RosBREED

Enabling marker-assisted breeding in Rosaceae

Development and Preliminary Evaluation of the IStraw90 Axiom® Array in The Cultivated Strawberry (*Fragaria ×ananassa*)

Nahla Bassil, USDA-ARS,NCGR

PAG XXII

International Strawberry SNP Consortium

USA

- NCGR, Corvallis
  Nahla Bassil
- Univ. New Hampshire
  Tom Davis, Lise Mahoney, David Wood, Yilong Yang, Hailong Zhang
- Washington State Univ.
  Stephen Ficklin, Sook Jung, Dorrie Main, Cameron Peace
- Michigan State Univ.
  Amy Iezzoni, Umesh Rosyara
- Univ of Florida
  Vance Whitaker
- Affymetrix-USA
  Mike Mittmann, Ali Pirani, **Teresa Webster**, LuAnn Glaser, Lu Yan....

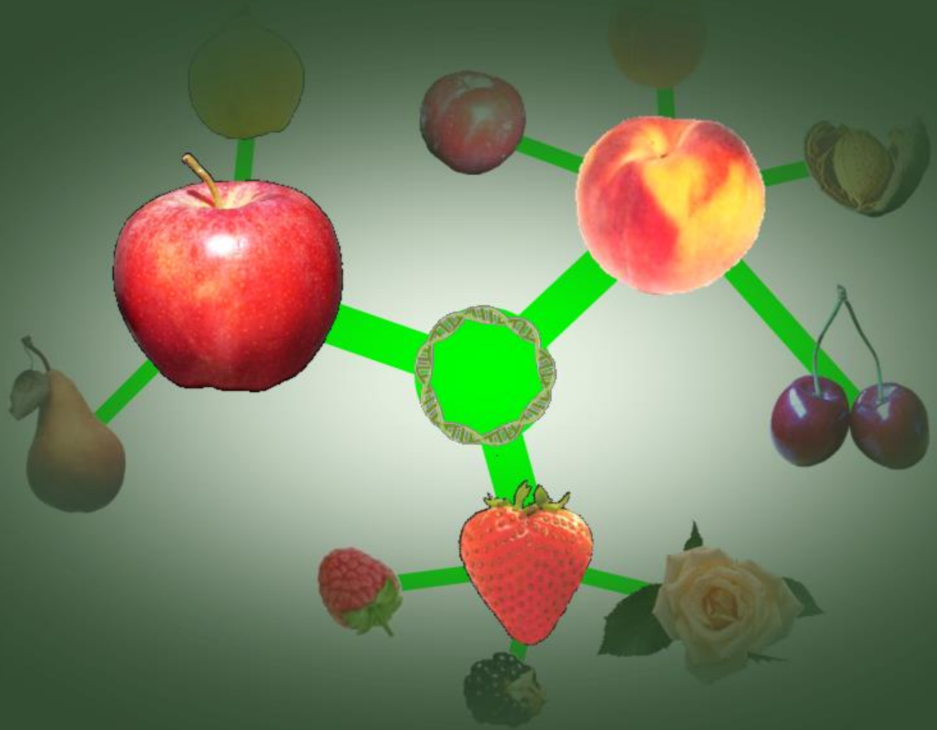
Europe

- Wageningen UR, NL
  Eric van de Weg, Thijs van Dijk
- IFAPA- Spain
  Iraida Amaya
- INRA-Bordeaux, France
  Beatrice Denoyes
- IRTA-Barcelona, Spain
  Amparo Monfort
- IASMA-FEM, Italy
  Dan Sargent
- Affymetrix-UK
  Fiona Brew

Present at PAG

Outline of Presentation

- Introduction
- IStraw90 description
- Preliminary evaluation
- Conclusions



Introduction

RosBREED

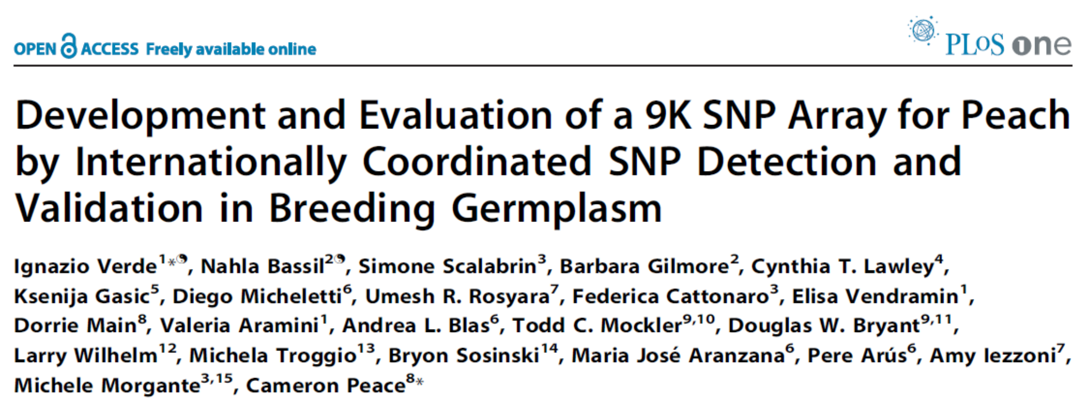
- RosBREED aims to accelerate and increase the efficiency of rosaceous cultivar release through marker-assisted breeding

- High throughput genotyping tools for QTL discovery: **SNP arrays** (through international collaborations)

- Now finding practical application in breeding programs: e.g., QTLs for resistance to bacterial spot in peach (Yang et al., 2013)

Illumina Infinium® Arrays

***Verde et al.* 2012**



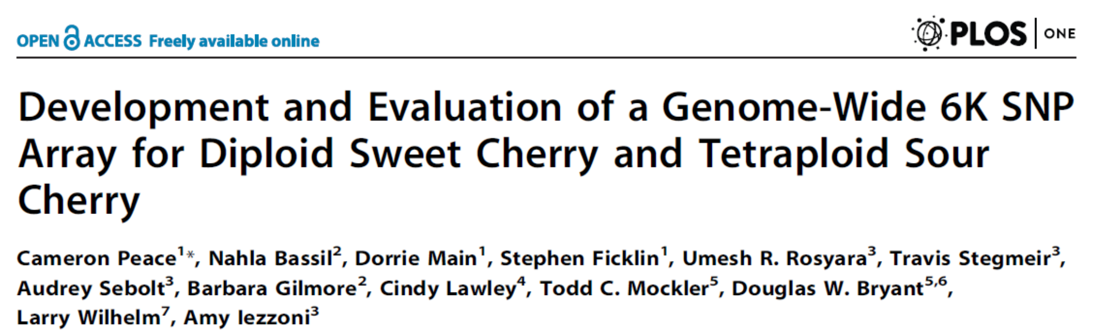
OPEN ACCESS Freely available online

PLoS one

Development and Evaluation of a 9K SNP Array for Peach by Internationally Coordinated SNP Detection and Validation in Breeding Germplasm

Ignazio Verde[1*¶], Nahla Bassil[2¶], Simone Scalabrin[3], Barbara Gilmore[2], Cynthia T. Lawley[4], Ksenija Gasic[5], Diego Micheletti[6], Umesh R. Rosyara[7], Federica Cattonaro[3], Elisa Vendramin[1], Dorrie Main[8], Valeria Aramini[1], Andrea L. Blas[6], Todd C. Mockler[9,10], Douglas W. Bryant[9,11], Larry Wilhelm[12], Michela Troggio[13], Bryon Sosinski[14], Maria José Aranzana[6], Pere Arús[6], Amy Iezzoni[7], Michele Morgante[3,15], Cameron Peace[8*]

***Peace et al.* 2012**



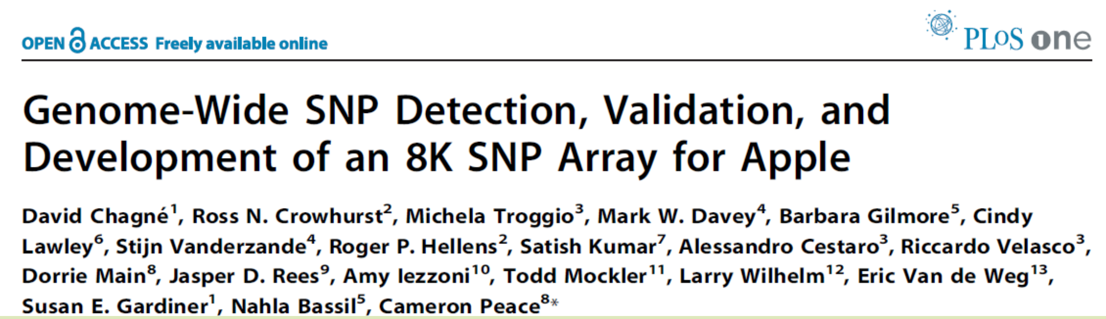
OPEN ACCESS Freely available online

PLOS | ONE

Development and Evaluation of a Genome-Wide 6K SNP Array for Diploid Sweet Cherry and Tetraploid Sour Cherry

