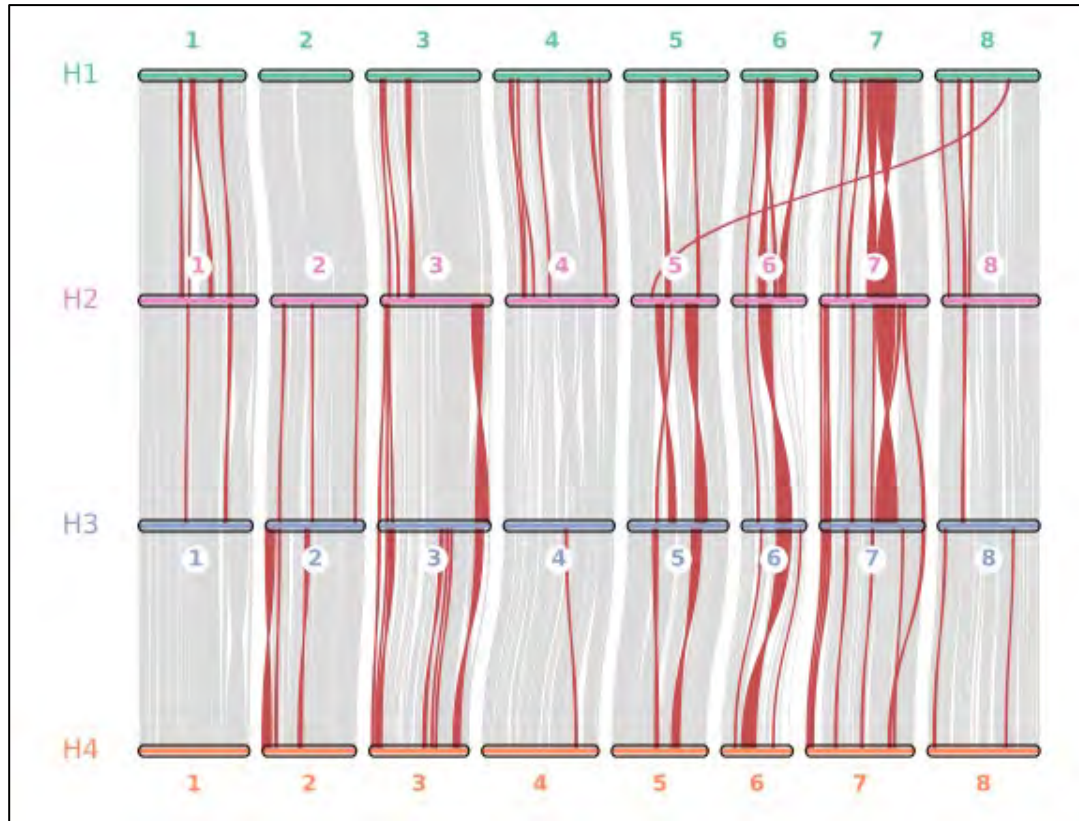
Regen-SY27x (version 1.0) Genome Assembly

Metric	All Haplogroups	Haplogroup 1	Haplogroup 2	Haplogroup 3	Haplogroup 4
Number of pseudo-chromosomes	32	8	8	8	8
Total number of Scaffolds	6,021	1,043	1,033	1,035	2,910
Total number of Contigs	7,162	1,331	1,348	1,306	3,177
Scaffold Length (bp)	3,286,073,882	789,244,119	789,460,494	793,088,424	914,280,845
Contig N50 (bp)	3,406,605	4,276,997	2,935,843	3,505,361	3,106,875
Scaffold N50 (bp)	79,699,470	79,699,470	73,507,436	75,653,446	81,593,309
Total number of genes	221,688	54,412	55,178	54,512	57,586
BUSCO completeness (%)	99.3	94.9	95.9	94.4	96.9
Number of R genes only	3,696	970	924	888	914

- Repeat annotation revealed that 62.65% of the assembly consists of interspersed repeats, dominated by long terminal repeat (LTR) retrotransposon (36.7%) with 5.5% helitron content.
- A total of 3,696 NLR genes were identified dominated by TNL and RxNL subclasses and frequently arranged in large tandem clusters.
- Nearly all (98.6%) of the BUSCOs were duplicated within the assembly.
- 75% of the found BUSCOs were present in exactly four copies and,
- 82% had at least one copy per haplotype.

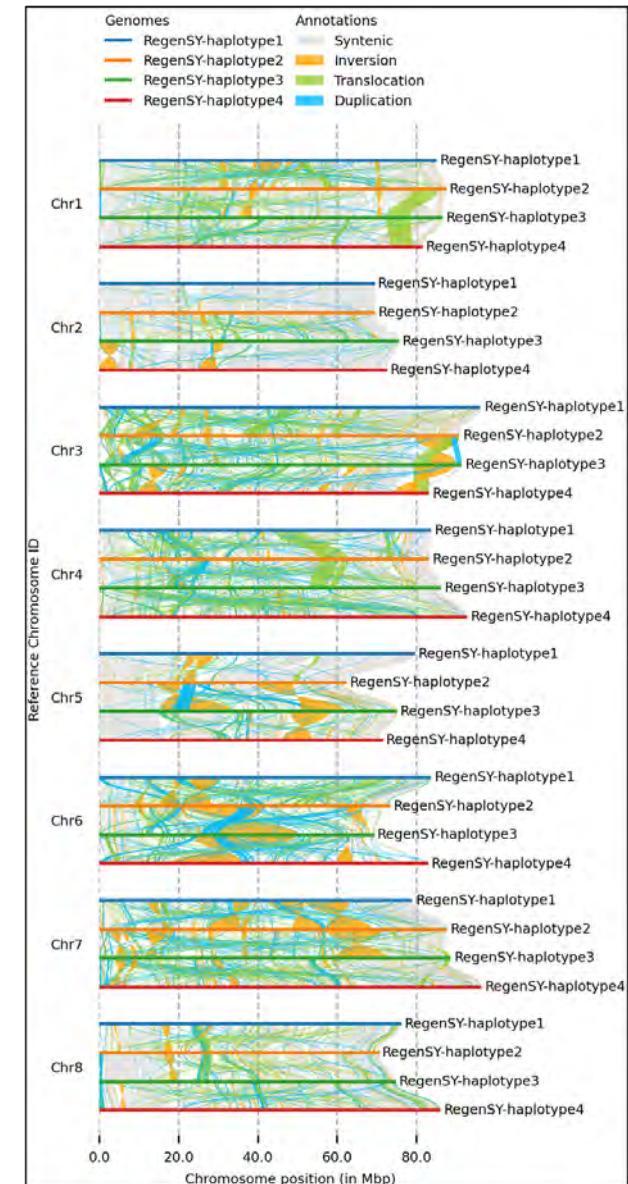
Kaur, H., Connor, C., Farmer, A., Mudge, J., Shannon, L. M., and Samac, D. A. (2026). Haplotype-resolved genome assembly of an autotetraploid *Medicago sativa* cv. RegenSY genotype uncovers large scale structural variation and gene family dynamics. G3: Genes|Genomes|Genetics. (Under Review)
<https://doi.org/10.64898/2026.05.01.722254> (Preprint).

