E

BMT-TWA/Maize/2/3 Add.
ORIGINAL: English
DATE: December 3, 2007

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
GENEVA

AD HOC CROP SUBGROUP ON MOLECULAR TECHNIQUES FOR MAIZE

Second Session

Chicago, United States of America, December 3, 2007

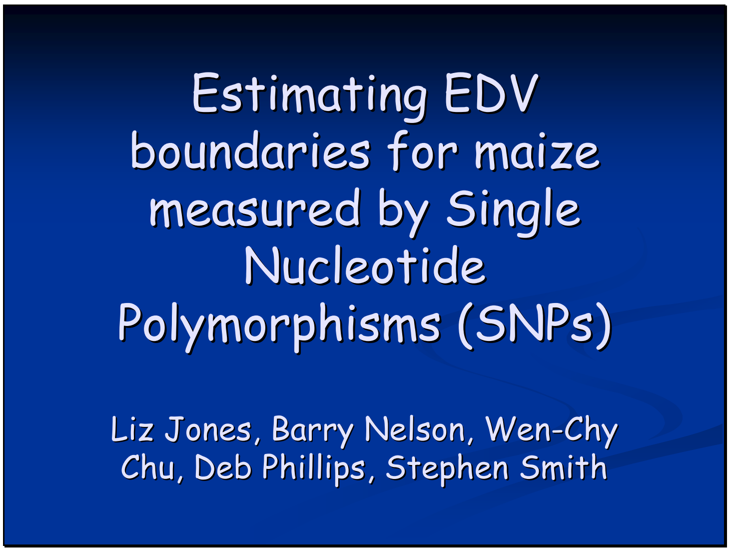
ADDENDUM TO DOCUMENT BMT-TWA/MAIZE/2/3

ANALYSIS OF PVPD INBREDS WITH PUBLIC SNP MARKERS TO ESTIMATE EQUIVALENT EDV THRESHOLDS COMPARED WITH SSR MARKERS

Document prepared by experts from Pioneer Hi-Bred International

This document is an addendum to document BMT-TWA/Maize/2/3 “Analysis of PVPd Inbreds with Public SNP Markers to Estimate Equivalent EDV Thresholds Compared with SSR Markers” and contains a copy of the presentation made by experts from Pioneer Hi-Bred International at the second session of the *Ad Hoc* Crop Subgroup on Molecular Techniques for Maize.

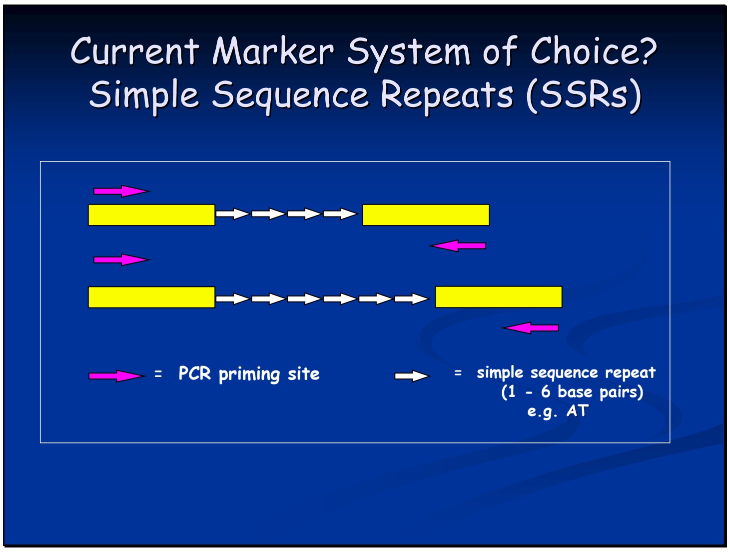
Slide 1



Estimating EDV boundaries for maize measured by Single Nucleotide Polymorphisms (SNPs)

Liz Jones, Barry Nelson, Wen-Chy Chu, Deb Phillips, Stephen Smith

Slide 2

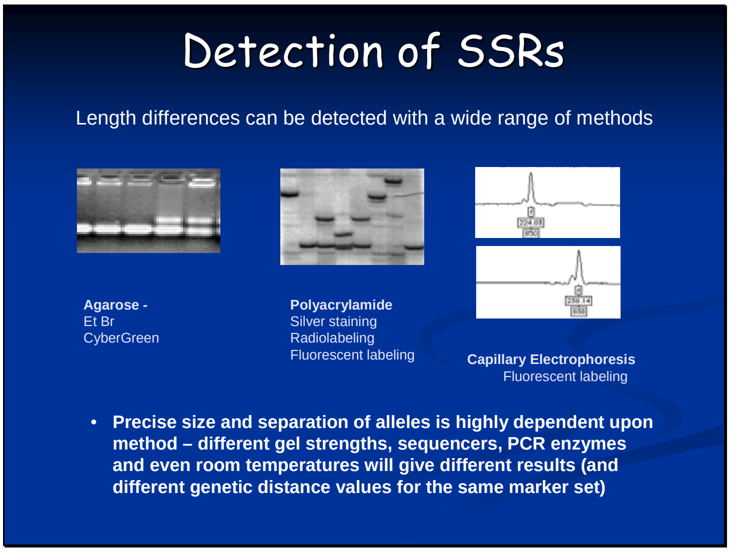


Current Marker System of Choice?
Simple Sequence Repeats (SSRs)