Cameron Peace[1*], Nahla Bassil[2], Dorrie Main[1], Stephen Ficklin[1], Umesh R. Rosyara[3], Travis Stegmeir[3], Audrey Sebolt[3], Barbara Gilmore[2], Cindy Lawley[4], Todd C. Mockler[5], Douglas W. Bryant[5,6], Larry Wilhelm[7], Amy Iezzoni[3]

***Chagné et al.* 2012**



OPEN ACCESS Freely available online

PLoS one

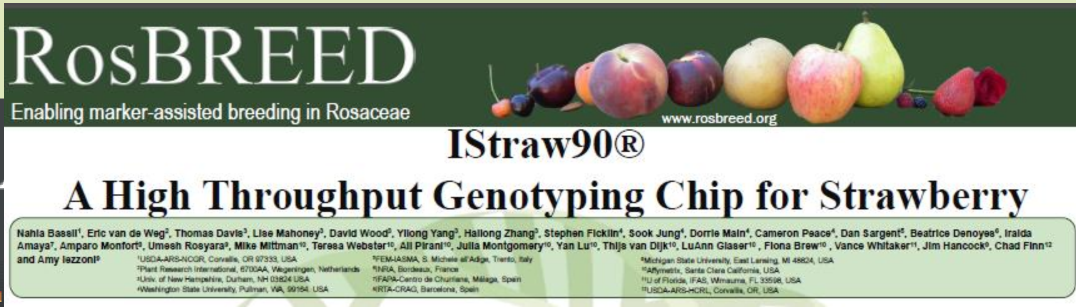
Genome-Wide SNP Detection, Validation, and Development of an 8K SNP Array for Apple

David Chagné[1], Ross N. Crowhurst[2], Michela Troggio[3], Mark W. Davey[4], Barbara Gilmore[5], Cindy Lawley[6], Stijn Vanderzande[4], Roger P. Hellens[2], Satish Kumar[7], Alessandro Cestaro[3], Riccardo Velasco[3], Dorrie Main[8], Jasper D. Rees[9], Amy Iezzoni[10], Todd Mockler[11], Larry Wilhelm[12], Eric Van de Weg[13], Susan E. Gardiner[1], Nahla Bassil[5], Cameron Peace[8*]

Affymetrix Axiom® Strawberry Array

The IStraw90® Array

Released Oct 22, 2013



RosBREED

Enabling marker-assisted breeding in Rosaceae

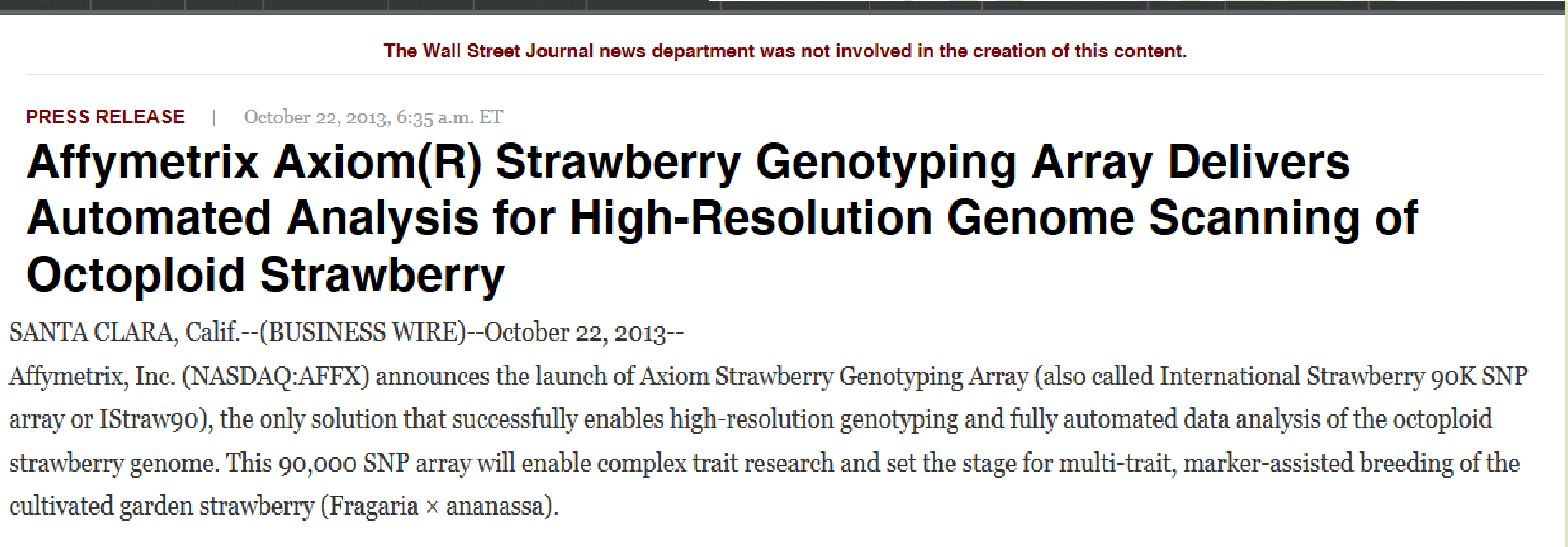
www.rosbreed.org

IStraw90®

A High Throughput Genotyping Chip for Strawberry

Nahla Bassil[1], Eric van de Weg[2], Thomas Davis[3], Lise Mahoney[3], David Wood[3], Yilong Yang[3], Hailong Zhang[3], Stephen Ficklin[4], Sook Jung[4], Dorrie Main[4], Cameron Peace[4], Dan Sargent[5], Beatrice Denoyes[6], Iraida Amaya[7], Amparo Monfort[8], Umesh Rosyara[9], Mike Mittman[10], Teresa Webster[10], Ali Pirani[10], Julia Montgomery[10], Yan Lu[10], Thijs van Dijk[10], LuAnn Glaser[10], Fiona Brew[10], Vance Whitaker[11], Jim Hancock[9], Chad Finn[12] and Amy Iezzoni[9]

[1]USDA-ARS-NCGR, Corvallis, OR 97333, USA
[2]Plant Research International, 6700AA, Wageningen, Netherlands
[3]Univ. of New Hampshire, Durham, NH 03824 USA
[4]Washington State University, Pullman, WA, 99164 USA
[5]FEM-IASMA, S. Michele all'Adige, Trento, Italy
[6]INRA, Bordeaux, France
[7]IFAPA-Centro de Churriana, Málaga, Spain
[8]IRTA-CRAG, Barcelona, Spain
[9]Michigan State University, East Lansing, MI 48824, USA
[10]Affymetrix, Santa Clara California, USA
[11]U of Florida, IFAS, Wimauma, FL 33598, USA
[12]USDA-ARS-HCRL, Corvallis, OR, USA



THE WALL STREET JOURNAL

U.S. EDITION | Tuesday, October 22, 2013 As of 6:35 AM EDT

The Wall Street Journal news department was not involved in the creation of this content.