Comparative Analyses

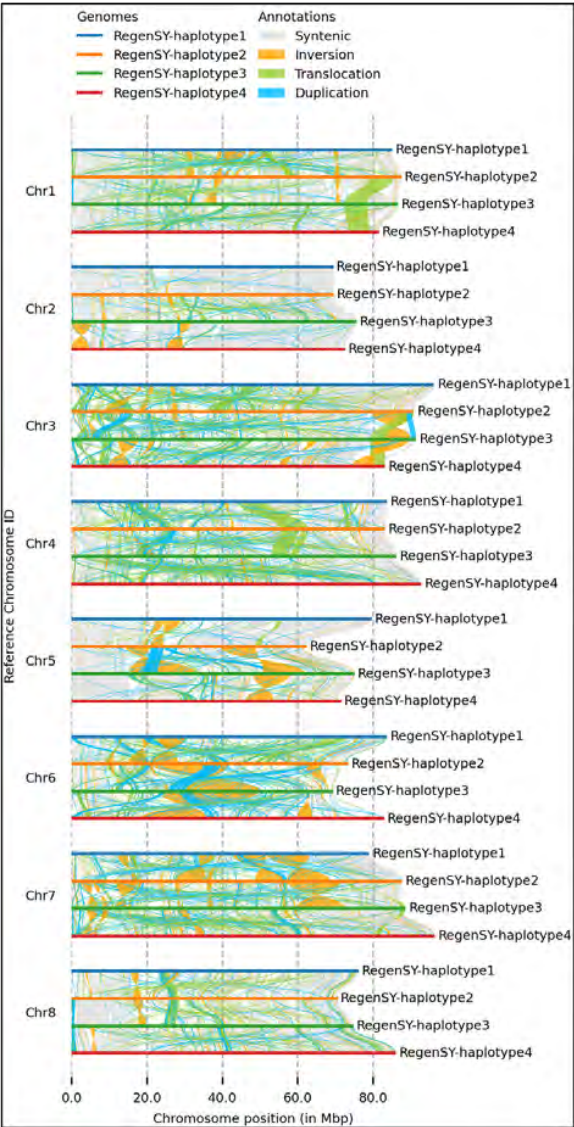


RegenSY-27x Gene Based Synteny between 4 haplotypes

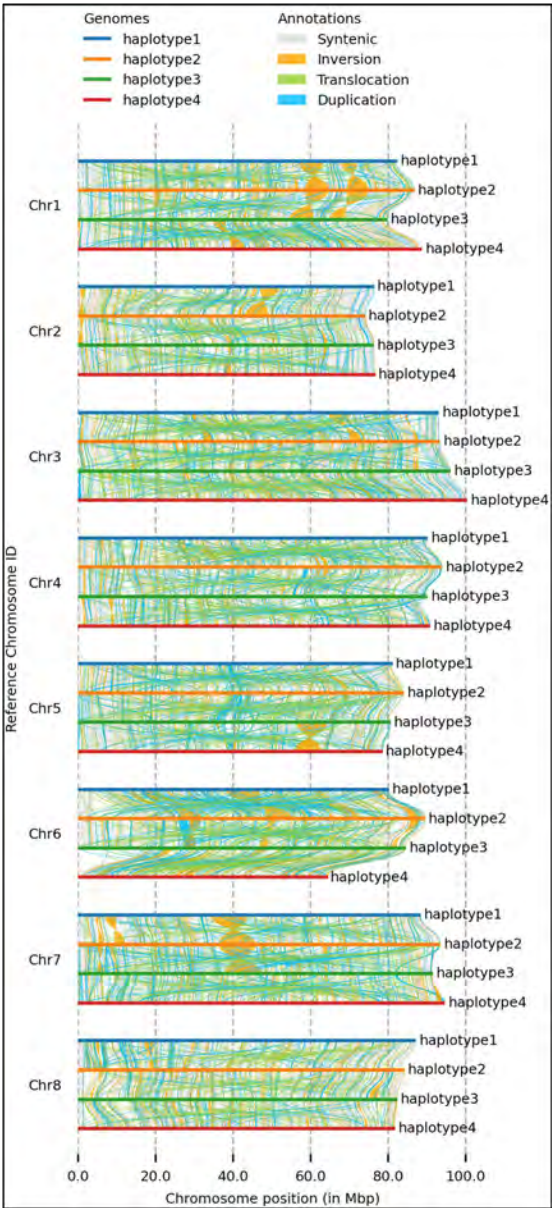
Whole genome sequence comparison



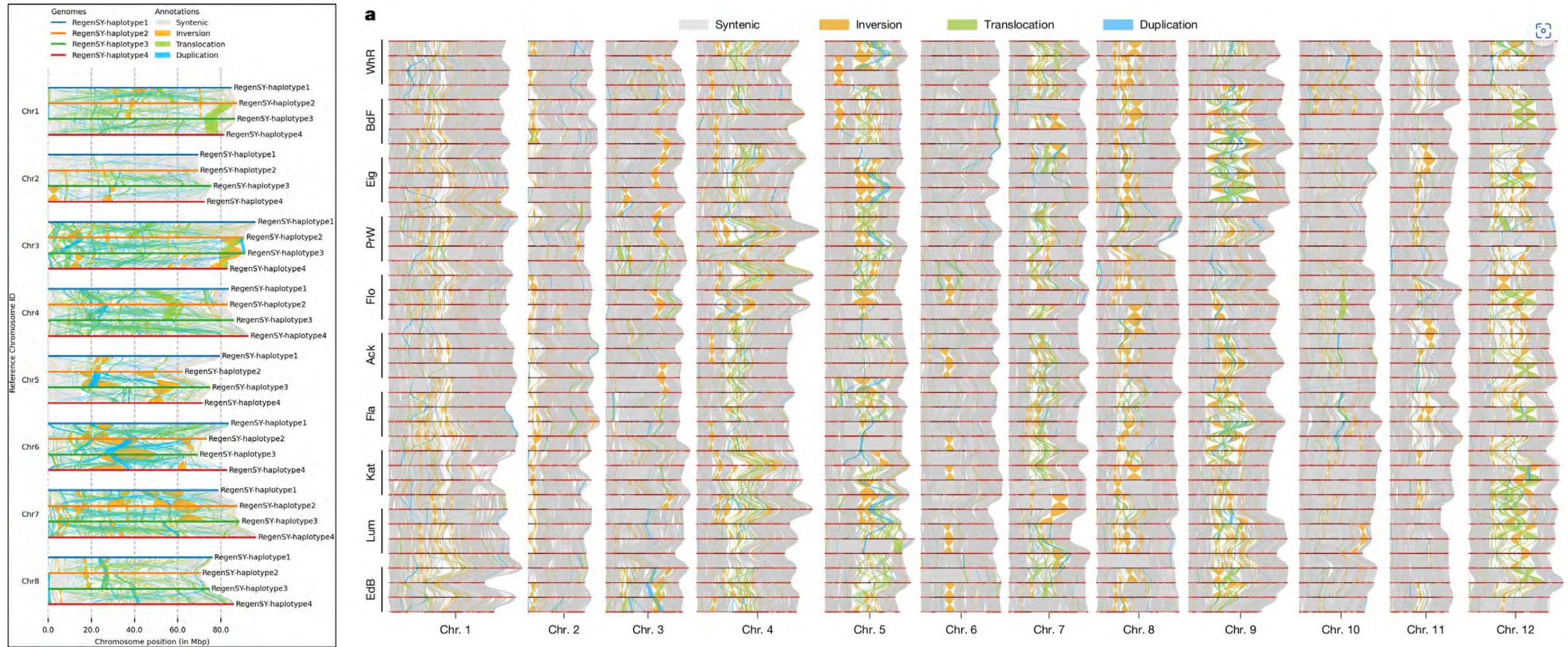
RegenSY-27x



Another alfalfa genome- Xin Jiang DaYe (Chen et al. 2020)



Another autotetraploid highly heterozygous assembly- Potato genomes



Source: Sun, H., Tusso, S., Dent, C.I. et al. The phased pan-genome of tetraploid European potato. *Nature* 642, 389–397 (2025). <https://doi.org/10.1038/s41586-025-08843-0>

Utilize Structural Variations

- Assembling other 8 alfalfa assemblies