→ = PCR priming site

→ = simple sequence repeat
(1 - 6 base pairs)
e.g. AT

Slide 3



Detection of SSRs

Length differences can be detected with a wide range of methods

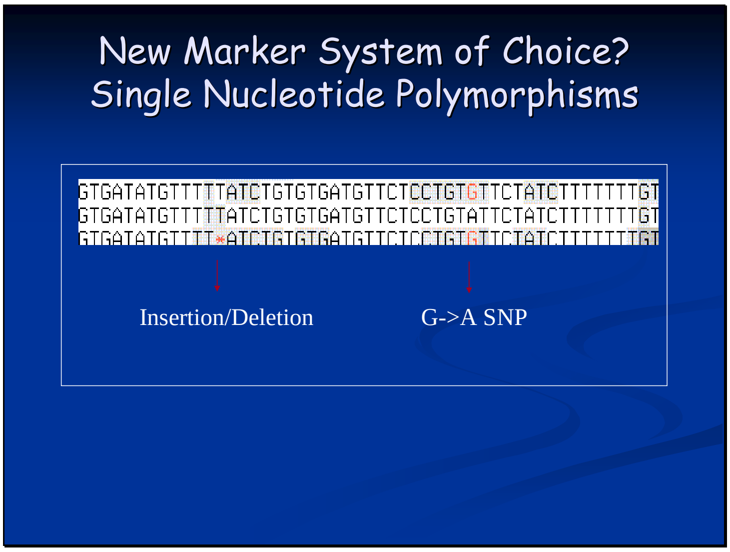
Agarose -
Et Br
CyberGreen

Polyacrylamide
Silver staining
Radiolabeling
Fluorescent labeling

Capillary Electrophoresis
Fluorescent labeling

- **Precise size and separation of alleles is highly dependent upon method – different gel strengths, sequencers, PCR enzymes and even room temperatures will give different results (and different genetic distance values for the same marker set)**

Slide 4



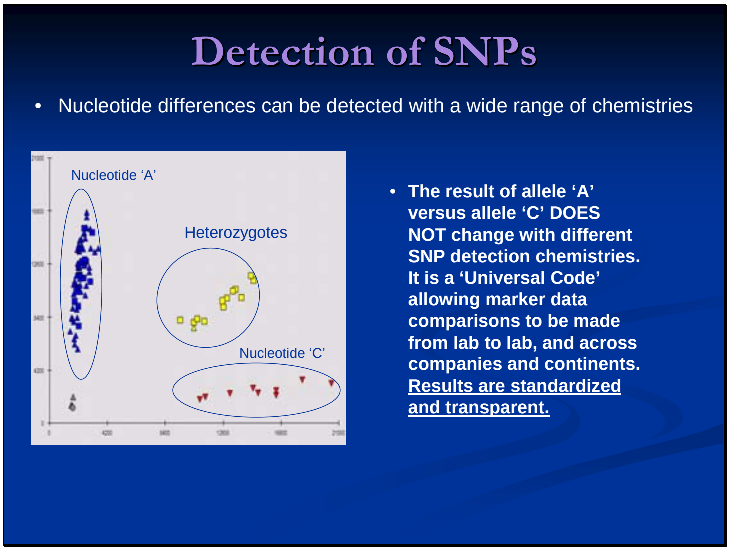
New Marker System of Choice? Single Nucleotide Polymorphisms

```
GTGATATGTTTTTATCTGTGTGATGTTCTCCTGTGTTCTATCTTTTTTTGT
GTGATATGTTTTTATCTGTGTGATGTTCTCCTGTATTCTATCTTTTTTTGT
GTGATATGTTTT*ATCTGTGTGATGTTCTCCTGTGTTCTATCTTTTTTTGT
```