PRESS RELEASE | October 22, 2013, 6:35 a.m. ET

Affymetrix Axiom(R) Strawberry Genotyping Array Delivers Automated Analysis for High-Resolution Genome Scanning of Octoploid Strawberry

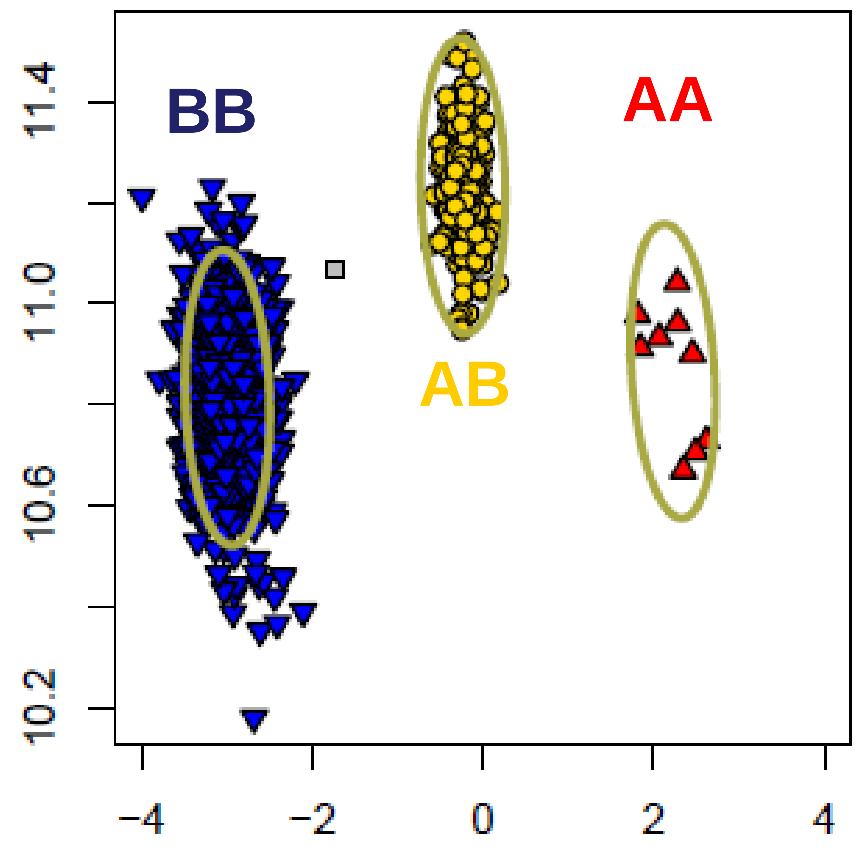
SANTA CLARA, Calif.--(BUSINESS WIRE)--October 22, 2013--

Affymetrix, Inc. (NASDAQ:AFFX) announces the launch of Axiom Strawberry Genotyping Array (also called International Strawberry 90K SNP array or IStraw90), the only solution that successfully enables high-resolution genotyping and fully automated data analysis of the octoploid strawberry genome. This 90,000 SNP array will enable complex trait research and set the stage for multi-trait, marker-assisted breeding of the cultivated garden strawberry (Fragaria × ananassa).

AxiomGT1 SNP Cluster Plot

Size = Signal intensity:

[log2(A)+log2(B)]/2



AxiomGT1 Genotype Calls
Blue = BB
Gold = AB
Red = AA
Grey = NoCall

⬯ = Location of genotype cluster determined by AxiomGT algorithm

Contrast = Ratio between the two label-colors: log2(A/B)

Genotype Notation: Genotypes of subgenomes are separated by dashes

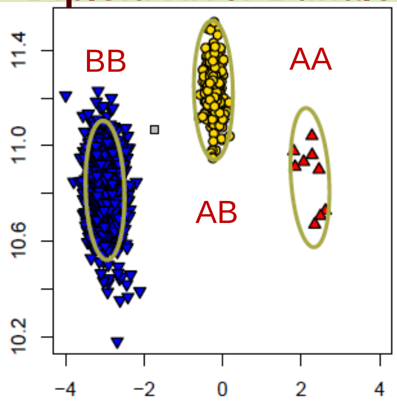
- Subgenome with the segregating allele is first (red),
- Subgenomes with fixed alleles follow (black)

Approaches for SNP Selection

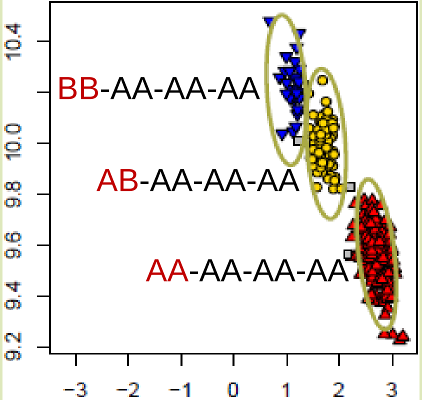
- ❖ Focus on **high coverage (≥16x)** and removal of duplicate reads- **9 SNP Detection Panel**

- ❖ **Reduced ploidy markers** to address challenges of cluster compression (8x -> 2x-6x)

Diploid River Buffalo



Octoploid Strawberry



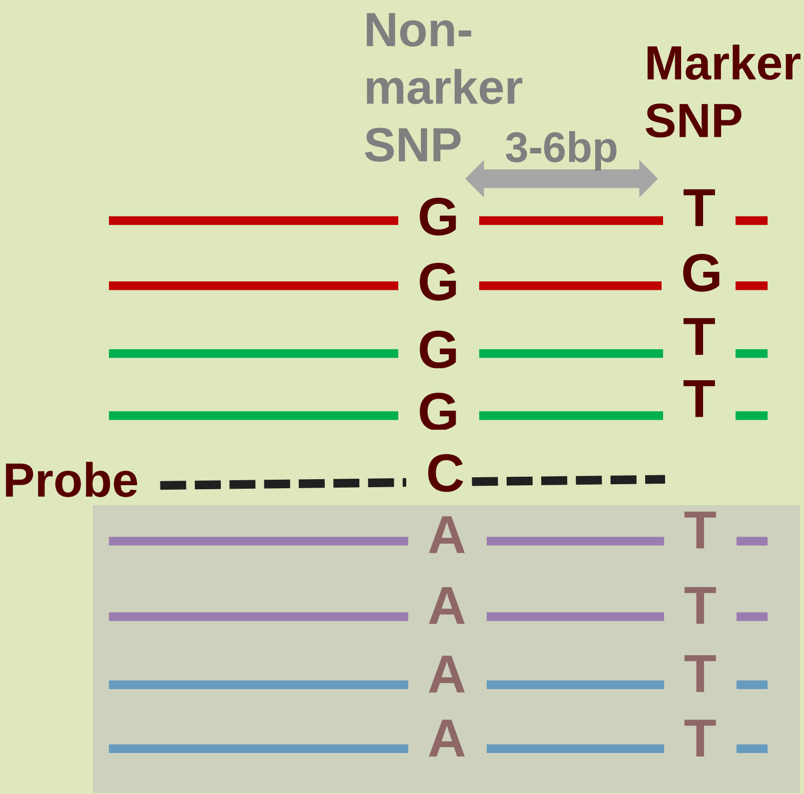
- ❖ Extensive use of **bioinformatic** tools and in-house scripts for SNP filtering

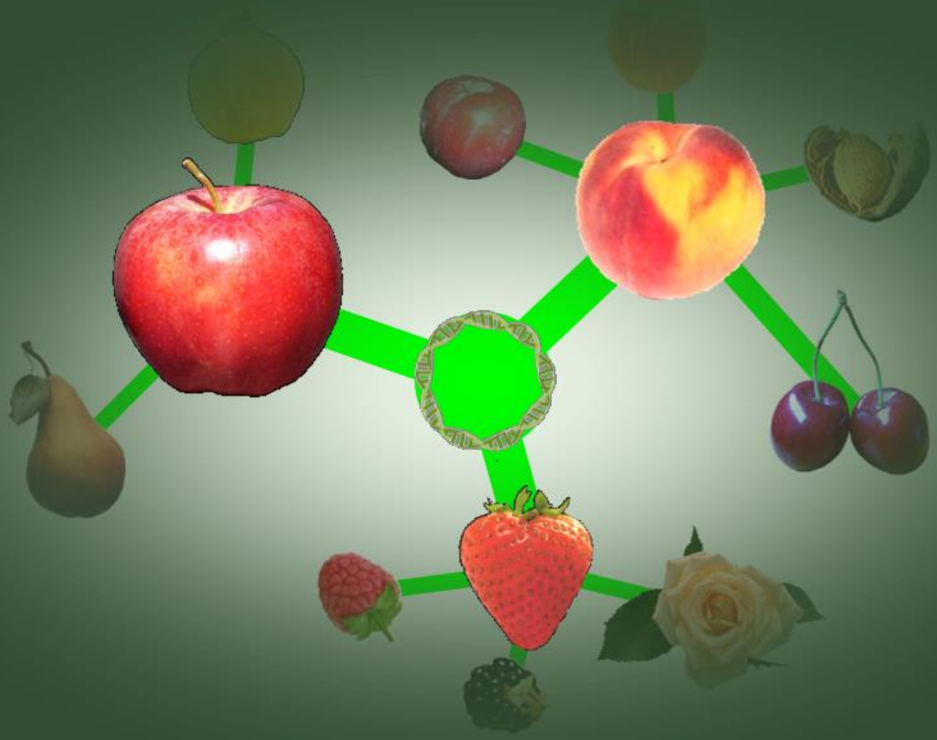
- ❖ **Large number of SNPs: > 90K** (permits experimentation and reduced conversion rate)

