

ROSBREED

Enabling marker-assisted breeding in Rosaceae

ROSBREED'S SNP DETECTION PIPELINE AND COMMUNITY- AVAILABLE GENOMIC RESOURCES

Nahla Bassil, B. Gilmore, D. Main, C. Peace, T. Mockler, L. Wilhelm, E. van de Weg, T. Davis, D. Chagne, S. Gardiner, R. Crowhurst, I. Verde, B. Sosinski, P. Arus, R. Velasco, M. Troggio, A. Cestaro, G. Fazio, J. Norelli, J. Rees and A. Iezzoni



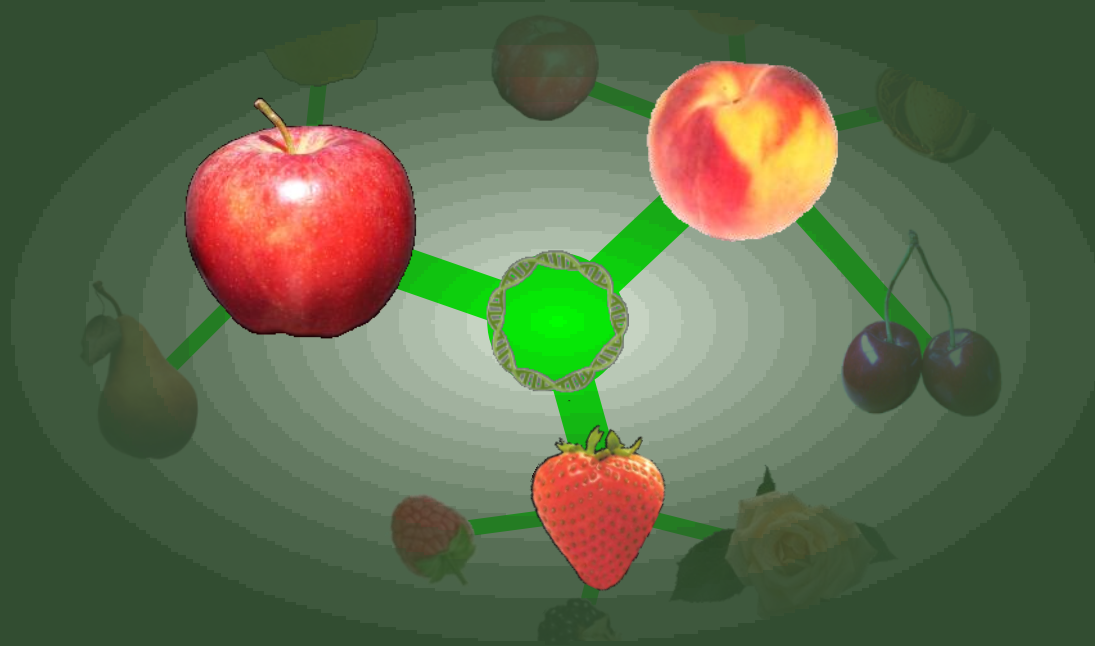
Available Rosaceae Genomic Resources

- ❖ SNP detection pipeline-to upload on GDR
- ❖ Short sequence reads and SNPs for apple, peach, cherry and strawberry-GenBank
- ❖ Validated apple, peach and cherry SNPs
- ❖ Infinium SNP chip for apple (9k), peach (9k) and cherry (6k)



Outline of Presentation

- ❖ SNP detection pipeline
- ❖ Available sequences and SNPs for peach, cherry, apple and strawberry
- ❖ SNP Validation using the GoldenGate assay



SNP Detection Pipeline

SNP Calling Workflow (SOAP-Based)

Filter, trim, name illumina reads
(fastq)

Index reference genome for
SOAP

Align reads to indexed reference using SOAP

Break up SOAP output by scaffold/contig

Sort each SOAP output file by position

Run soapSNP on each sorted SOAP output file

Calculate statistics (average, stdev. of coverage depth; frequency distribution of Q scores) on Soap output

SNP Calling Workflow (SOAP-Based) – Cont.

Filter each soapSNP output file

For each Filtered SNP

❖ Determine nucleotides contributed by each genotype to the coverage at the SNP position

❖ Analyze data- did genotype contribute uniform (homozygous) or non-uniform (heterozygous) data?