Insertion/Deletion

G->A SNP

Slide 5



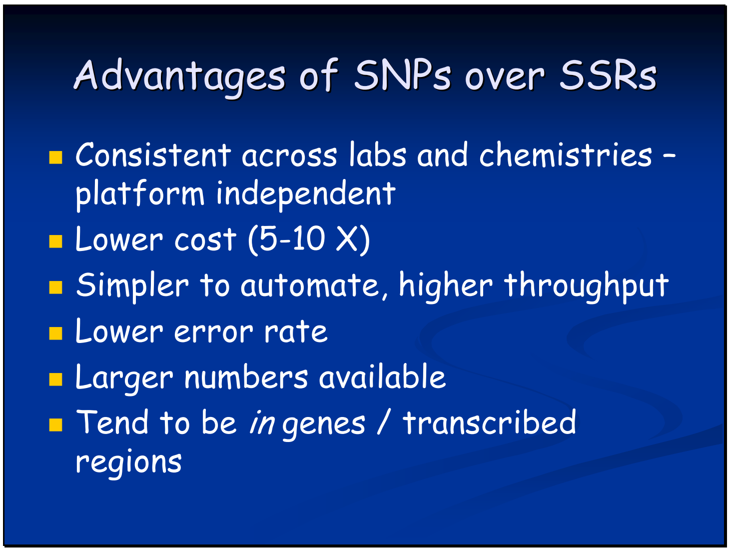
Detection of SNPs

- Nucleotide differences can be detected with a wide range of chemistries

- **The result of allele 'A' versus allele 'C' DOES NOT change with different SNP detection chemistries. It is a 'Universal Code' allowing marker data comparisons to be made from lab to lab, and across companies and continents. Results are standardized and transparent.**

Here a segregating population of maize individuals are being interrogated as to whether they are homozygous for the A allele (blue) or homozygous for the C allele (red) or heterozygous (yellow). By the time profiles from 30 or more SNP loci are interrogated each maize inbred essentially has a fingerprint.

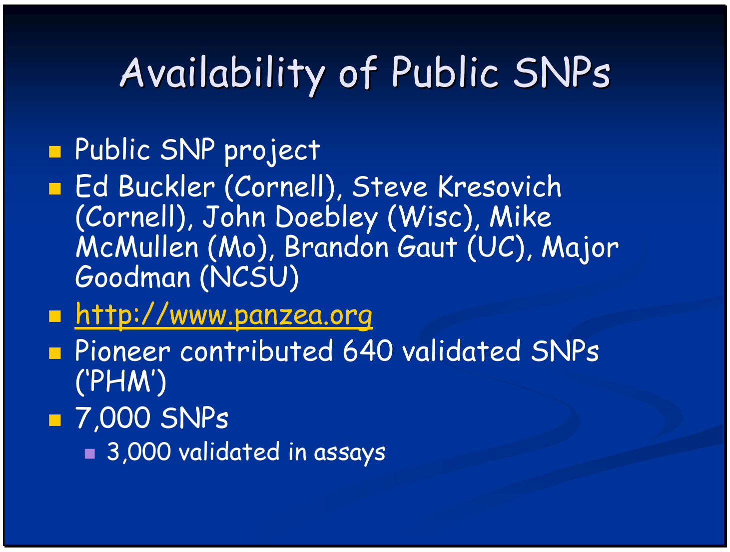
Slide 6



Advantages of SNPs over SSRs

- Consistent across labs and chemistries – platform independent
- Lower cost (5-10 X)
- Simpler to automate, higher throughput
- Lower error rate
- Larger numbers available
- Tend to be *in* genes / transcribed regions

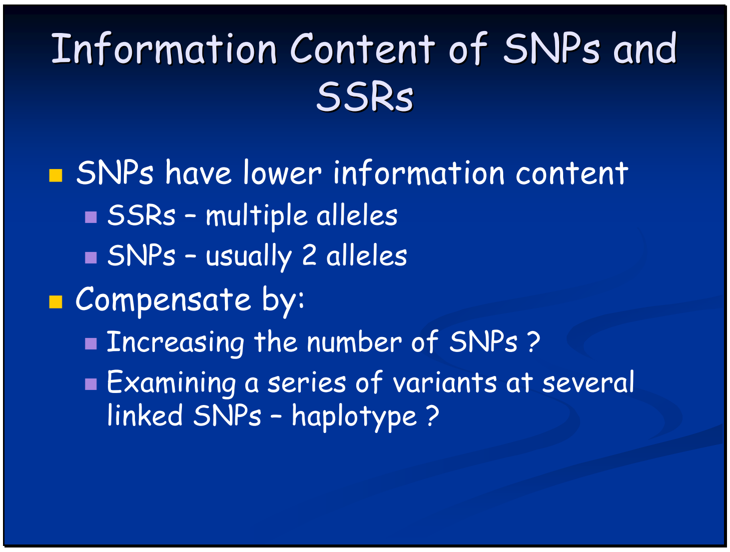
Slide 7



Availability of Public SNPs

- Public SNP project
- Ed Buckler (Cornell), Steve Kresovich (Cornell), John Doebley (Wisc), Mike McMullen (Mo), Brandon Gaut (UC), Major Goodman (NCSU)
- http://www.panzea.org
- Pioneer contributed 640 validated SNPs ('PHM')
- 7,000 SNPs
  - 3,000 validated in assays