IStraw90® Axiom Array

Strawberry SNP Classes	# SNPs	# Probes
Classical SNPs	**63,263**	**86,817**
Reduced ploidy	**22,401**	**41,282**
del	4,913	5,361
ins	4,615	5,197
SNP in Del	836	932
SNP in Ins	2,007	2,444
Indel-SNP	1,177	1,206
mSNP	1,761	19,050
SNP-SNP	7,092	7,092
Diploid (*F. iinumae*) SNPs		
F. iinumae SNPs	3,751	4,000
Sequence-independent	5,648	6,000
Total	**95,063**	**138,099**

SNP-SNP Markers





Preliminary Evaluation

Strawberry Individuals: 384

Strawberry Samples	No.	Classification
1. Diploid	**27**	
F. iinumae	24	Mapping population
F. vesca	2	Available sequence data
F. mandschurica	1	Available sequence data
2. Octoploid		
F. ×ananassa	**306**	
	176 (4)	Mapping populations
	2 (4,2)	Technical replicates
	9	SNP detection panel
	49	Pedigree-linked populations-MSU-ORUS
	66	Other
NON *F. ×ananassa*	**51**	
	38	Pedigree-linked populations-NH
	2	*F. chiloensis* sequenced
	2	*F. virginiana* sequenced
	9	Reconstituted *F. ×ananassa* (FVC)

First Round of Axiom Genotypic Analysis Workflow

Step 1. Group samples into diploid vs octoploid

Octoploid

Diploid

Step 2. Genotype each group with the Affymetrix Power Tool or Genotyping Console

Genotype Calls Table

Genotype Calls Table

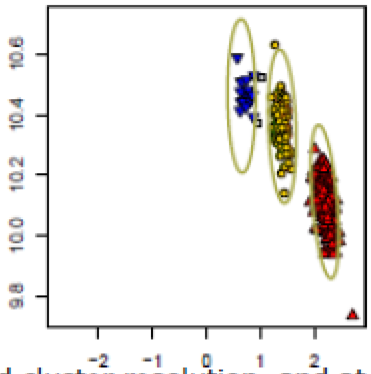
Step 3. Sort SNPs into six quality categories with *SNPolisher* software

Next slide

(PAG Poster #437)
)

Six SNP Categories

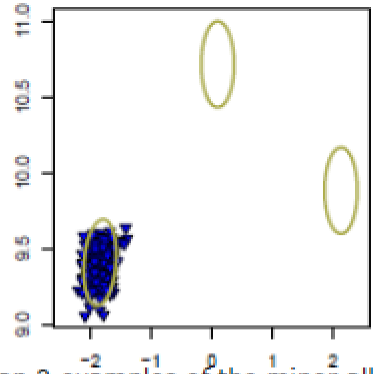
PolyHighResolution



Good cluster resolution, and at least 2 example of the minor allele.

23.4%

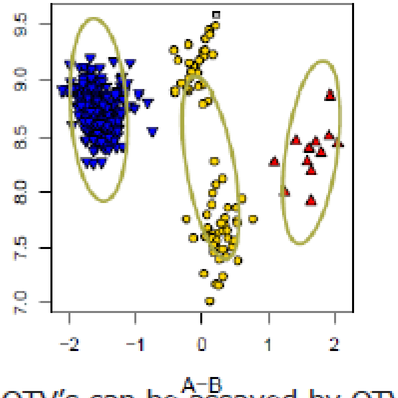
MonoHighResolution



Less than 2 examples of the minor allele usually due to low MAF samples, but possible cluster fusion / compressio1

19.0%

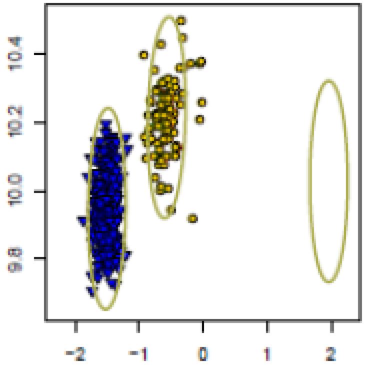
OTV or "Null" allele



OTV's can be assayed by OTV genotyping (see appendix)

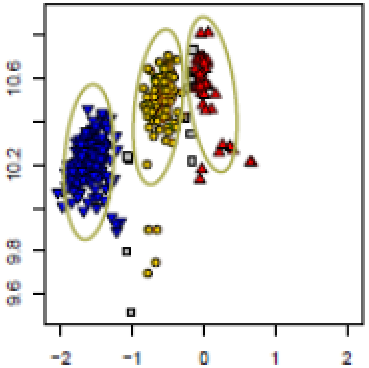
1.1%

No Minor Hom **37.1%**



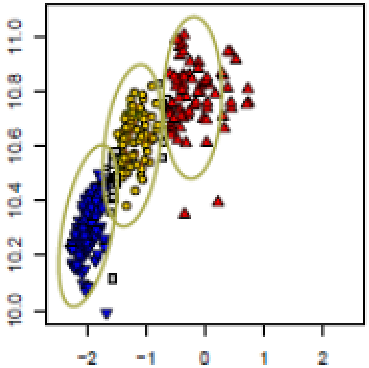
Two clusters with no examples of the minor homozygous genotypes. Het cluster not in OTV location

CallRate Below Thresh. **6.5%**



SNP callrate is below threshold, but other cluster properties are above threshold.

Other **12.7%**

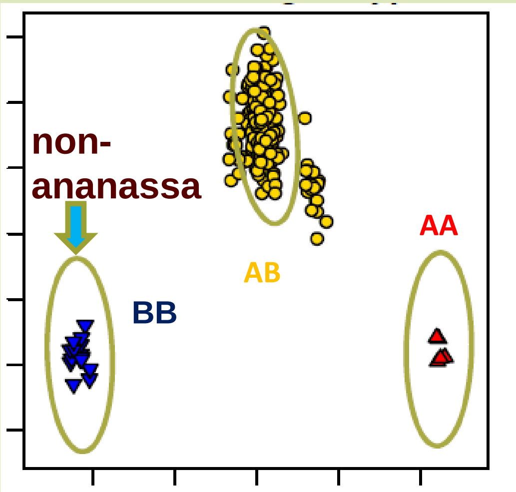


One or more cluster properties are below threshold.
Expect lower quality genotypes

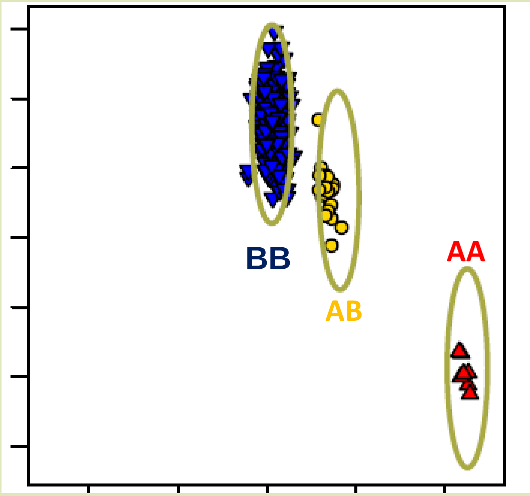
Steps Adopted to Identify 3 Well-Resolved Genotype Clusters