 - Varying in Fall Dormancy levels
 - Five new mapping populations and their linkage maps developed using GBS data

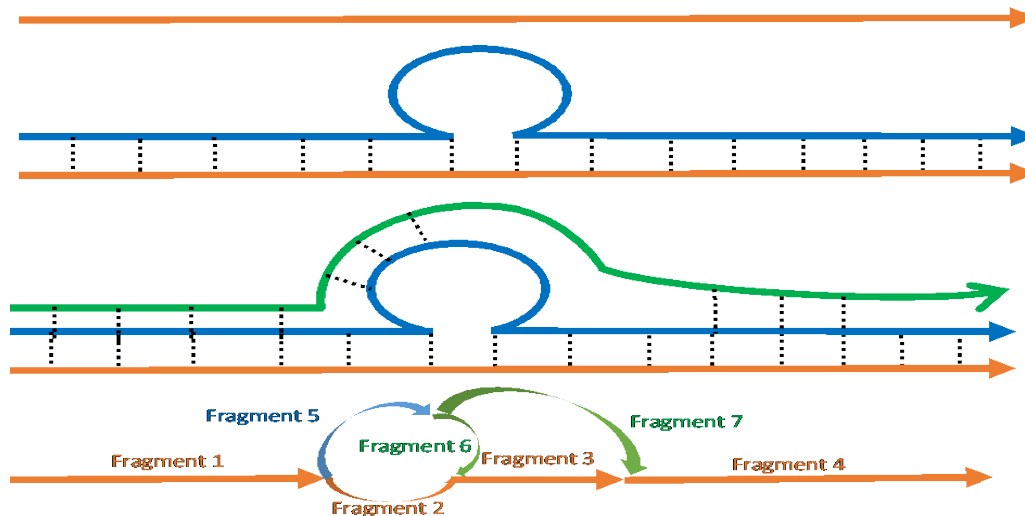
Population	Populations Size
Legacy4 × S&W7	372
Legacy5 × GAMS1404	372
RegenSY27x × FL99	372
S&W6 × CW1010	372
S&W4 × FL99	562

Will help us identify SVs at larger scale- to find diversity for different traits using linkage/QTL mapping, association studies

How to use Phased Genome Assemblies?

- It has advantages – Allele Specific Gene Editing
- But lot of modern tools can only use one haplotype as reference genome
- Solution – Pangenome (Kaur et al. 2024)