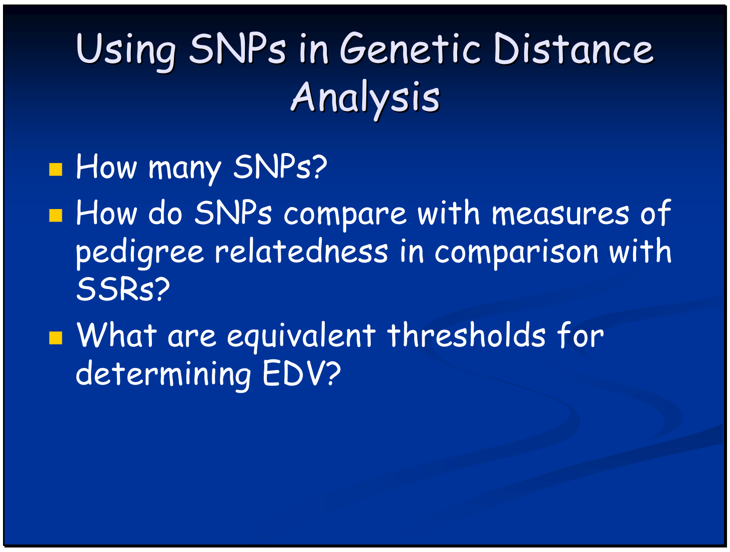
Slide 8



Information Content of SNPs and SSRs

- SNPs have lower information content
  - SSRs – multiple alleles
  - SNPs – usually 2 alleles
- Compensate by:
  - Increasing the number of SNPs ?
  - Examining a series of variants at several linked SNPs – haplotype ?

Slide 9



Using SNPs in Genetic Distance Analysis

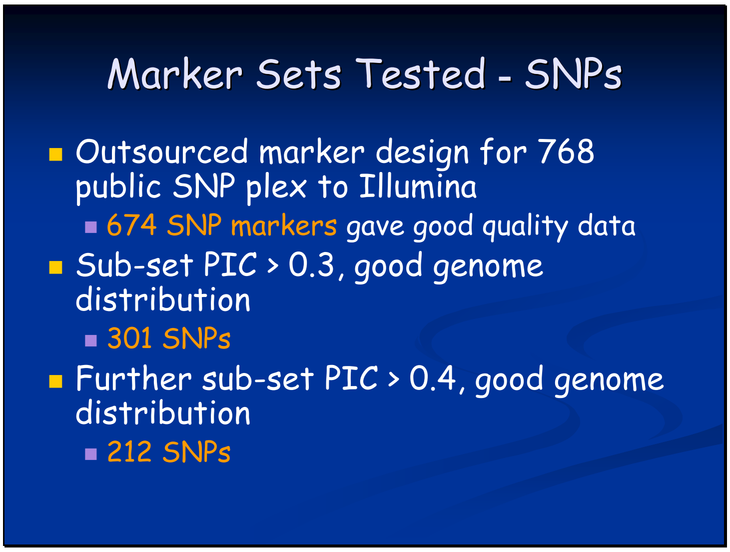
- How many SNPs?
- How do SNPs compare with measures of pedigree relatedness in comparison with SSRs?
- What are equivalent thresholds for determining EDV?

Slide 10

Marker Sets Tested - SSRs

- SEPROMA SSRs
  - 163 recommended, only had data for 90 SSRs
- Standard set of Pioneer SSRs used for fingerprinting
  - 314 SSRs
  - Good genome coverage
  - Good quality data under high throughput conditions
  - All public
- Sub-set of above marker set that meet recommended criteria ISF
  - Distance > 5 cM
  - > 80% genome coverage
  - PIC > 0.3
    - Average PIC across all markers > 0.6
  - = 177 SSRs

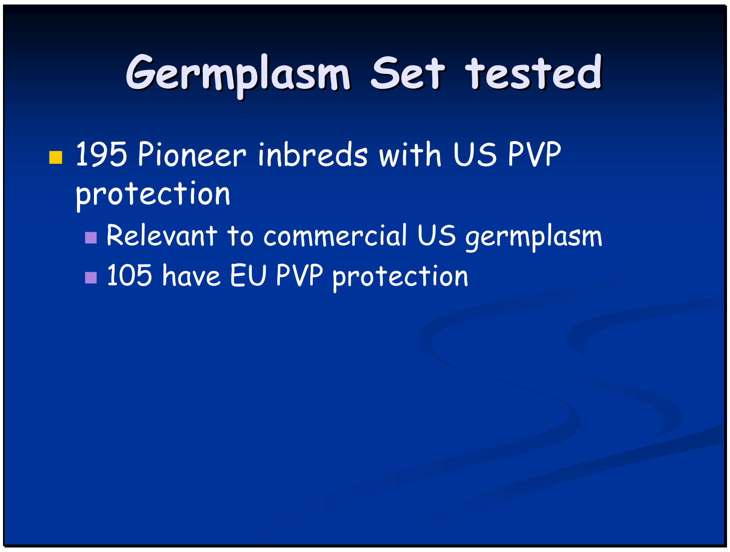
Slide 11



Marker Sets Tested - SNPs

- Outsourced marker design for 768 public SNP plex to Illumina
  - 674 SNP markers gave good quality data
- Sub-set PIC > 0.3, good genome distribution
  - 301 SNPs
- Further sub-set PIC > 0.4, good genome distribution
  - 212 SNPs