1. Genotype *F. ×ananassa* separately from other octoploid species

F. ×ananassa + wild progenitors



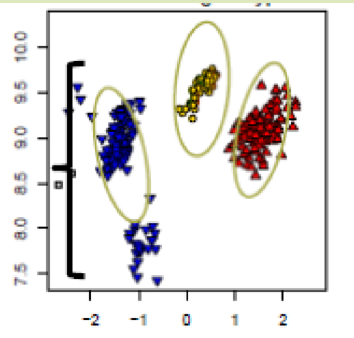
F. ×ananassa only



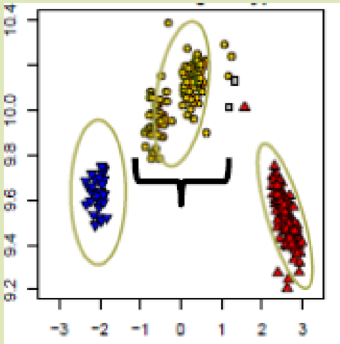
Steps Adopted to Identify 3 Well-Resolved Genotype Clusters

2. Further sub-division of the PolyHighResolution Class
    - ➢ Three additional filtering tools modified for strawberry data

A. **≥ 3** examples of the **minor homozygote genotypes**
B. **100% reproducibility** among technical replicates
C. Not more than 3 genotype clusters
    - ➢ Select through use of “variance” filters



Excessive Variance of the BB cluster

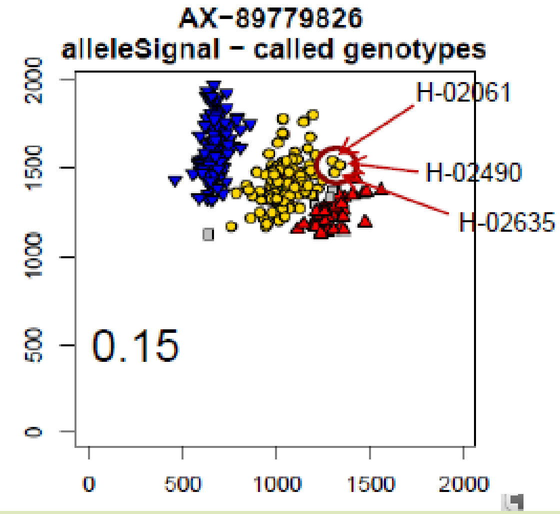


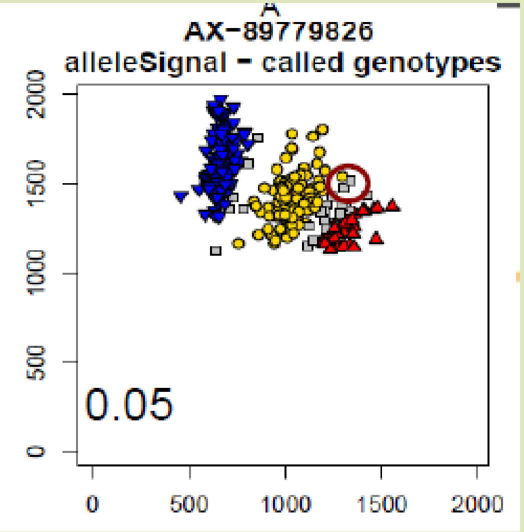
Excessive Variance of the AB cluster

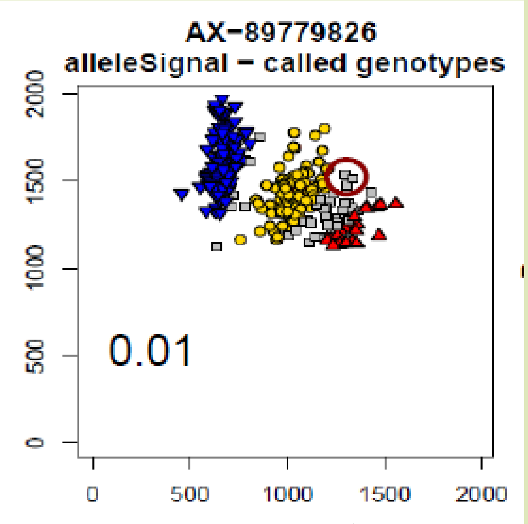
Steps Adopted to Identify 3 Well-Resolved Genotype Clusters

3. No calls for samples that cluster on margins
   - ➢ Increasing stringency for making genotype calls

Blue = BB **Red = AA**
Gold = AB **Grey = NoCall**







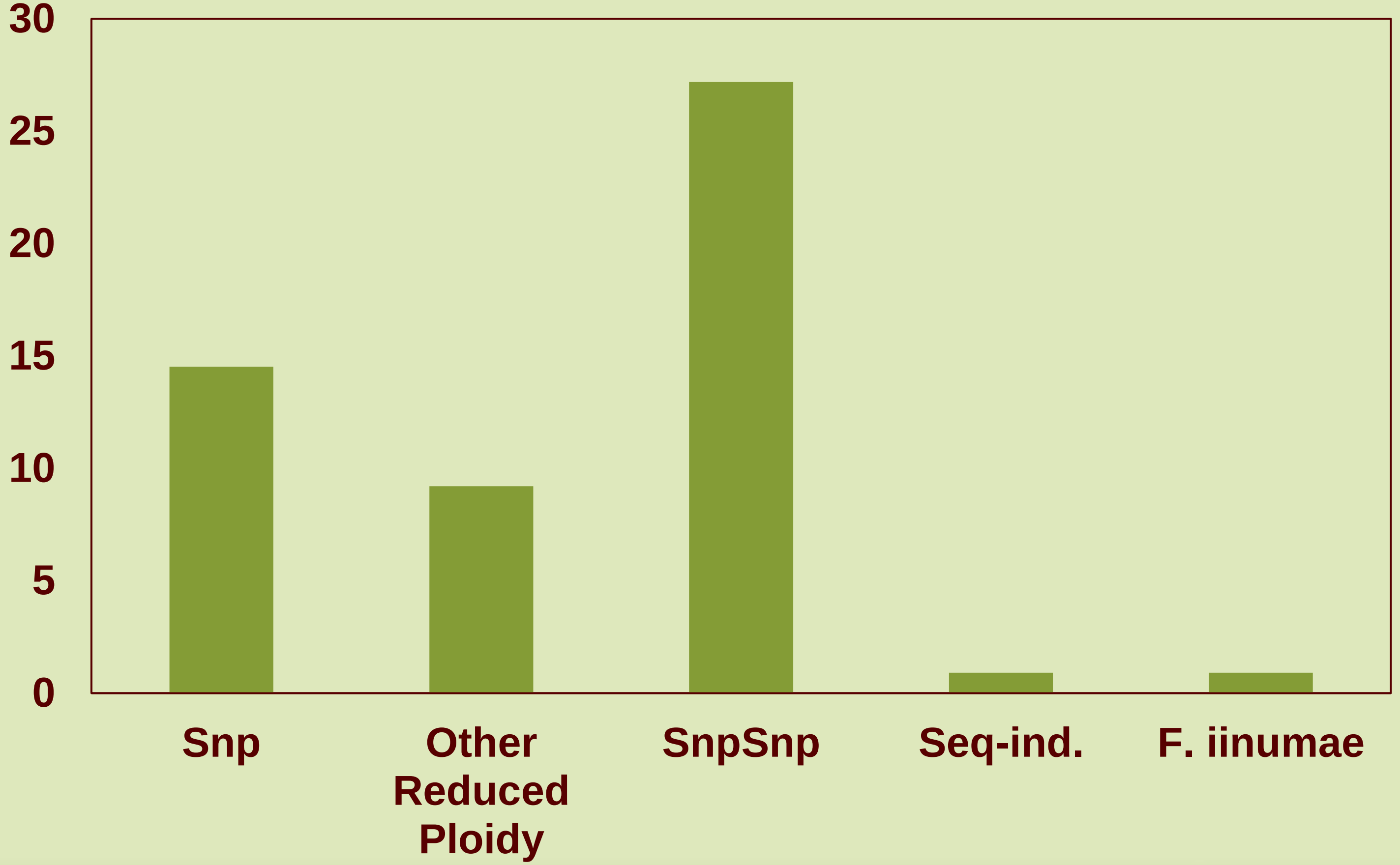
Effect of Increasing Call Stringency by decreasing the Confidence Score

Genotyping 306 *F. ×ananassa* Samples

284 samples pass threshold: 97% Genotyping Call (GC)