GACCTAAGGTC
GACC^AAAGG^CC
GACC^AAAGG^CC
GACCTAAGG^CC
GACCTAAGG^AC

Reference

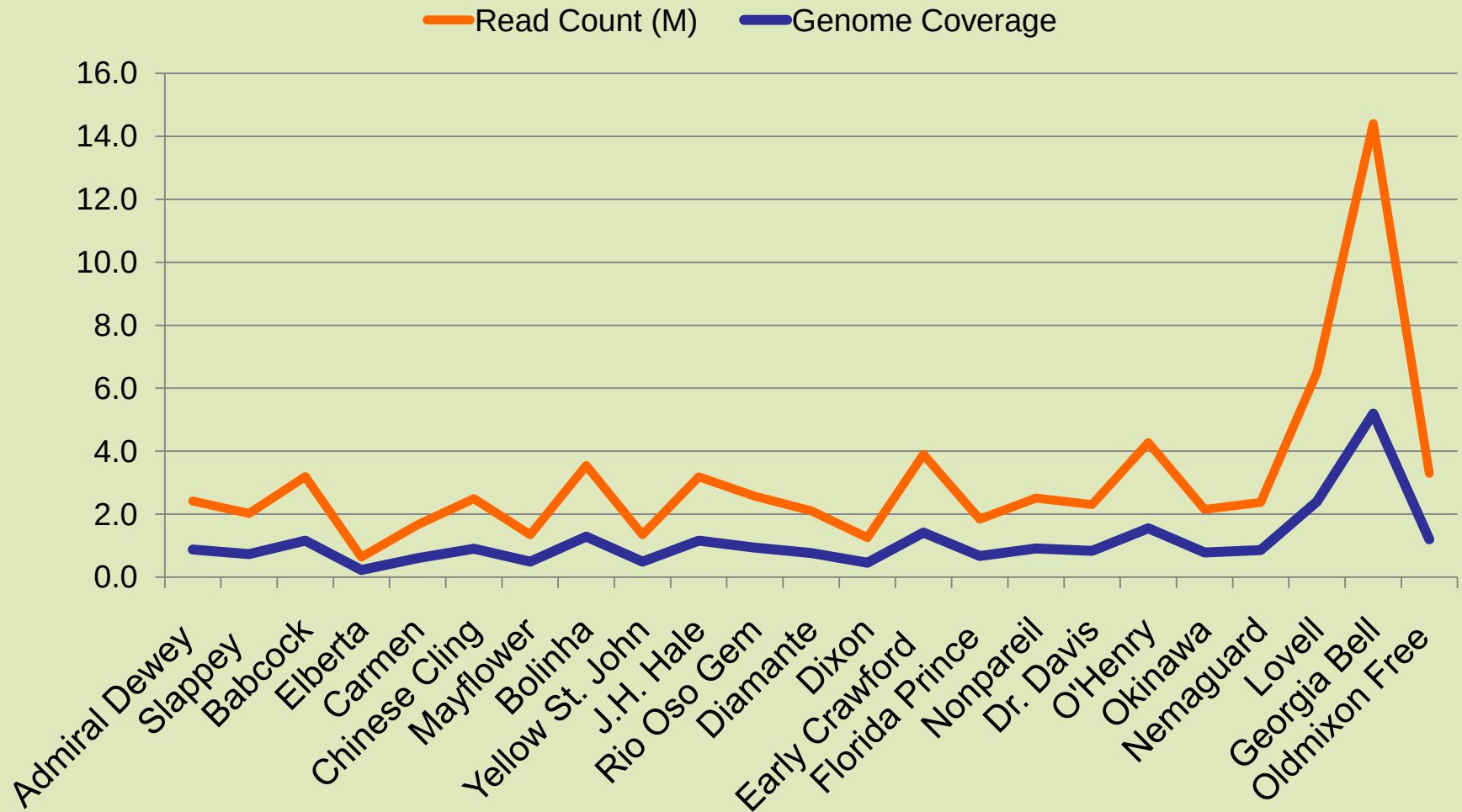
Accession A

Accession B

Homozygous
Accession A
SNP

Homozygous
Accession A SNP
Heterozygous
Accession B SNP

Peach SNP Detection Panel



5.7 Gb of data, 1.1 x coverage, on average



Addition of Eight Peach Accessions

 Libraries constructed from 8 varieties (Binaced, Big Top, O'Henry, Catherin, Sweet Cap, Venus, Elegant Lady, Nectaros) using nuclear DNA digested with *AluI* and size-selected for ~400-500bp fragments

 454 GS FLX sequencing with MID-labeled libraries

SNP Detection in Peach

Scaffold	Total SoapSnps	Filtering Parameters	
		Filter A	Filter B
scaffold_1	59798	7243	10043
scaffold_2	48193	6022	8789
scaffold_3	36144	5473	7554
scaffold_4	56427	6413	9004
scaffold_5	29583	3859	5261
scaffold_6	45496	6847	9403
scaffold_7	32042	2503	3513
scaffold_8	29796	2564	3455
other	16202	1600	599
Total	353,681	42524	57621

Filtering Parameters

Min # Reads: **8** (A)

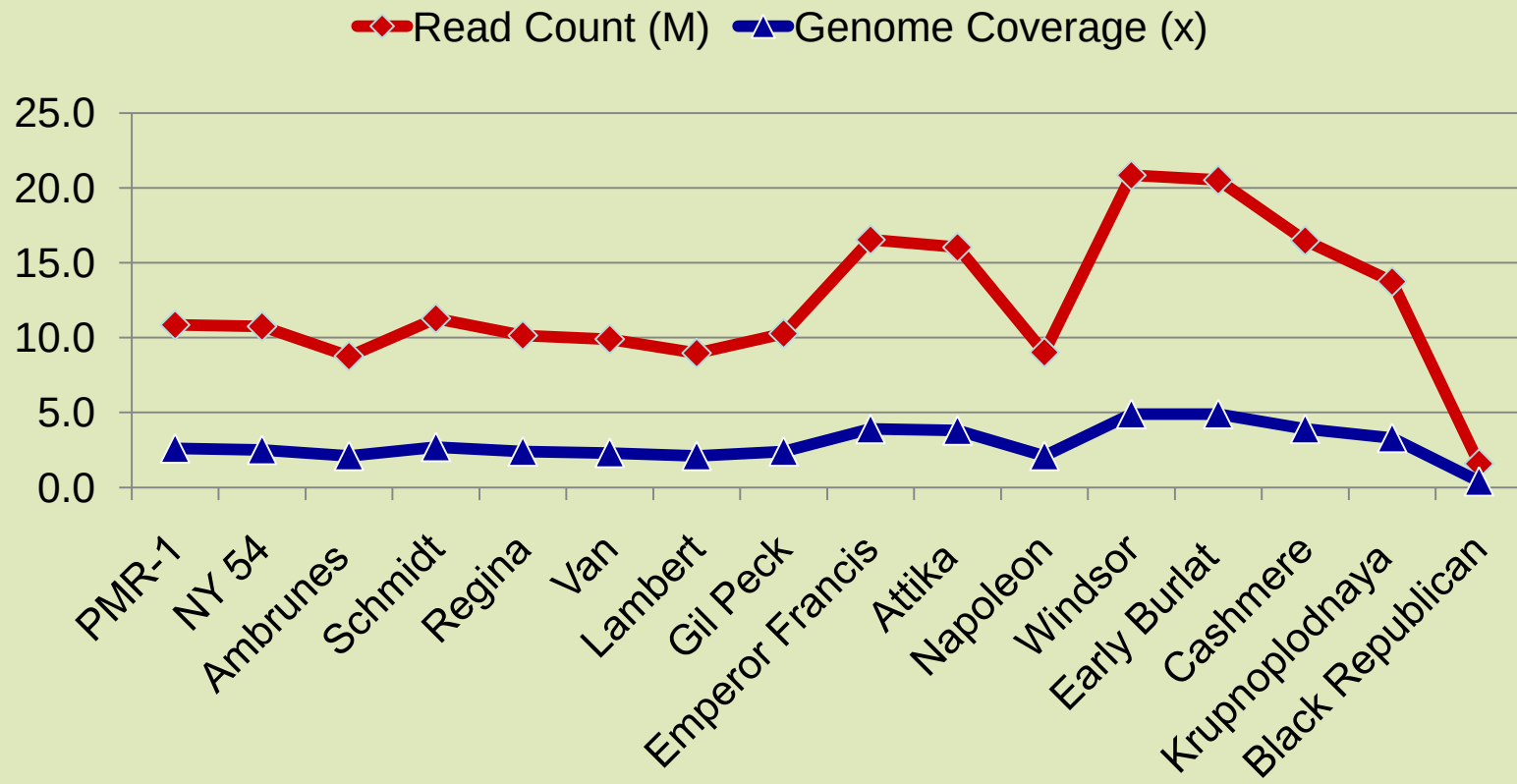
5 (B)