Slide 12



Germplasm Set tested

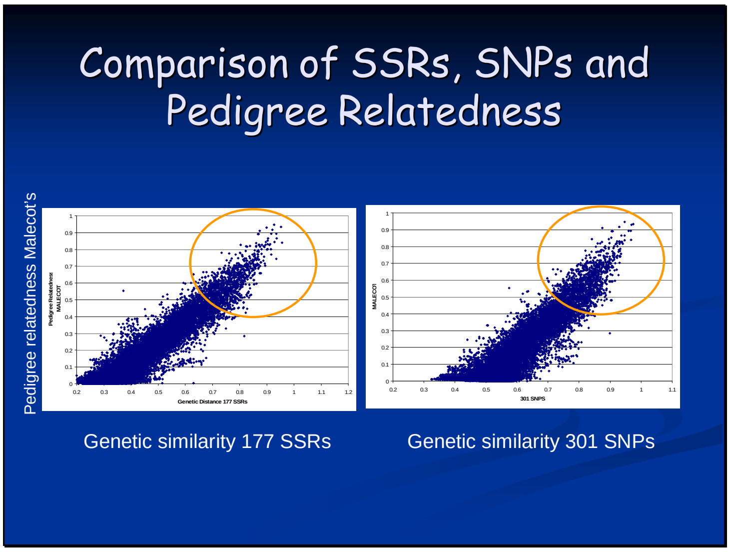
- 195 Pioneer inbreds with US PVP protection
  - Relevant to commercial US germplasm
  - 105 have EU PVP protection

Slide 13

Distance Analyses

- Genetic distances with Band method (Lynch, 1990) in NTSYS
  - Equivalent to Nei and Li (1979)
- Pedigree relatedness with Malecot’s coefficient
- Correlations in NTSYS with Mantel’s T-test
- Line of best fit in Excel

Slide 14



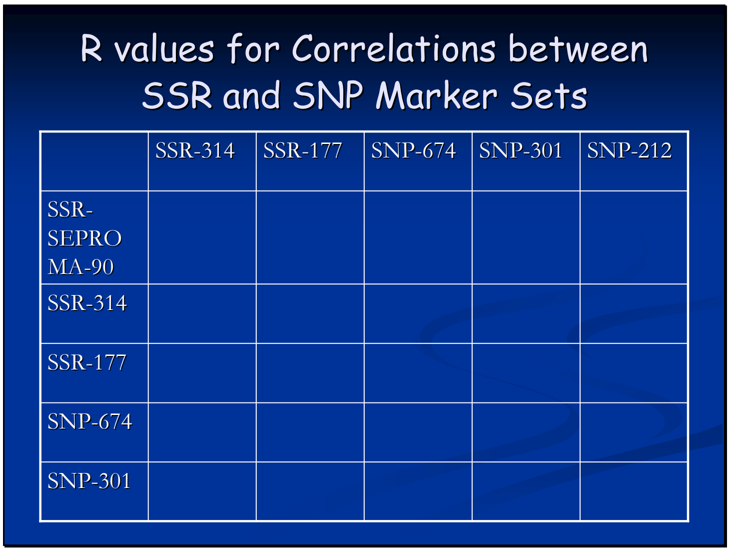
Comparison of SSRs, SNPs and Pedigree Relatedness