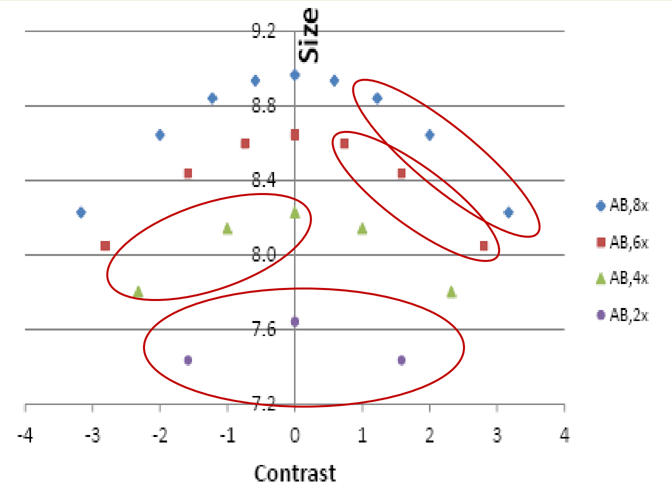
	SNP Class					
	Snp	**Other Reduced Ploidy**	**SnpSnp**	***F. iinumae***	**Seq-ind.**	**Total**
Total # SNPs	63,263	15,309	7,091	3,751	5,648	**95,062**
PolyHighRes-Start	16,891	2,945	3,101	157	261	**23,355**
Additional Filters						
- nMinorHom	1,153	311	53	33	61	1,611
- Reproducibility	2,568	385	182	55	73	3,263
- Variance	3,984	839	938	35	76	5,872
PolyHighRes-Filtered	9,186	1,410	1,928	34	51	**12,609**

Snp-Snp Markers Have the Highest Percent of Conversion [PolyHighResolution (PHR) Class]



Axiom® Genotyping at different levels of allo-ploidy

Simulated Cluster Modeling

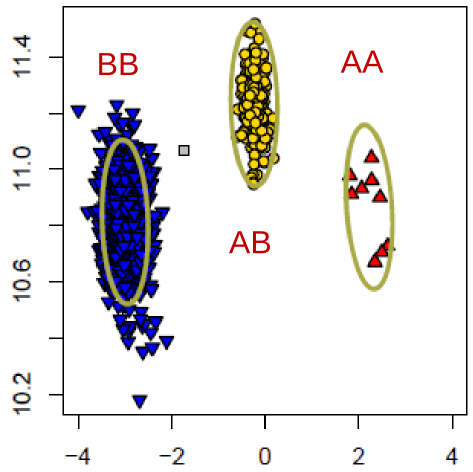


Affymetrix developed procedures to easily identify SNPs that perform at the diploid level

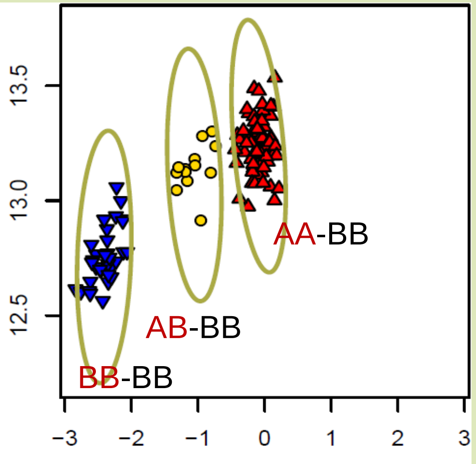
- Homozygous Ratio Offset (HomRO)

41% of PHR SNPs have diploid clusters

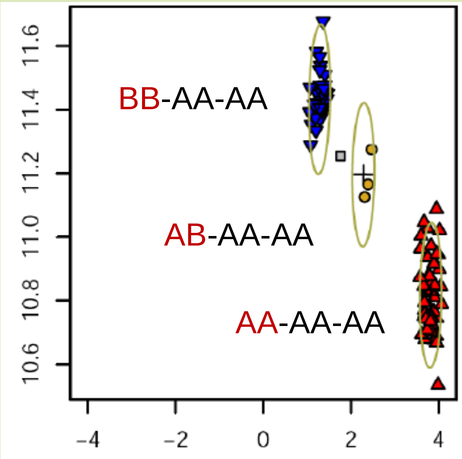
Diploid River Buffalo



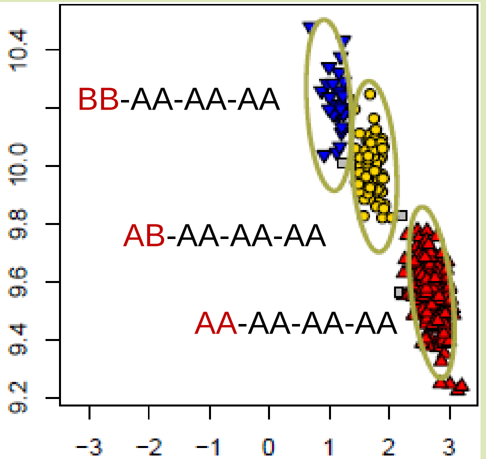
Allotetraploid Rapeseed



Allohexaploid Wheat

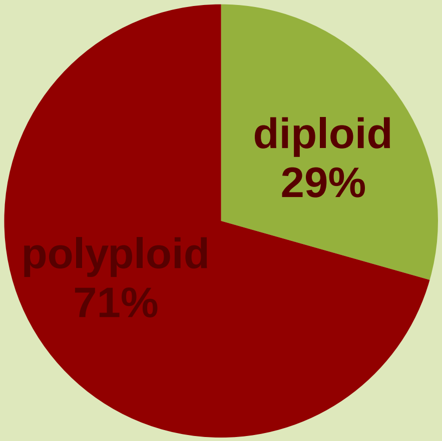


Octoploid Strawberry

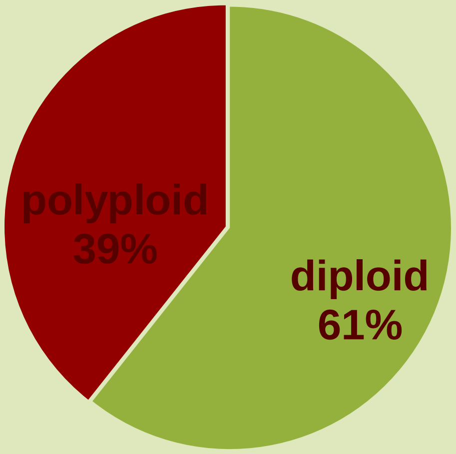


SNP Design & Effective Ploidy Level

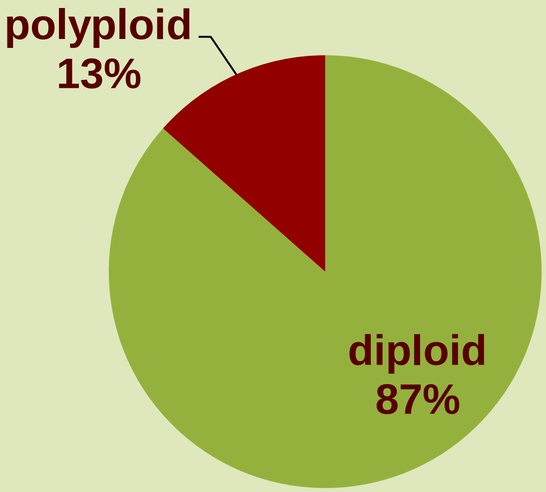
Snp



Other Reduced Ploidy



SnpSnp



- Classical SNP: 29% at diploid level
- Reduced ploidy strategies: effective!
  - Majority has diploid-like clusters