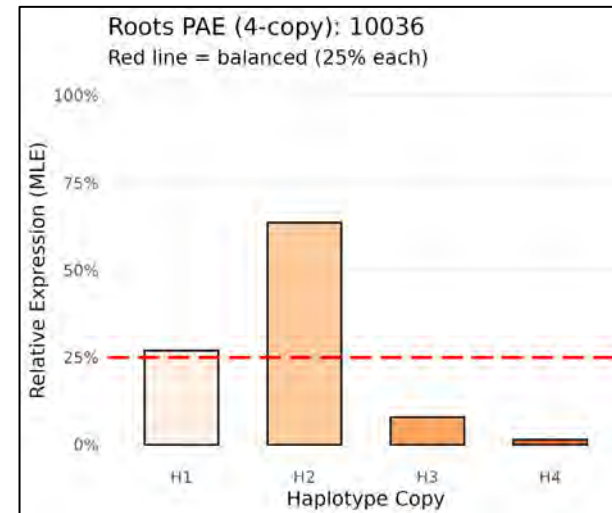
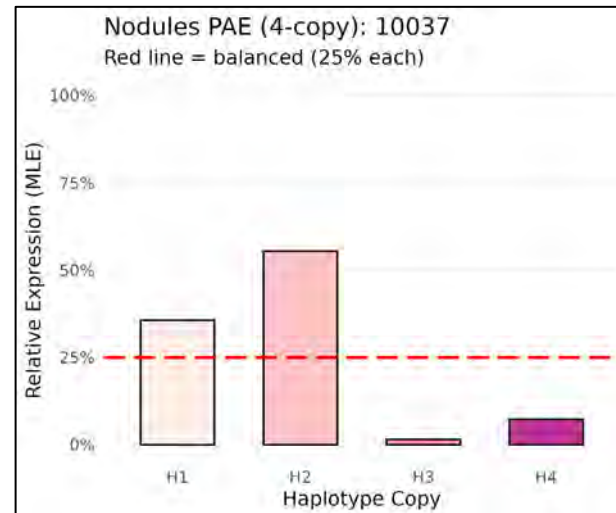
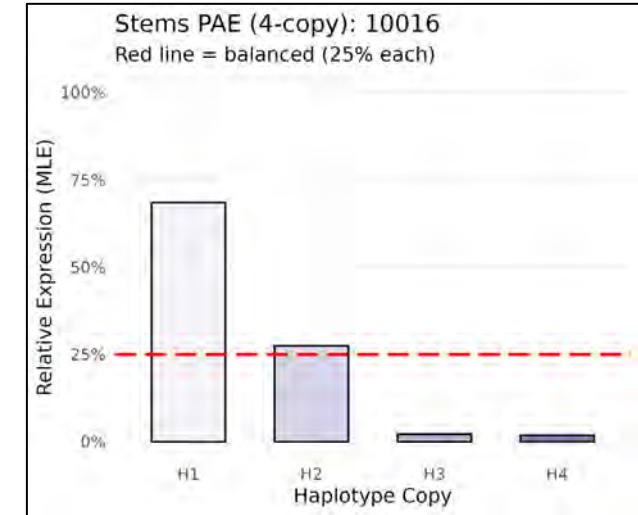
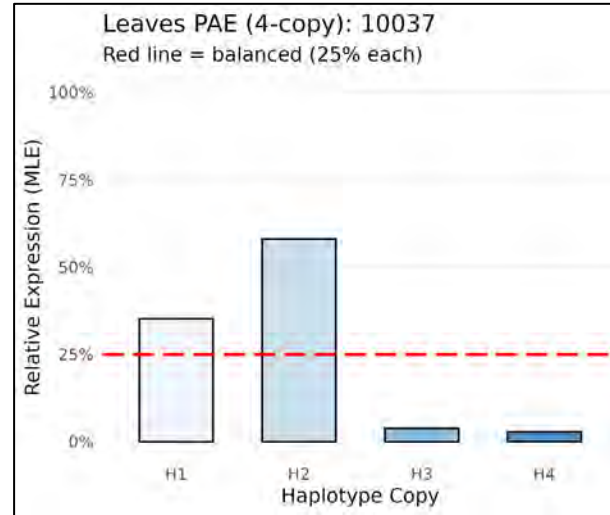
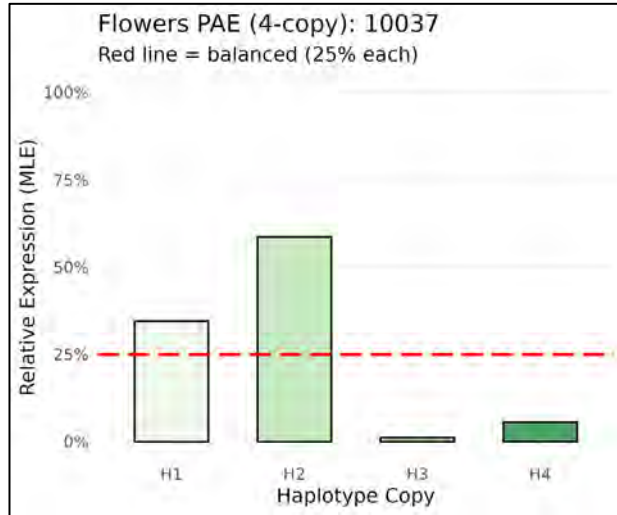
Kaur, H., Shannon, L. M., & Samac, D. A. (2024). A stepwise tool guide for pan-genome development in crop plants improvement: An alfalfa (*Medicago sativa*) case study. BMC Genomics 25, 1022. <https://doi.org/10.1186/s12864-024-10931-w>



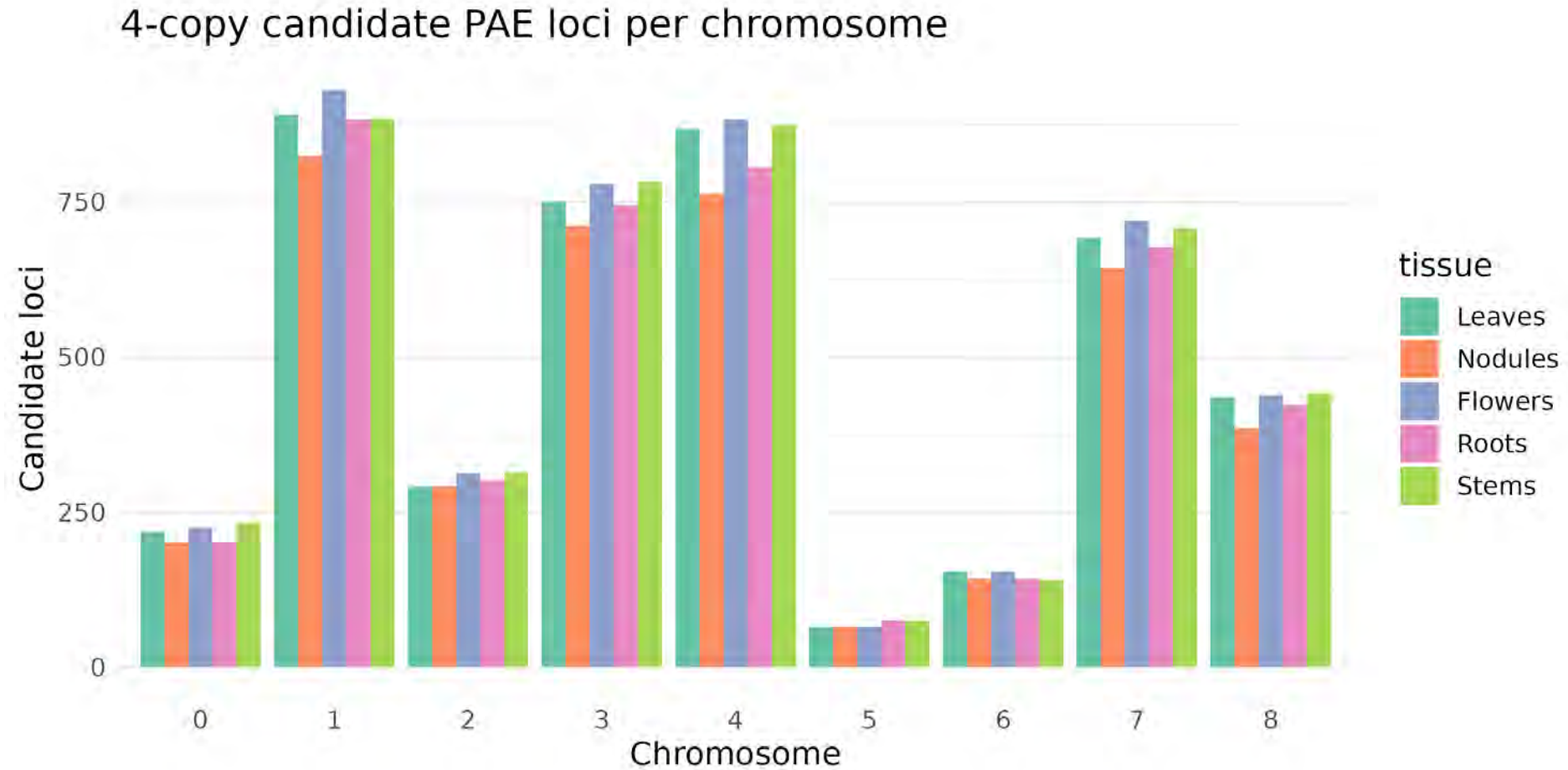
- Linearized form of graph-based genome
 - Haplotype 1 + sequences from other haplotypes
- Used in linkage Maps
- Exploring its utility in Preferential allele analyses