Pedigree relatedness Malecot’s

Genetic similarity 177 SSRs

Genetic similarity 301 SNPs

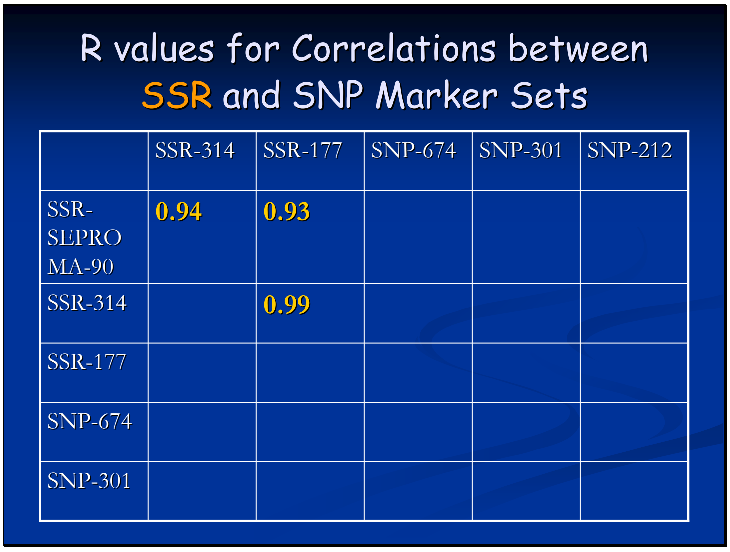
Slide 15



R values for Correlations between SSR and SNP Marker Sets

	SSR-314	SSR-177	SNP-674	SNP-301	SNP-212
SSR-SEPRO MA-90					
SSR-314					
SSR-177					
SNP-674					
SNP-301					

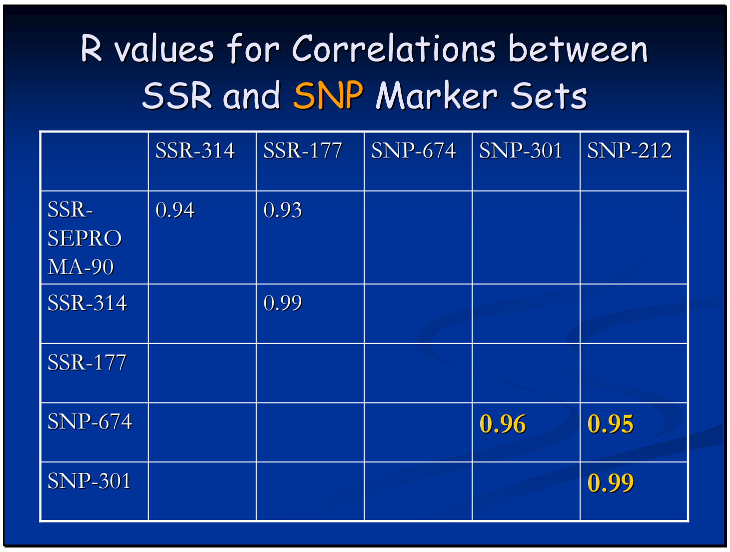
Slide 16



R values for Correlations between SSR and SNP Marker Sets

	SSR-314	SSR-177	SNP-674	SNP-301	SNP-212
SSR-SEPRO MA-90	**0.94**	**0.93**			
SSR-314		**0.99**			
SSR-177					
SNP-674					
SNP-301					

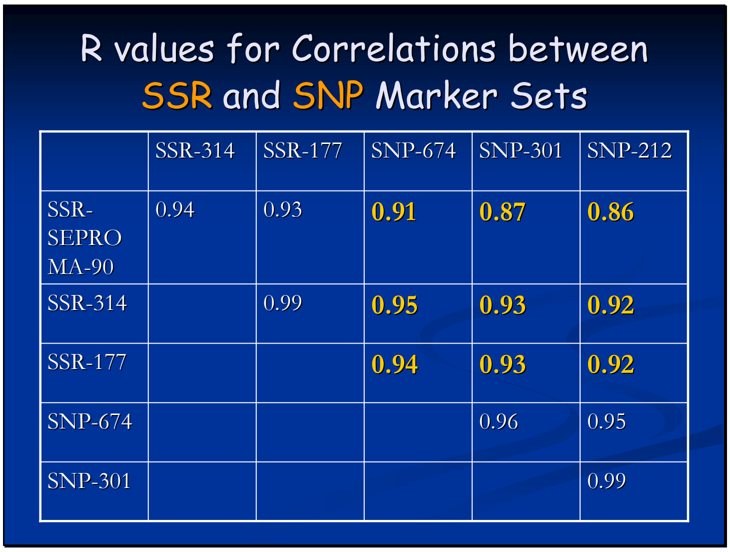
Slide 17



R values for Correlations between SSR and SNP Marker Sets

	SSR-314	SSR-177	SNP-674	SNP-301	SNP-212
SSR-SEPRO MA-90	0.94	0.93			
SSR-314		0.99			
SSR-177					
SNP-674				**0.96**	**0.95**
SNP-301					**0.99**

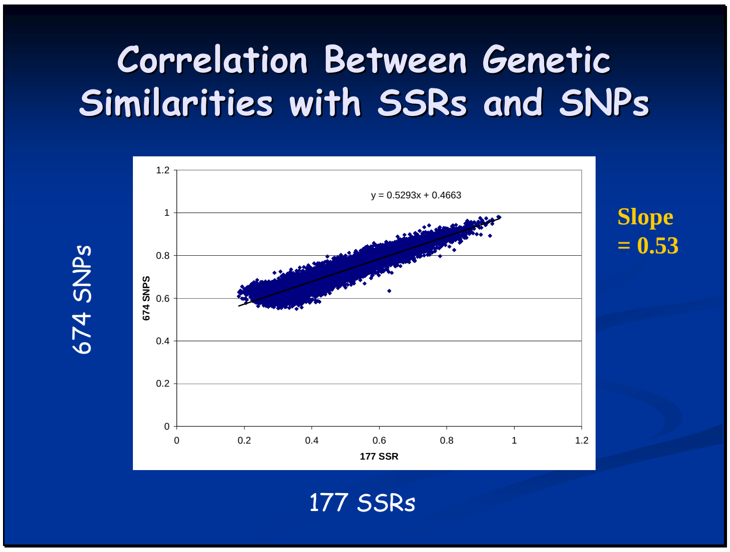
Slide 18



R values for Correlations between SSR and SNP Marker Sets

	SSR-314	SSR-177	SNP-674	SNP-301	SNP-212
SSR-SEPRO MA-90	0.94	0.93	**0.91**	**0.87**	**0.86**
SSR-314		0.99	**0.95**	**0.93**	**0.92**
SSR-177			**0.94**	**0.93**	**0.92**
SNP-674				0.96	0.95
SNP-301					0.99