Validation of Genotyping Accuracy

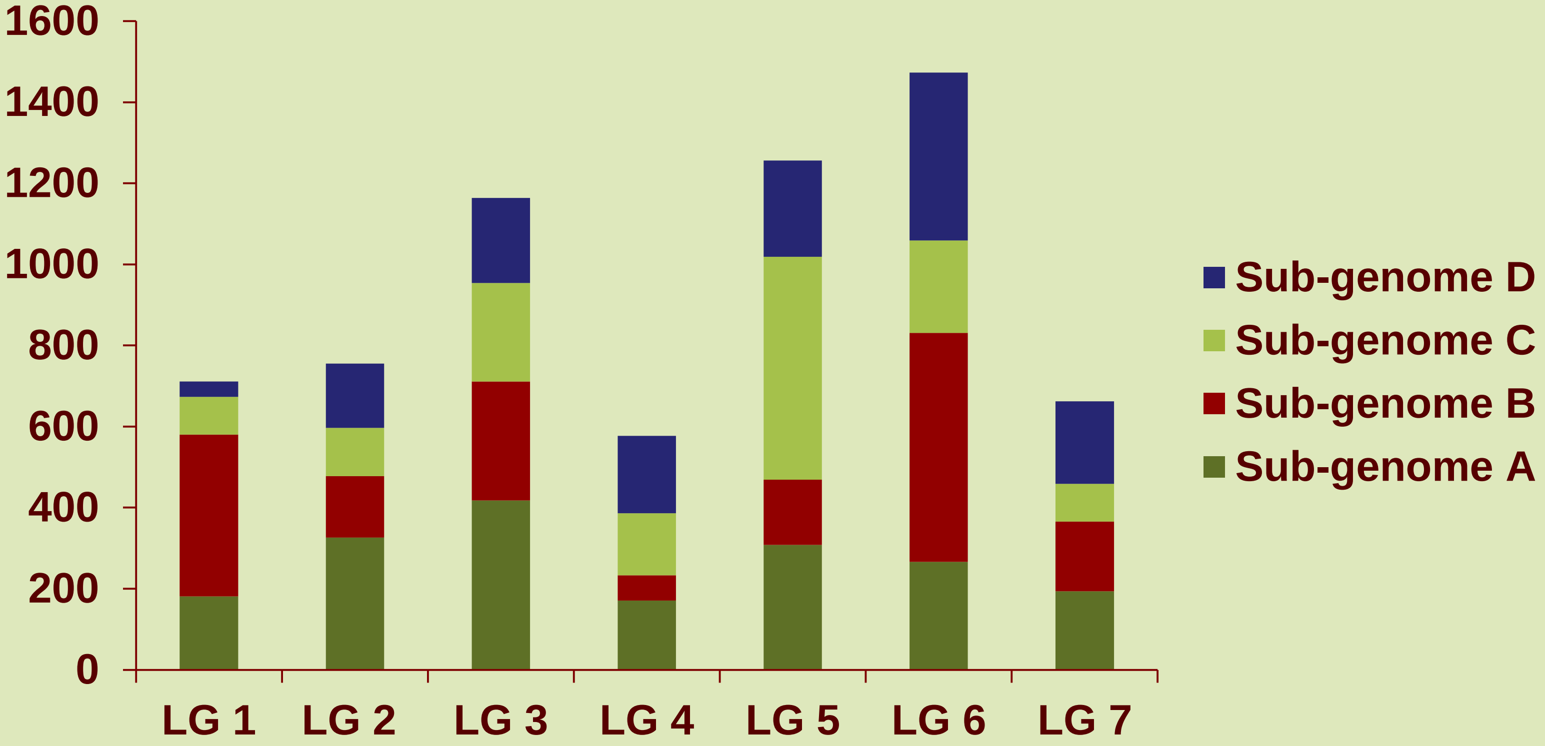
Using **inheritance error** and **mapping performance** in four mapping populations

❖ Holiday x Korona	HxK	**n = 79**
❖ Tribute x Honeoye	TxH	n = 22 P1/♂ lacking
❖ Redgauntlet x Hapil	RxH	n = 38 P1/♀ lacking
❖ Capitola x CF1116	CxCF	n = 20

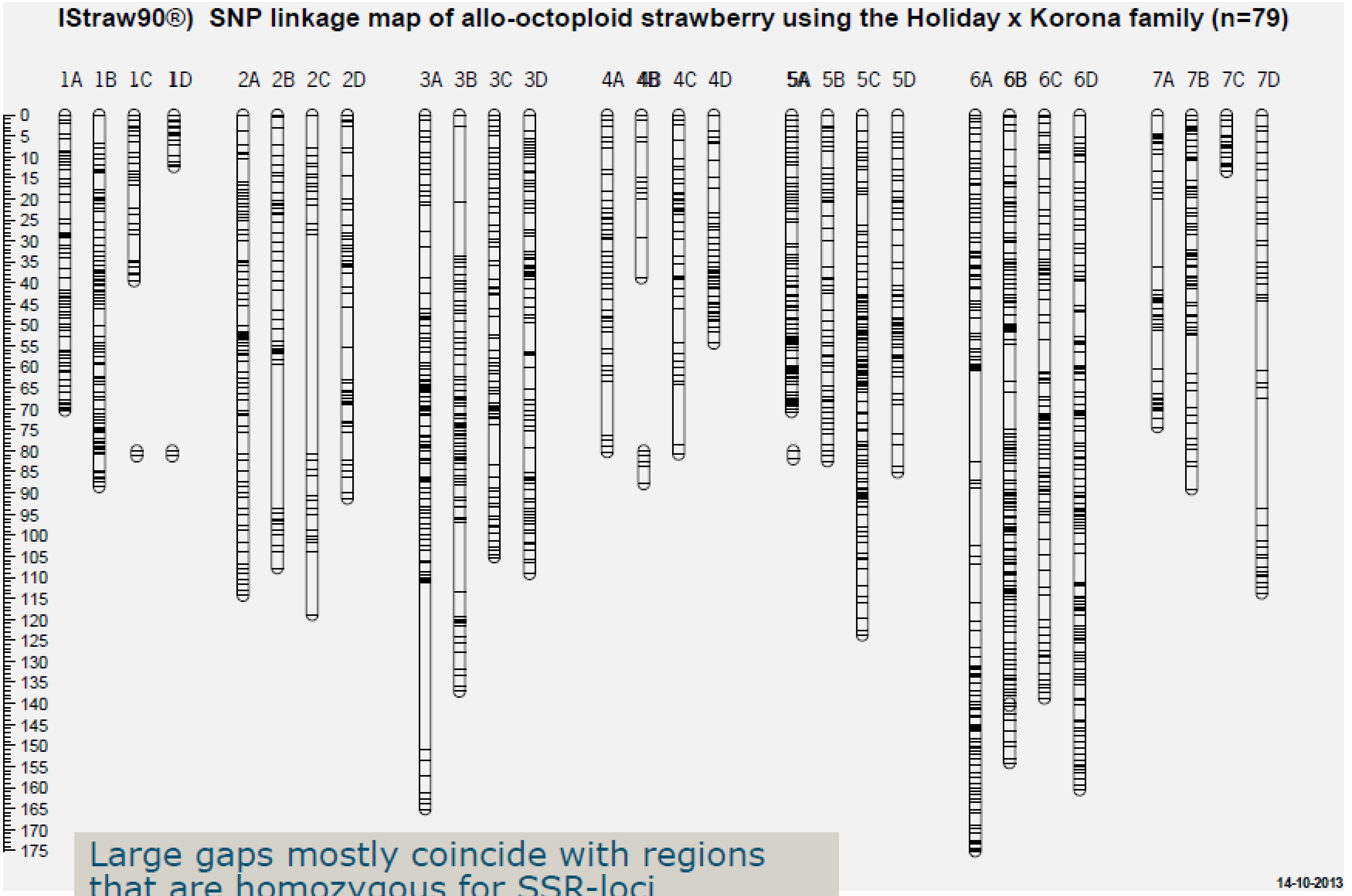
Validation in ‘Holiday’ x ‘Korona’

Start: **8,084 polymorphic markers**

Mapped: **6,589** (=81%) **were mapped giving all the expected 28 LGs**



Validation in ‘Holiday’ x ‘Korona’



IStraw90®) SNP linkage map of allo-octoploid strawberry using the Holiday x Korona family (n=79)