Max # Reads: **380**

Min GC: **30**

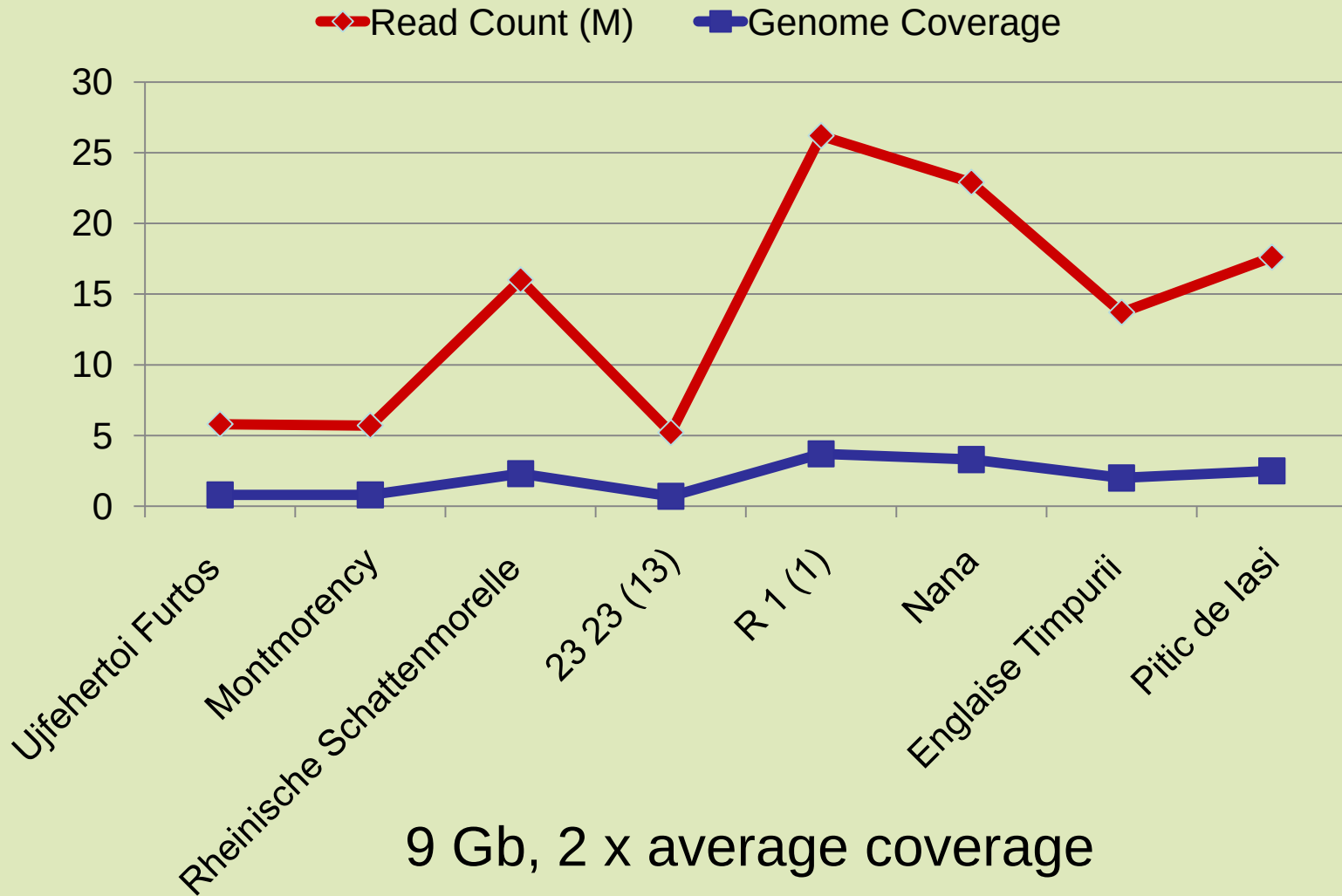


Sweet Cherry SNP Detection Panel



15.7 Gb, 2.1 Gb aligned, 2.9 x average coverage

Tart Cherry SNP Detection Panel



SNP Detection in Cherry

Scaffold	All SOAP Snps	Filtering Parameters		
		Filter A	Filter B	Filter C
scaffold_1	413,068	13,227	25,049	225,068
scaffold_2	222,921	6,595	12,574	114,147
scaffold_3	198,446	6,026	11,478	104,361
scaffold_4	232,552	6,401	12,405	117,869
scaffold_5	170,333	5,489	10,428	93,308
scaffold_6	252,525	8,116	15,334	135,591
scaffold_7	195,544	6,103	11,718	104,839
scaffold_8	182,445	6,035	11,176	95,758
other	32,861	794	1,509	14,719
Total	1,900,695	58,786	111,671	1,005,660

Filtering Parameters

Min # Reads: **8** (C)

20 (B)

30 (A)

Max # Reads: **380** (A, B)

1254 (C)

Min GC: **30** (A, B)

35 (C)