Slide 19



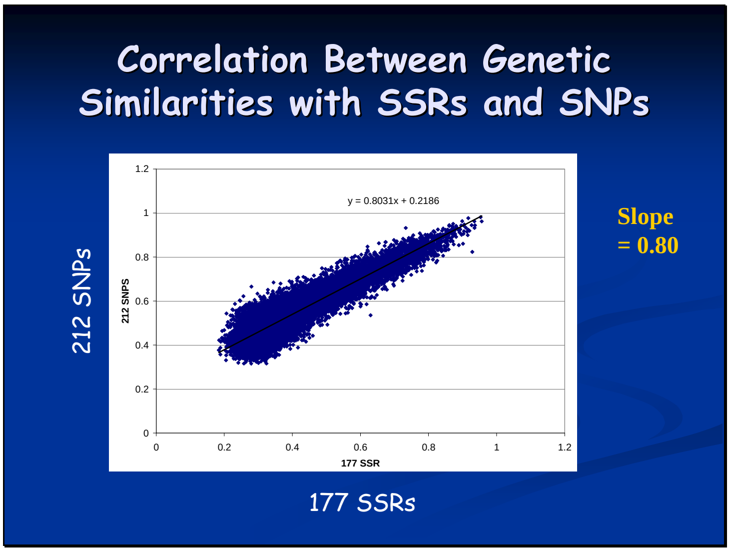
Correlation Between Genetic Similarities with SSRs and SNPs

674 SNPs

177 SSRs

Slope = 0.53

Slide 20



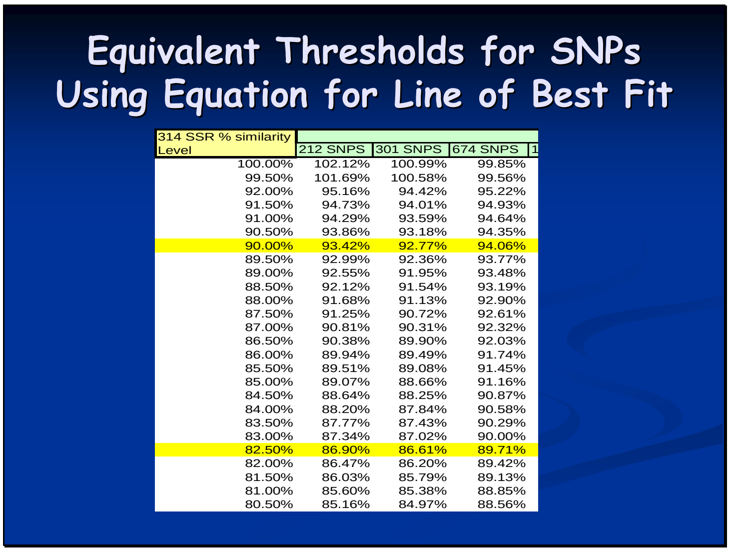
Correlation Between Genetic Similarities with SSRs and SNPs