Large gaps mostly coincide with regions that are homozygous for SSR-loci

14-10-2013

WAGENINGEN UR
For quality of life

affymetrix
Biology for a better world

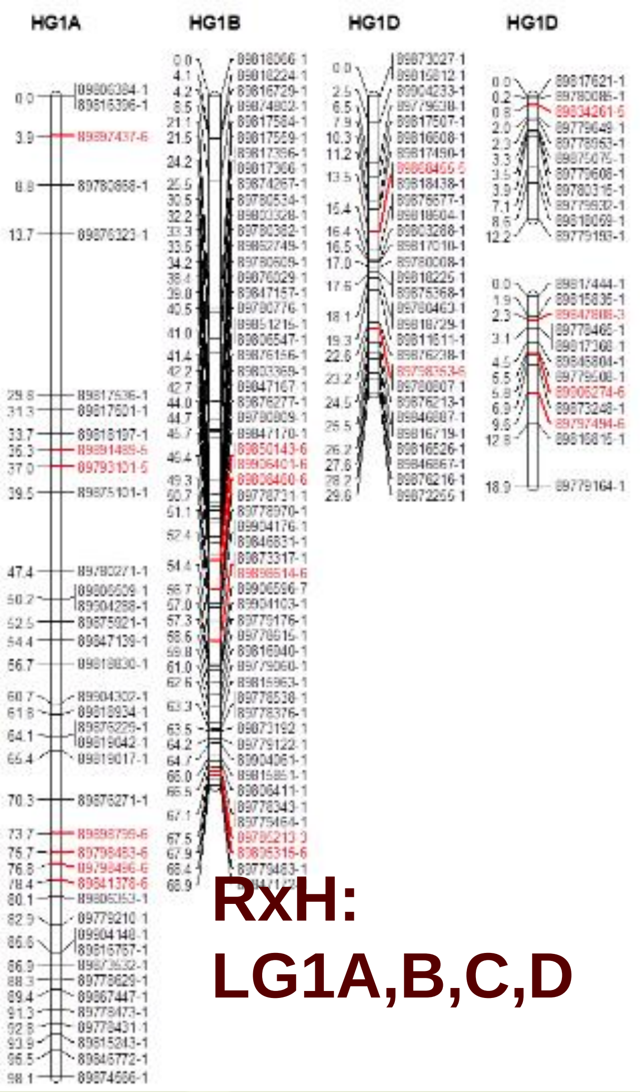
RosBREED
Enabling marker-assisted breeding in Rosaceae

USDA
NIFA

Validation in Smaller Mapping Populations

- ❖ RxH: mapped **1769** SNPs
  - ➢underestimation due to absence of genotype data for ♀ parent
- ❖ TxH: mapped **2852** SNPs
  - ➢underestimation due to absence of genotype data for ♂ parent
- ❖ CxCF: In progress

- ❖ Next:
  - ➢ Include the lacking parents
  - ➢ Scrutinize genotype calls of individual samples
  - ➢ Increase family sizes



RxH: LG1A,B,C,D

Mapping of Diploid SNPs

Interspecific hybrid, F_1, and F_2 population

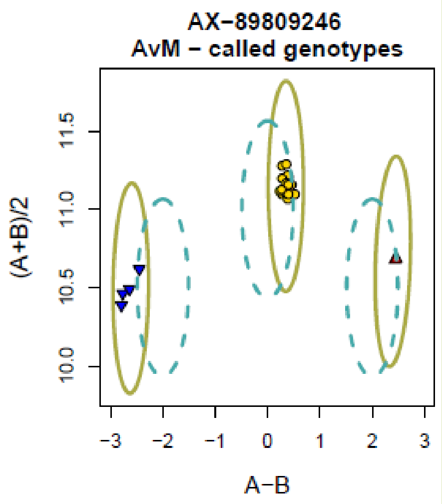
F1D Illumina sequence used for SNP Discovery

J17 × J4

F1D (self)

F2D (n=150)

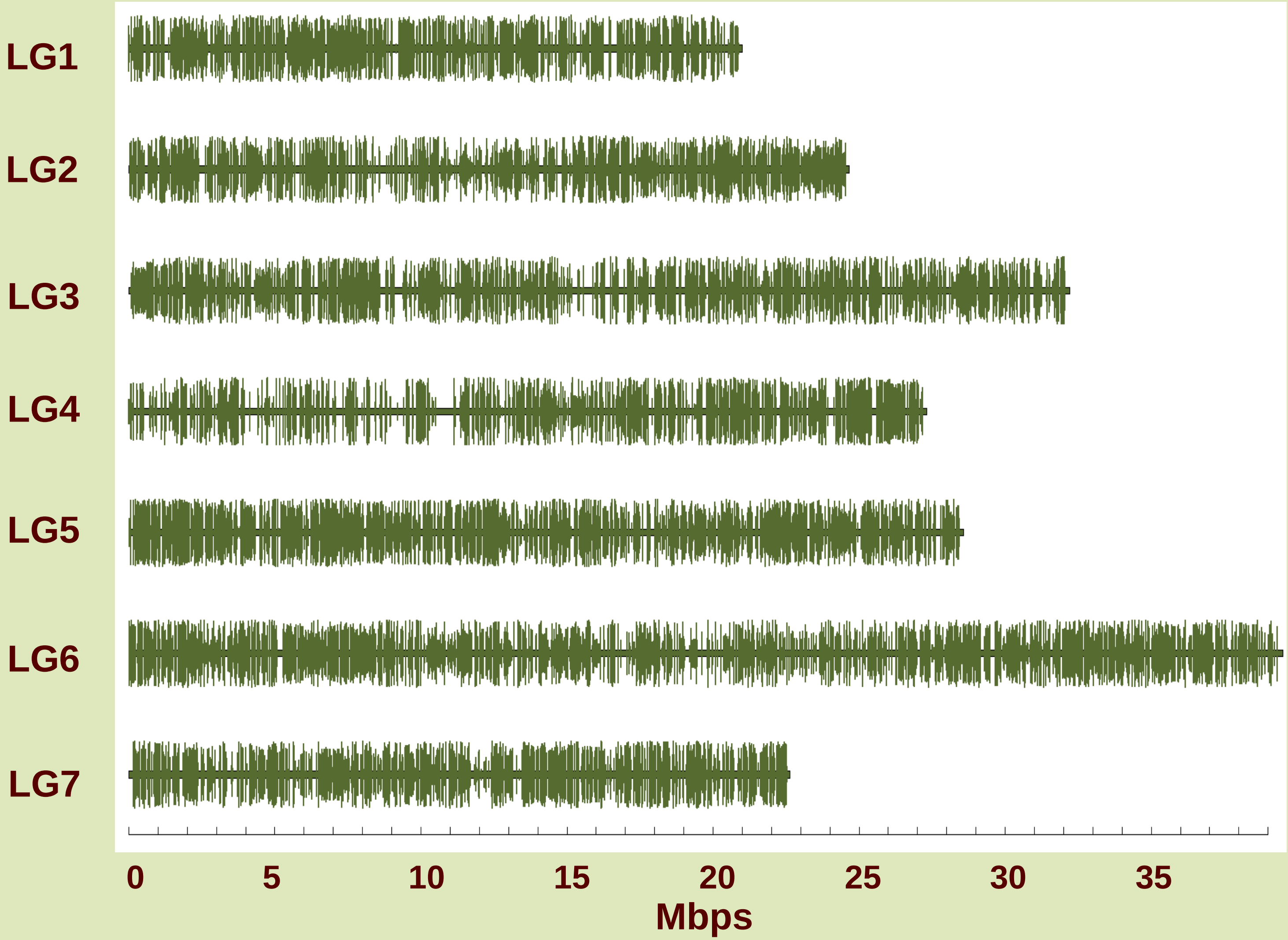
4000 SNPs	IStraw 90
J4	1
J17	1
F1D	1
F2Ds	21



***2928 F1D SNPs successfully map** (PAG Poster #437)

- ❖ 188 SNPs designed for the octoploids map in *F. iinumae*
- ❖ 34 SNPs designed for *F. iinumae* are in the filtered PolyHighRes. category in the octoploids

SNP Distribution and Polymorphism in Strawberry Germplasm (52)



Distribution of SNPs among 7 LGs as illustrated by minor allele frequency according to physical location on the *F. vesca* Hawaii v. 1 reference genome in 52 diverse strawberry accessions

Conclusions

- Commercial availability of the first array for an octoploid organism, IStraw90, provides new opportunities for genetic research in polyploids
- Affymetrix genotyping tools further empowered
- Over 6,000 SNPs mapped in ‘Holiday’ x ‘Korona’
- SNP design can efficiently reduce effective ploidy level, Snp-snp approach being most effective
- Validation of genotype calls across diverse germplasm is in progress

RosBREED

Enabling marker-assisted breeding in Rosaceae

Portfolio of Impact Statements

January 2014

RosBREED was a multi-state, multi-institution project dedicated to genetic improvement of U.S. rosaceous crops by targeted applications of genomics knowledge and tools to accelerate and increase the efficiency of breeding programs. This Coordinated Agricultural Project was funded through the USDA's National Institute of Food and Agriculture - Specialty Crop Research Initiative by a combination of federal and matching funds, Grant 2009-51181-05808.
www.rosbreed.org

Questions?