Kaur, H., Shannon, L. M., & Samac, D. A. (2026). Impact of increased population size and three reference genomes on the development of linkage maps in alfalfa. Crop Science. (Under review).

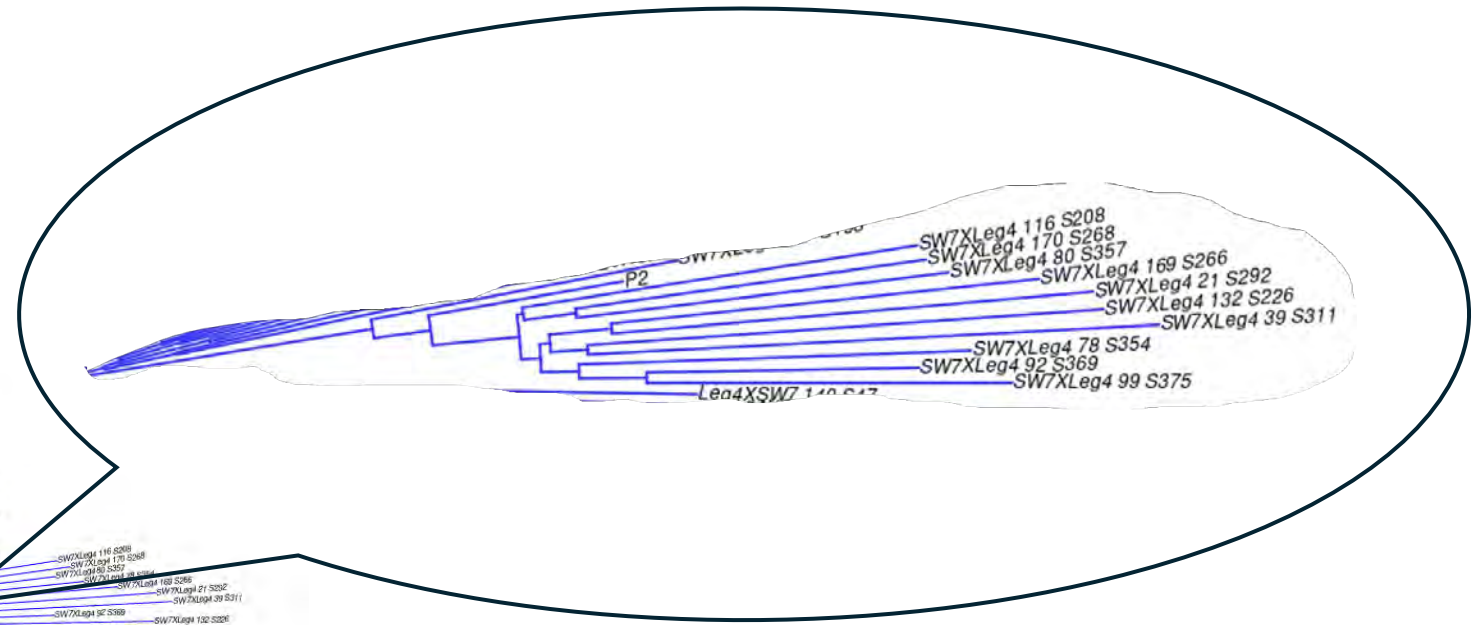
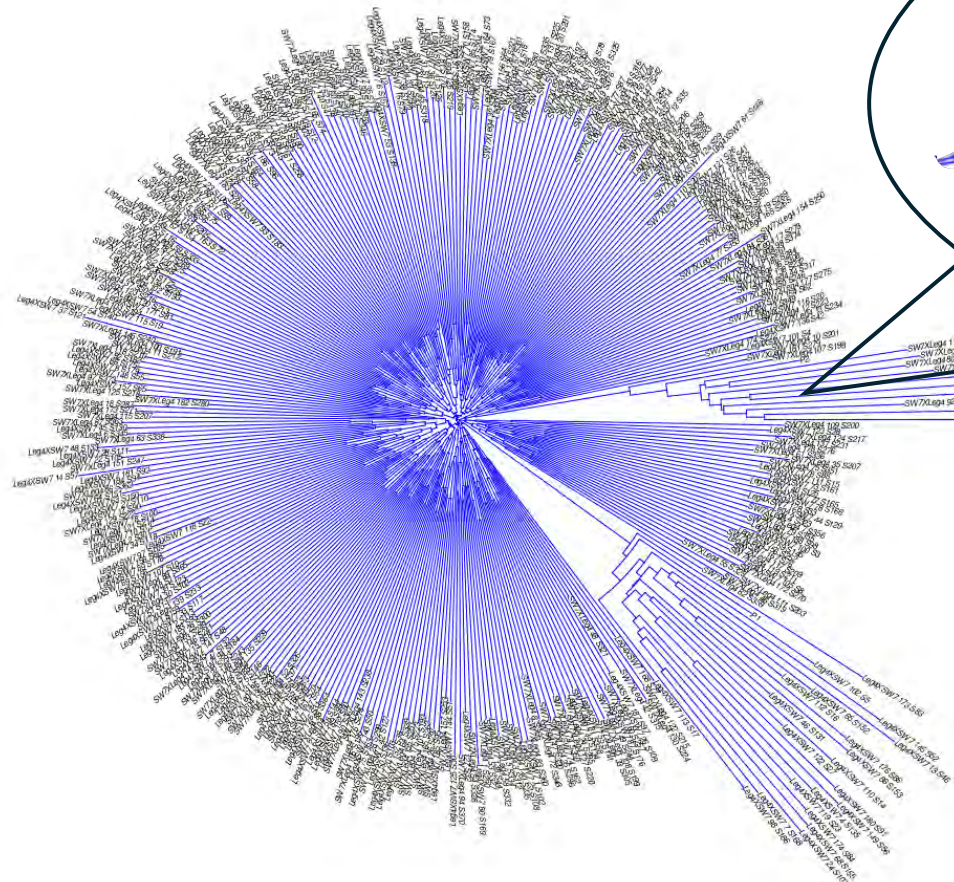
Preferential Allele Specific Analyses - RNAseq