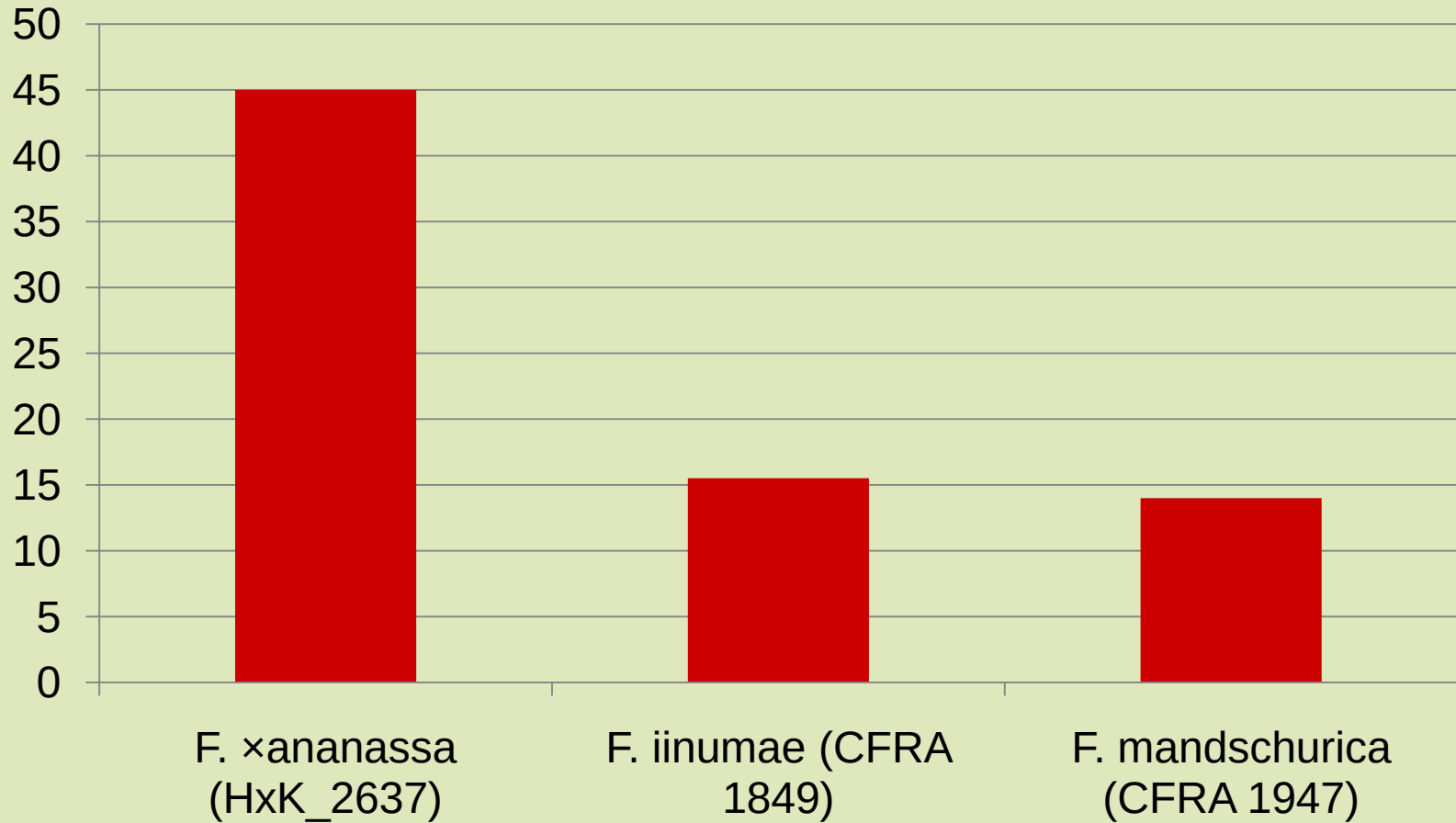
Apple SNP Detection Panel

RosBREED, US	ARCSA, SA	PFR, NZ
Crimson Crisp	Fuji	Delicious (red)
Honeycrisp	Cox's Orange Pippin	Cox's Orange Pippin
Haralson	Goldrush	Geneva
McIntosh	Frostbite	Braeburn
Rome Beauty Law	Zestar	Granny Smith
Delicious	Cripps Pink	James Grieve
Golden Delicious	Dolgo	Splendour
Jonathan	Northern Spy	Idared
F ₂ 26829-2-2	Wagener	Red Dougherty
Duchess of Oldenburg	Co-op 15	Ralls Janet



Fragaria Sequences

Genome Coverage of *Fragaria* (x)

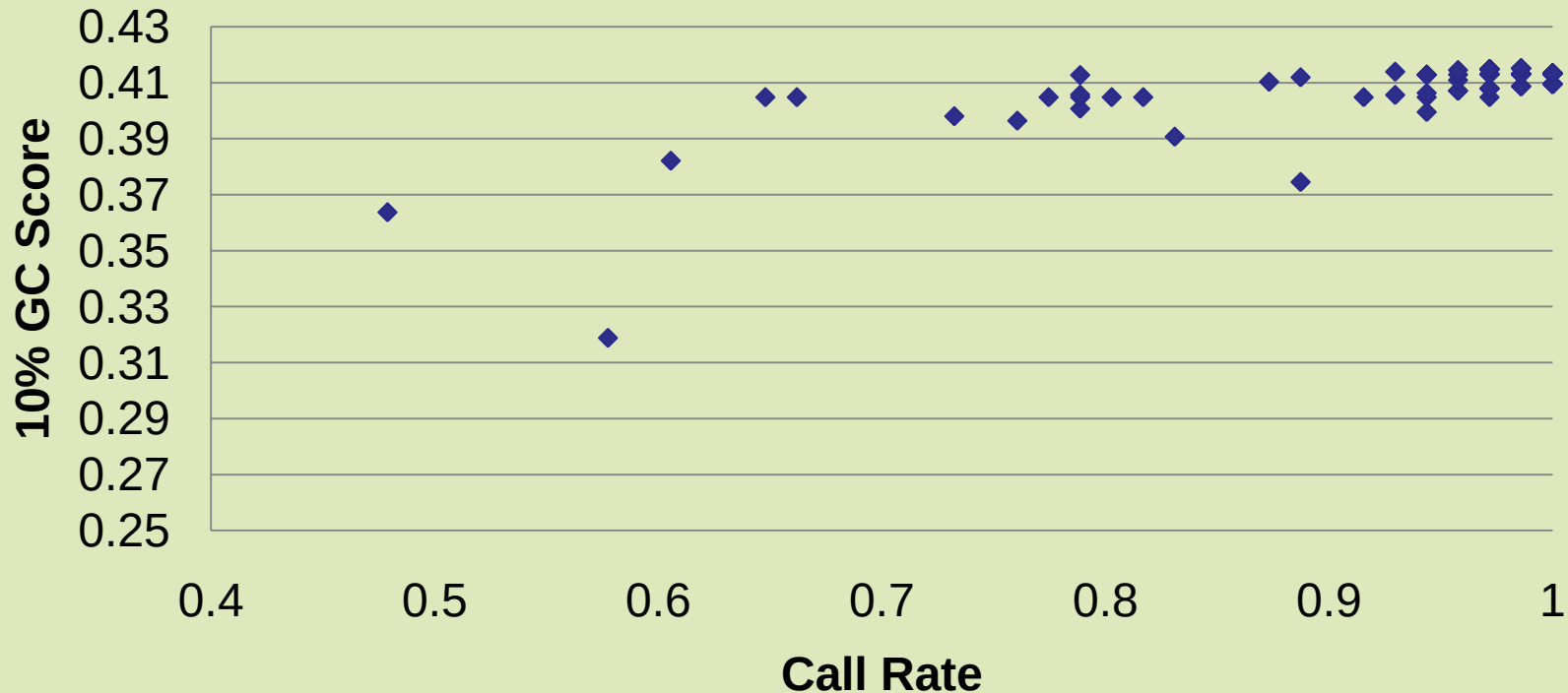


Criteria for SNP Validation Using GoldenGate Assay

- Equidistant on each scaffold 144, 96, 96, 48 in apple, peach, sweet and tart cherry
- Location in genome
(40%: genes; 20%: introns; 20% promoter; 20% intergenic)
- 20% accession-specific
- Zoomed-in region