212 SNPs

177 SSRs

Slope = 0.80

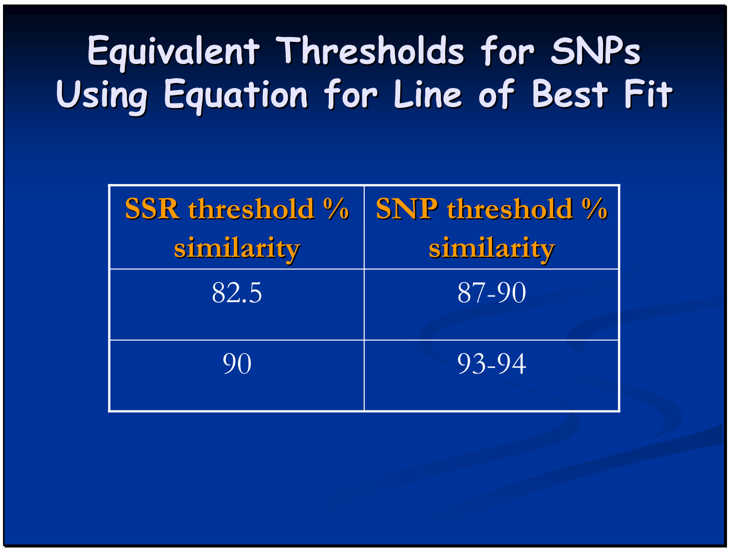
Slide 21



Equivalent Thresholds for SNPs Using Equation for Line of Best Fit

314 SSR % similarity Level	212 SNPS	301 SNPS	674 SNPS
100.00%	102.12%	100.99%	99.85%
99.50%	101.69%	100.58%	99.56%
92.00%	95.16%	94.42%	95.22%
91.50%	94.73%	94.01%	94.93%
91.00%	94.29%	93.59%	94.64%
90.50%	93.86%	93.18%	94.35%
90.00%	93.42%	92.77%	94.06%
89.50%	92.99%	92.36%	93.77%
89.00%	92.55%	91.95%	93.48%
88.50%	92.12%	91.54%	93.19%
88.00%	91.68%	91.13%	92.90%
87.50%	91.25%	90.72%	92.61%
87.00%	90.81%	90.31%	92.32%
86.50%	90.38%	89.90%	92.03%
86.00%	89.94%	89.49%	91.74%
85.50%	89.51%	89.08%	91.45%
85.00%	89.07%	88.66%	91.16%
84.50%	88.64%	88.25%	90.87%
84.00%	88.20%	87.84%	90.58%
83.50%	87.77%	87.43%	90.29%
83.00%	87.34%	87.02%	90.00%
82.50%	86.90%	86.61%	89.71%
82.00%	86.47%	86.20%	89.42%
81.50%	86.03%	85.79%	89.13%
81.00%	85.60%	85.38%	88.85%
80.50%	85.16%	84.97%	88.56%

Slide 22



Equivalent Thresholds for SNPs Using Equation for Line of Best Fit

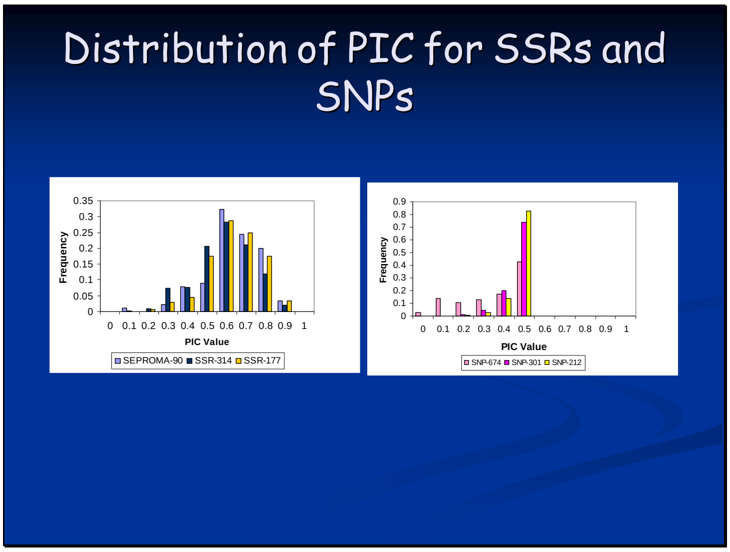
SSR threshold % similarity	**SNP threshold % similarity**
82.5	87-90
90	93-94

Slide 23

Future work

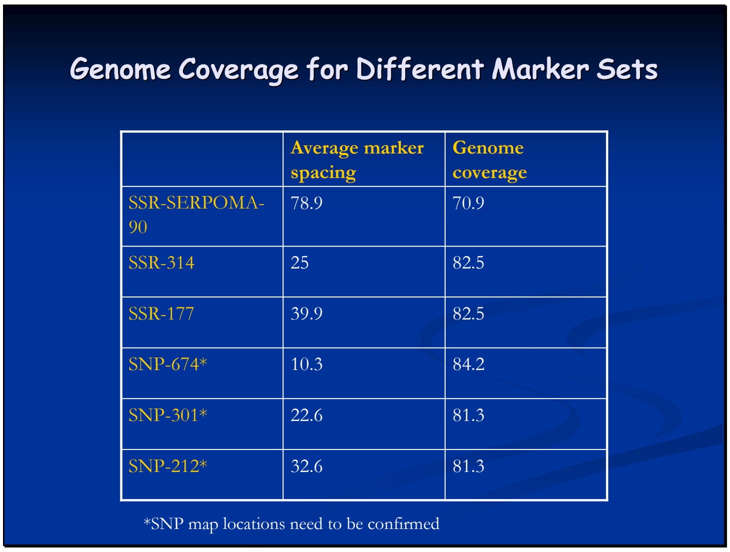
- Profile ASTA US public inbred set with public SNPs
  - Compare ASTA recommended SSR set to SNPs
- Link to SEPROMA study through shared SSRs/inbreds

Slide 24



Distribution of PIC for SSRs and SNPs

Slide 25



Genome Coverage for Different Marker Sets

	Average marker spacing	Genome coverage
SSR-SERPOMA-90	78.9	70.9
SSR-314	25	82.5
SSR-177	39.9	82.5
SNP-674*	10.3	84.2
SNP-301*	22.6	81.3
SNP-212*	32.6	81.3

*SNP map locations need to be confirmed

[End of document]