Preferential Allele Expression (PAE) - RNAseq



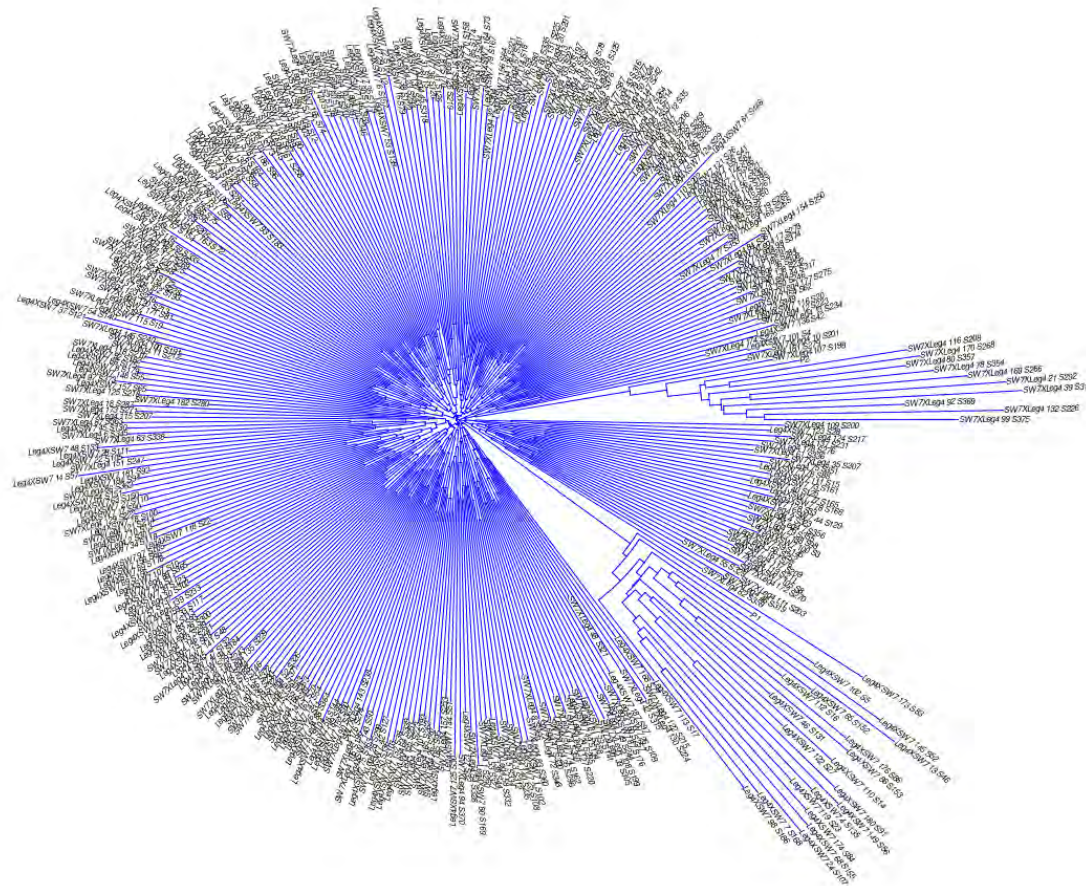
S&W7 × Legacy4



SNP-based genetic relationships identify candidate selfed progenies in the biparental population