Results of Validation in Peach

Call Rate x 10%_GC_Score

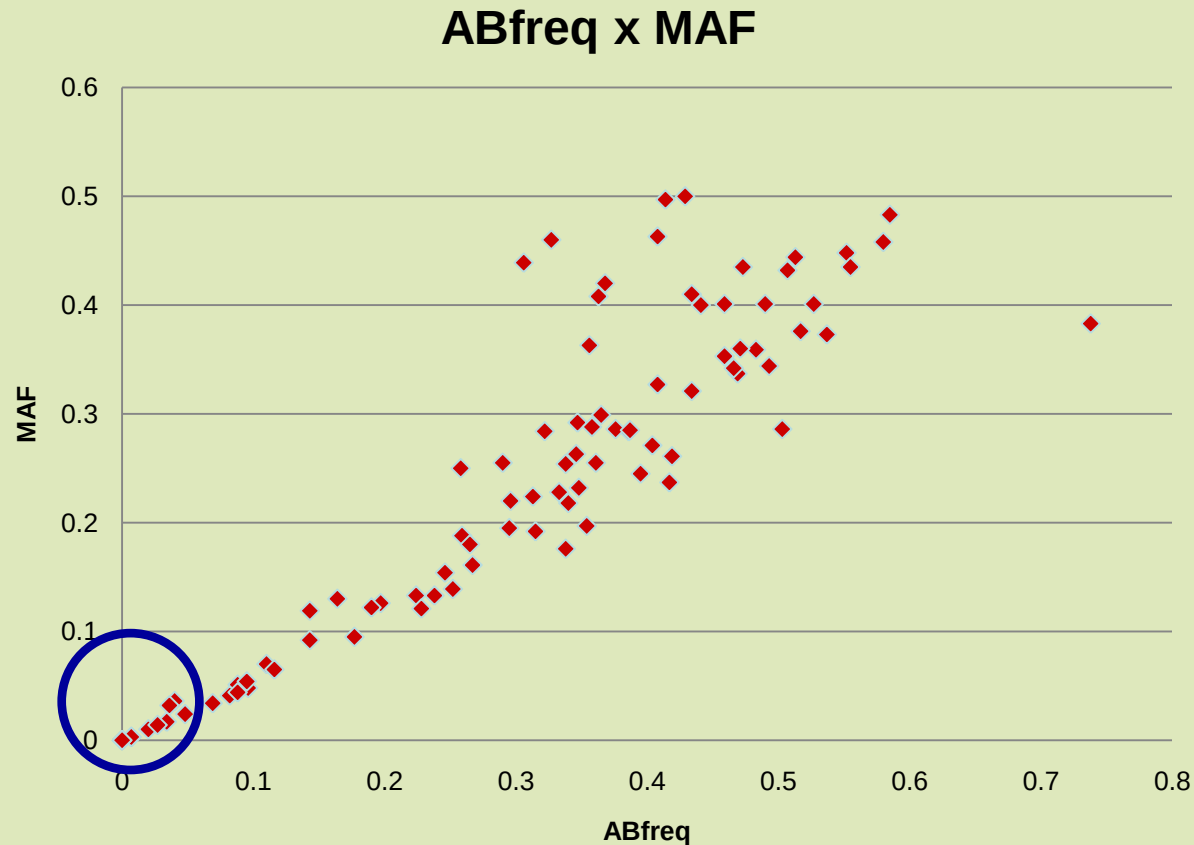


**Low performing samples: Call Rate < 0.82
& 10%GC Score < 0.39**

Thirteen Peach Accessions-Excluded from Analysis

DNA_Name	#No_Calls	#Calls	Call_Rate	10%_GC_Score
F8.5-159	30	41	0.5775	0.3188
Jordanolo	37	34	0.4789	0.3637
MB1-73 Hybrid F1	8	63	0.8873	0.3745
Nonpareil	28	43	0.6056	0.3821
S-37	17	54	0.7606	0.3964
F8.5-147	19	52	0.7324	0.398
5 (F2)	15	56	0.7887	0.4007
Mission	25	46	0.6479	0.4048
F10C.20-51	24	47	0.662	0.4048
Lovell	16	55	0.7746	0.4048
2005.17-207	15	56	0.7887	0.4048
83 (F2)	15	56	0.7887	0.4057
Vilmos	15	56	0.7887	0.4127

SNPs: Monomorphic or Polymorphic?



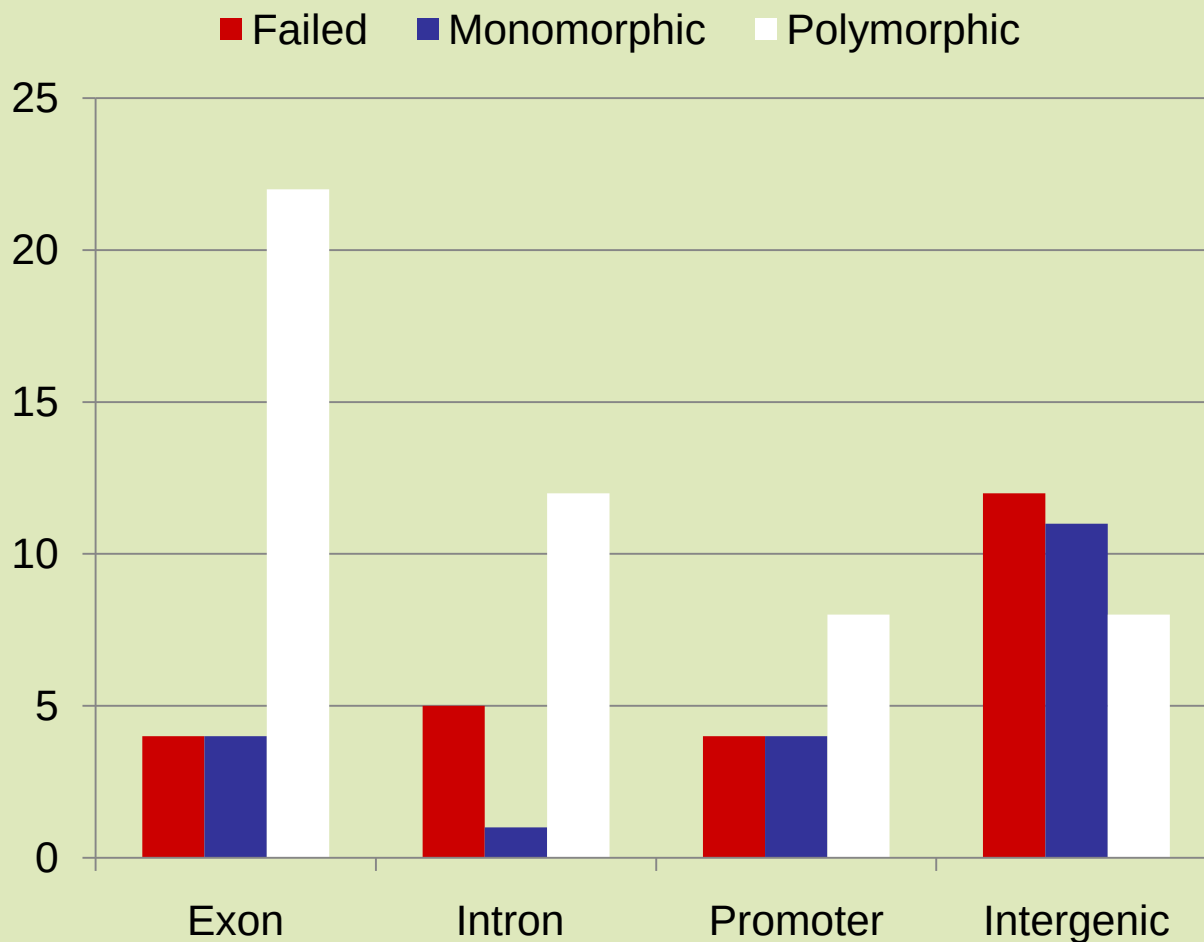
Monomorphic SNP: $MAF < 0.05$ OR A/B freq < 0.1