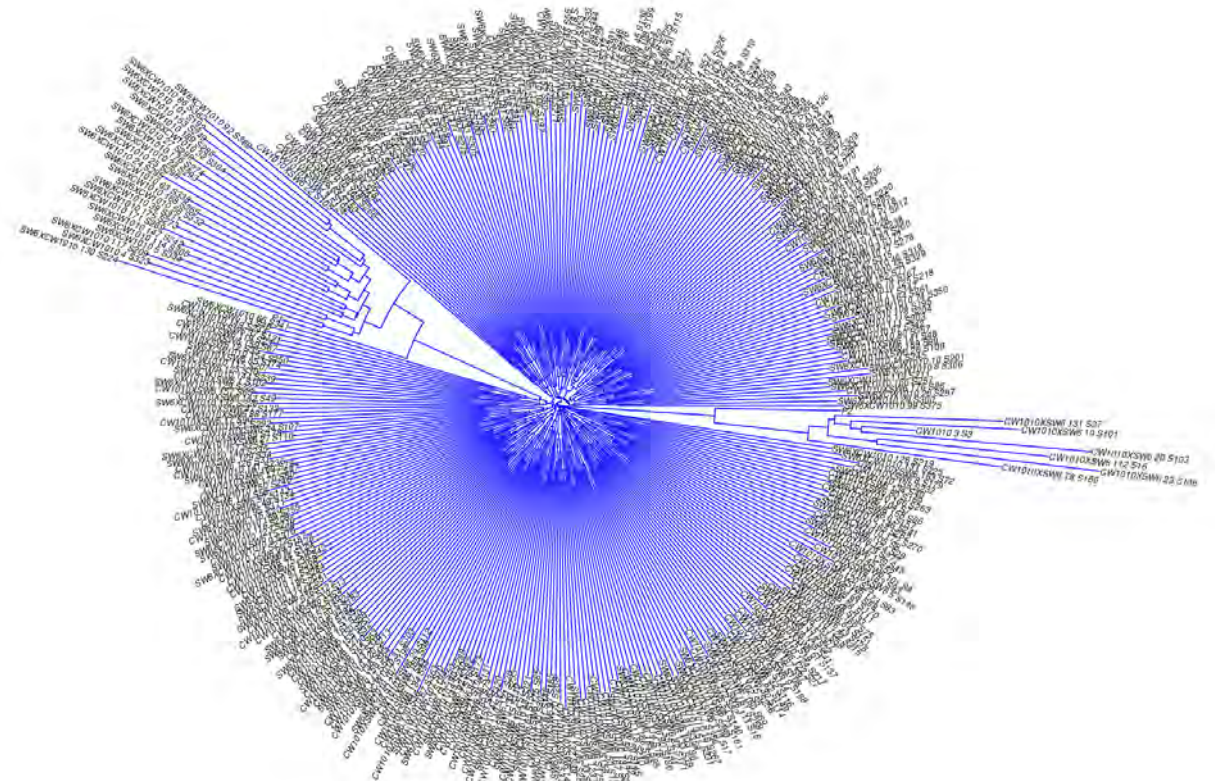
SNP- based Neighbor joining tree

S&W7 × Legacy4



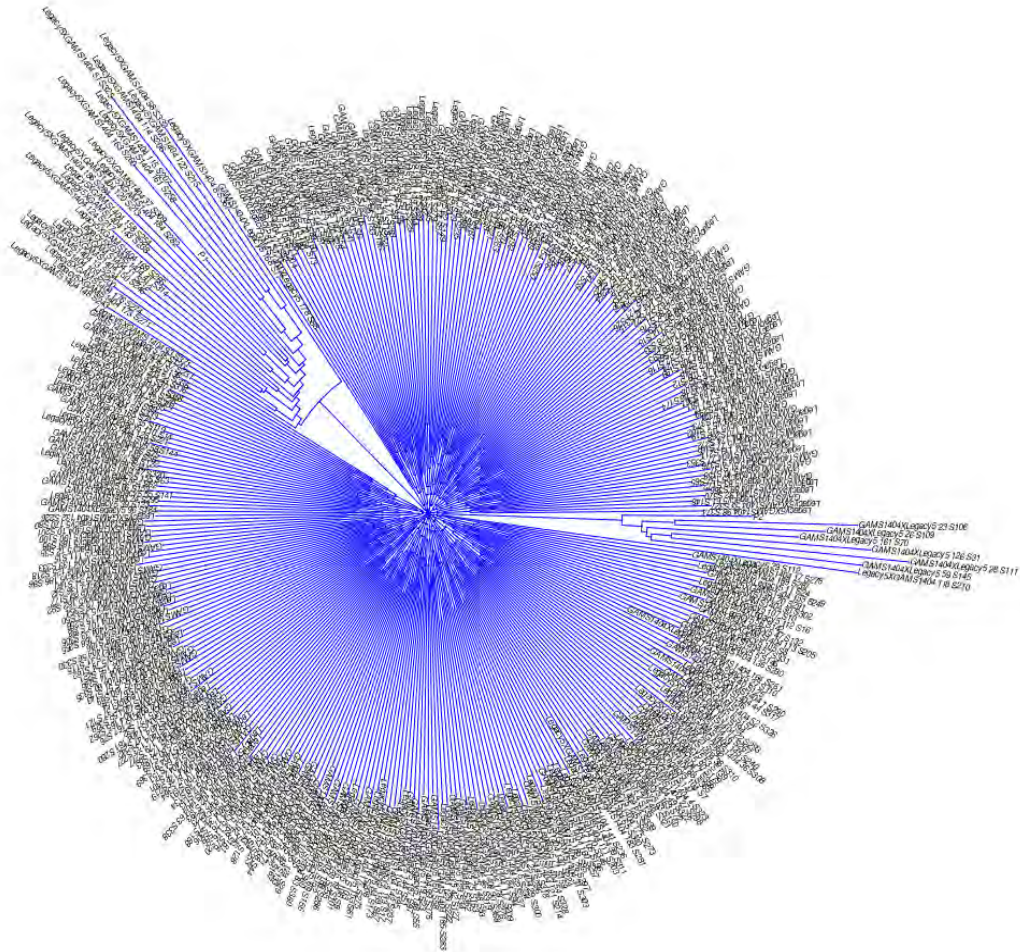
Population size - 372
putative selfed progenies - 30