Overview of GoldenGate Validation Results

Crop	Failed	Monomorphic	Polymorphic	Total
Apple	46	25	73	144
Peach	25	20	51	96
Sweet Cherry	38	30	28	96
Tart Cherry	13	20	15	48
Total	122	95	167	384

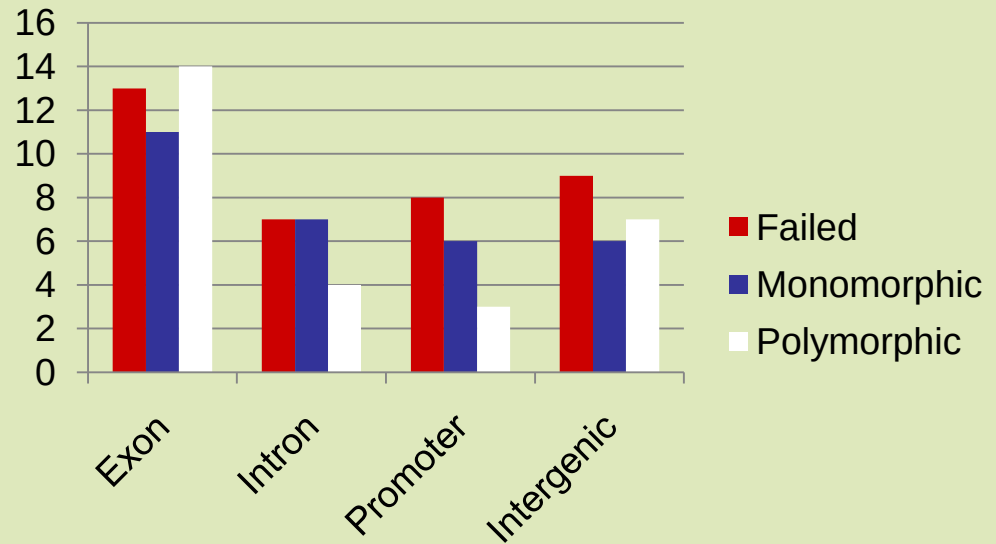


Effect of SNP Location (Peach)

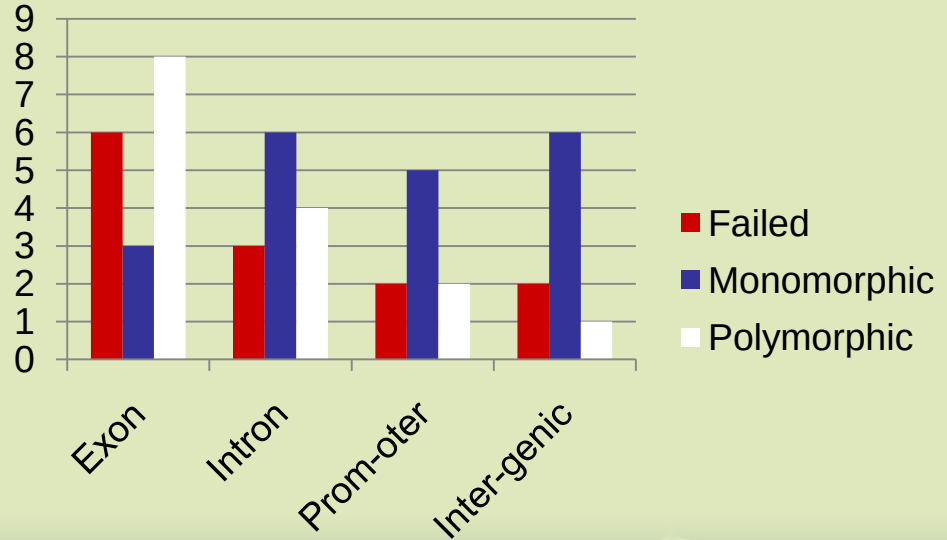


Effect of SNP Location (Cherry)