S&W6 × CW1010



Population size - 372
putative selfed progenies - 29

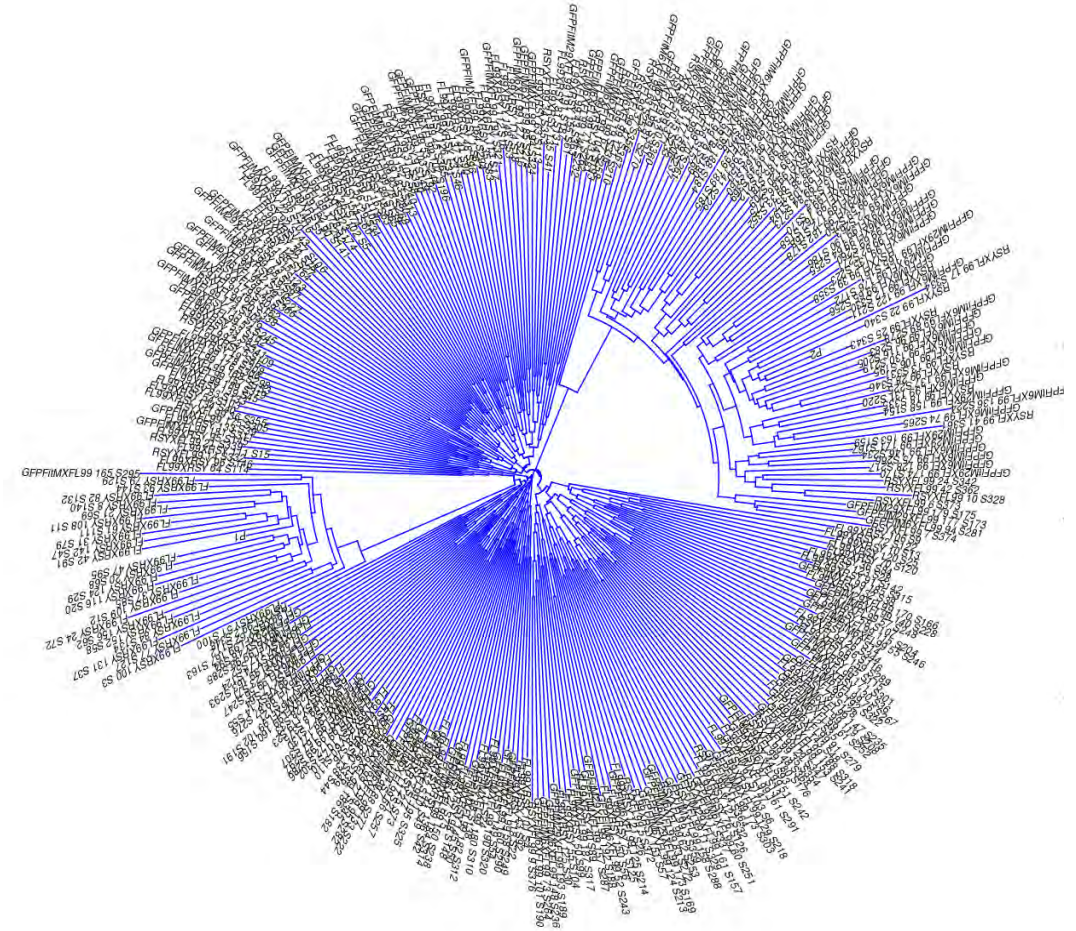
Legacy5 × GAMS1404



Population size - 372

putative selfed progenies - 30

RegenSY27x × FL99



Population size - 372

putative selfed progenies - 103

Improvement in Linkage Map (RegenSY-27x × FL99)

Before removing plausible selfs

LG	Map length (cM)	Markers/cM	Total	Max gap
1	257.93	2.34	603	13.27
2	155.06	1.97	306	12.08
3	217.65	2.36	513	16.22
4	195.12	2.72	530	29.23
5	53.04	2.60	138	8.81
6	146.88	1.82	268	8.86
7	201.85	2.65	535	16.28
8	201.07	2.18	438	10.34
Total	1428.6	2.33	3331	14.39

After removing plausible selfs

LG	Map length (cM)	Markers/cM	Total	Max gap
1	86.79	8.65	751	3.52
2	86.98	6.54	569	5.13
3	78.25	8.32	651	3.04
4	83.04	8.74	726	2.90
5	91.48	3.38	309	9.04
6	100.94	3.56	359	11.19
7	71.86	9.57	688	3.72
8	78.12	6.86	536	2.87
Total	677.46	6.95	4589	5.18

Distorted Segregation came down from ~30% to 10-12%

Improvement in Linkage Map (S&W7 × Legacy4)

Before removing putative selfs

LG	Map length (cM)	Markers/cM	Total	Max gap
1	522.87	1.24	646	13.25
2	289.89	1.51	437	10.98
3	399.02	1.17	465	27.49
4	318.65	1.95	620	7.53
5	394.47	1.28	506	11.72
6	272.03	1.22	331	15.74
7	402.64	1.45	582	8.42
8	377.53	1.25	471	22.3
Total	2977.1	1.38	4058	14.68

After removing putative selfs

LG	Map length (cM)	Markers/cM	Total	Max gap
1	82.55	9.98	824	3.05
2	83.55	8.64	722	2.47
3	79.12	7.76	614	4.42
4	83.58	9.83	822	4.45
5	80.07	7.42	594	4.12
6	80.29	5.65	454	4.79
7	77.83	8.38	652	2.56
8	99.65	6.26	624	6.55
Total	666.64	7.99	5306	4.05

Plausible Selfs

Female Parent	Male Parent	# Selfs	Total Putative Selfs	Total Crosses
S&W4	Legacy4	10	30 (8.1%)	372
Legacy4	S&W4	20	30 (8.1%)	372
GAMS1404	Legacy5	6	30 (8.1%)	372
Legacy5	GAMS1404	24	30 (8.1%)	372
S&W6	CW1010	23	29 (7.8%)	372
CW1010	S&W6	6	29 (7.8%)	372
FL99	RegenSY	24	103 (27.71%)	372
RegenSY	FL99	79	103 (27.71%)	372

Plausible Selfs Summary – % heterozygosity

Female Parent	Male Parent	Female Parent	Male Parent	Selfs	Crosses
S&W4	Legacy4	73.3	67.8	57.0–66.9	66.4–83.5
Legacy4	S&W4	67.8	73.3	52.9–60.8	66.7–86.9
GAMS1404	Legacy5	72.6	70.8	57.2–69.4	54.1–83.6
Legacy5	GAMS1404	70.8	72.6	57.1–62.7	68.2–82.3
S&W6	CW1010	69.9	62.1	43.3–55.0	62.7–76.7
CW1010	S&W6	62.1	69.9	43.4–50.9	63.5–70.2
FL99	RegenSY	67.8	69.3	52.4–57.7	64.2–71.9
RegenSY	FL99	69.3	67.8	54.6–60.1	65.4–73.1

Plausible Selves



Hybrids



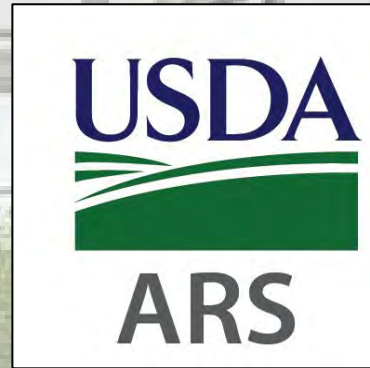
Also, based on Aphanomyces disease screening results- hard to differentiate between selves vs. hybrids

Summary

- Candidate selfs cluster close to one parent in the SNP-based tree.
- Candidate selfs have consistently lower heterozygosity than their maternal parent.
- Candidate selfs generally have lower heterozygosity than crossed progeny.
- Together, these patterns are consistent with unintended self-fertilization.
- During attempted cross-pollination, selfed and crossed offspring may have undergone postzygotic viability selection.
- Less-viable selfed embryos were likely eliminated through selective embryo abortion, leaving a subset of vigorous selfed progeny capable of surviving alongside the crossed progeny

Acknowledgements

- Debby Samac, USDA-ARS
- Mindy Dornbusch , USDA-ARS
- Laura Shannon, UMN
- Andrew Farmer, NCGR
- Joann Mudge, NCGR
- Cameron Connor, NCGR
- Adam Gomez, NCGR



A photograph of a large greenhouse filled with rows of potted plants. The plants are in yellow pots and are arranged in long, straight rows on black plastic trays. The greenhouse has a high, arched roof with a complex metal frame and translucent panels. The lighting is bright and even, suggesting a well-lit indoor environment. The text "Thank you!" is overlaid in the center of the image.

Thank you!