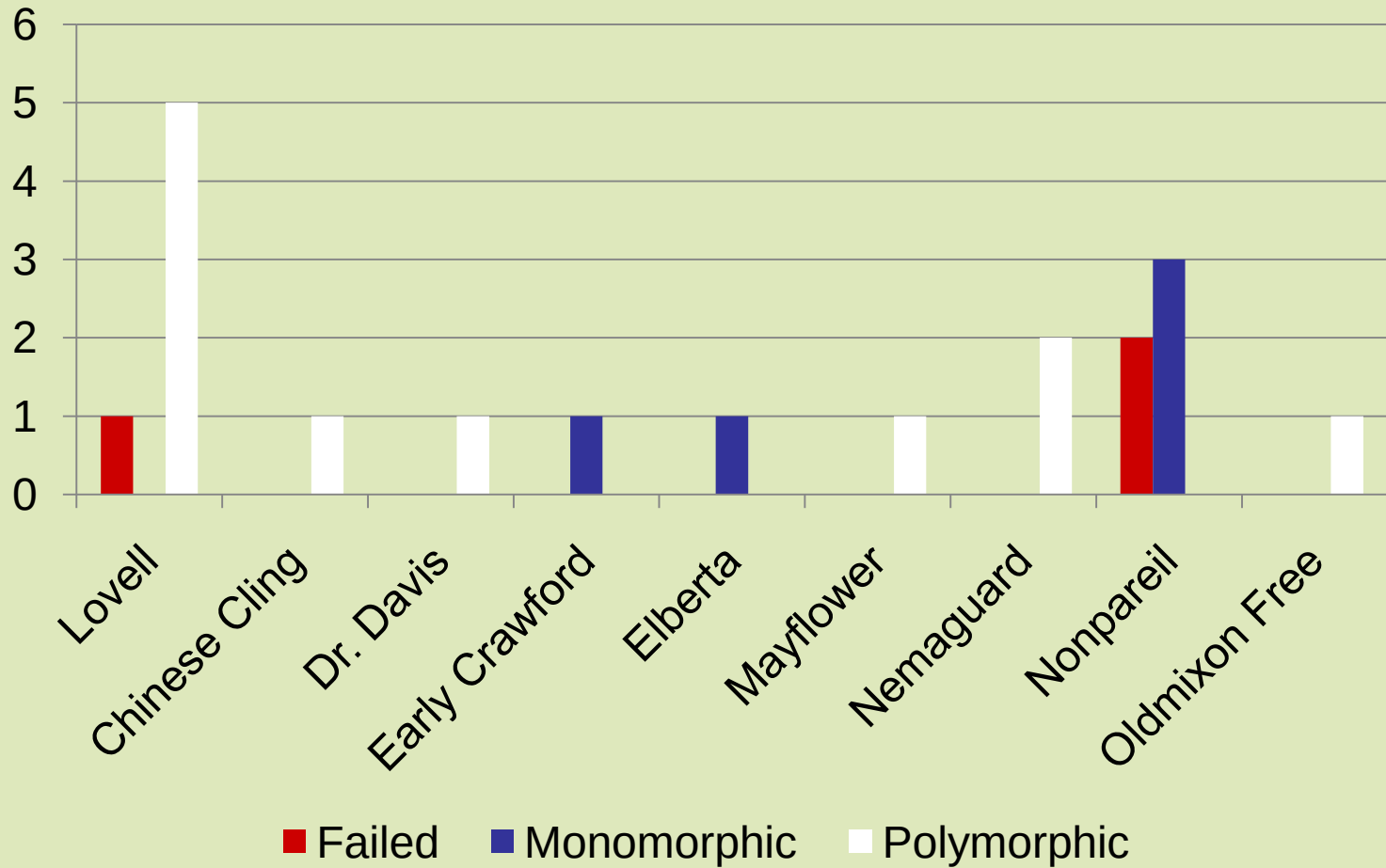
Sweet Cherry



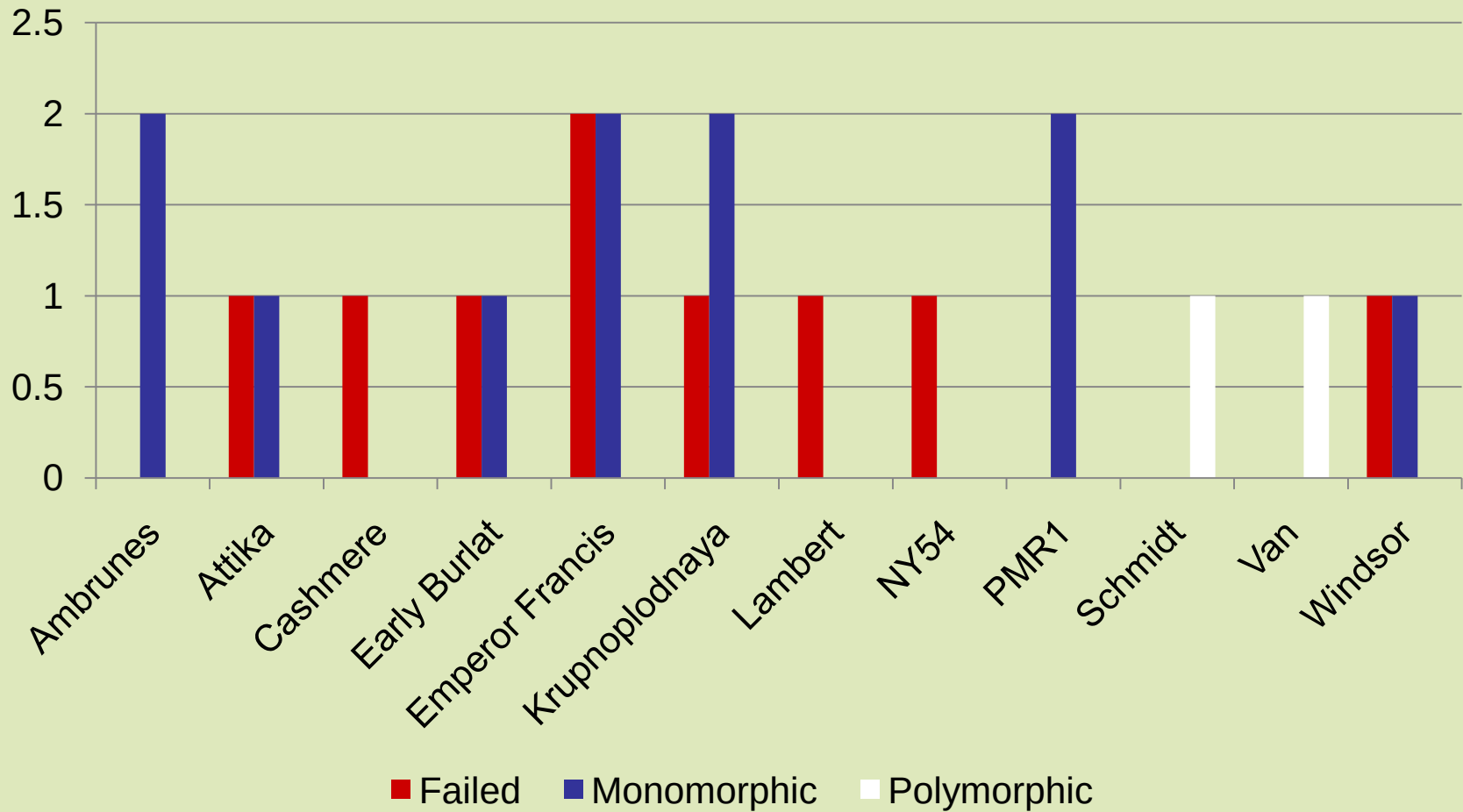
Tart Cherry



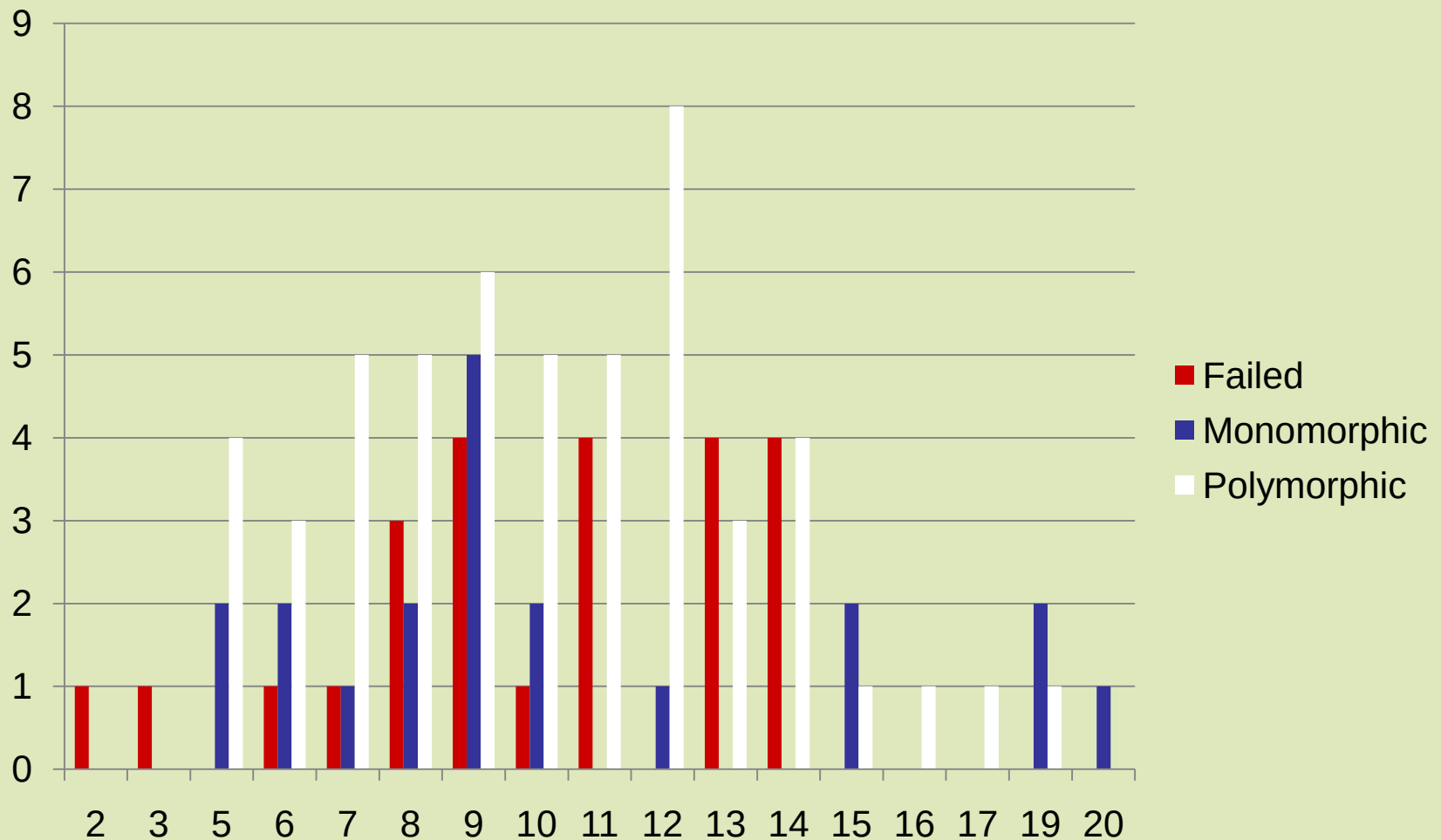
Performance of Accession-Specific SNPs Peach



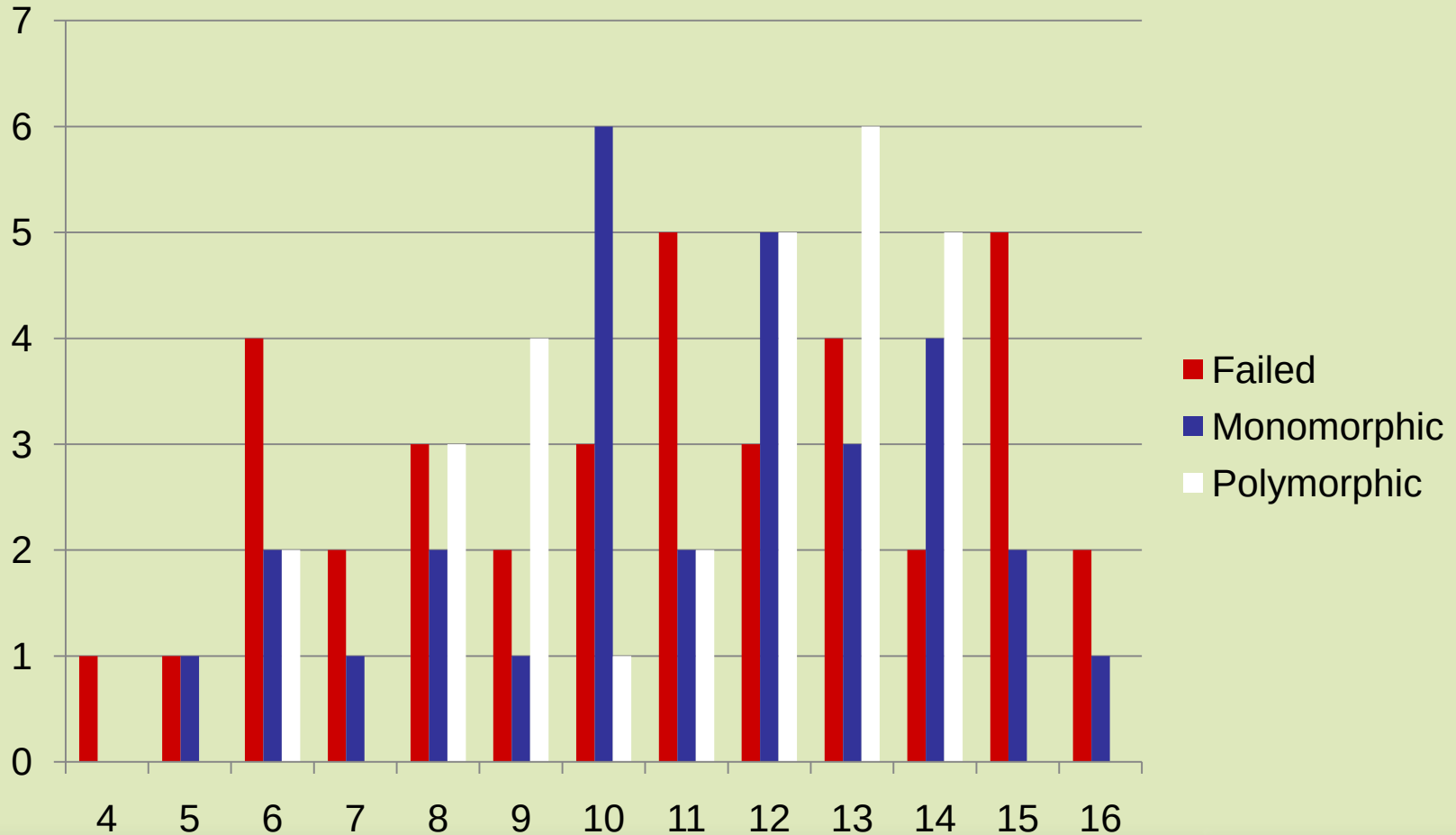
Performance of Accession-Specific SNPs- Sweet Cherry



Effect of Number of Peach Accessions Containing SNP data



Effect of Number of Sweet Cherry Accessions Containing SNP data





Available Rosaceae Genomic Resources

- ❖ SNP detection pipeline-to upload on GDR
- ❖ Short sequence reads and SNPs for apple, peach, cherry and strawberry-GenBank
- ❖ Validated apple, peach and cherry SNPs
- ❖ Infinium SNP chip for apple (9k), peach (9k) and cherry (6k)

Acknowledgements



United States Department of Agriculture
National Institute of Food and Agriculture

This project is supported by the
Specialty Crops Research Initiative of USDA's
National Institute of Food and Agriculture

Questions?

