



TEXAS A&M
UNIVERSITY

Identifying Potential Nesting of Fiber Quality Alleles within the Population Structure of Obsolete U.S. Cotton Cultivars

Mitchell Schumann Dr. C. Wayne Smith Dr. Alan Dabney

Fiber Quality in Upland Cotton

- ▶ **Cotton (*G. hirsutum*) 95% of global cotton fiber production**
 - ▶ US largest exporter of cotton
 - ▶ Texas largest export commodity
 - ▶ 1.6 Billion Dollars to Texas economy (Texasagriculture.gov, 2016)
- ▶ **Improving fiber quality is important**
 - ▶ listed as an important priority producers want from science (Hake 2016)
 - ▶ Allow for use in more efficient textile technologies
 - ▶ Produce better products
 - ▶ Obtain a premium for farmers

Fiber Quality in Upland Cotton

- ▶ **High Volume Instrumentation (HVI) was not widely implemented into breeding programs until the 1980s and 90s.**
 - ▶ 100 years of selections not using these technologies
 - ▶ Many sources of superior fiber quality alleles likely overlooked
- ▶ **Advanced Fiber Information System (AFIS) was developed in the 1990s (cost limits use/acceptance)**

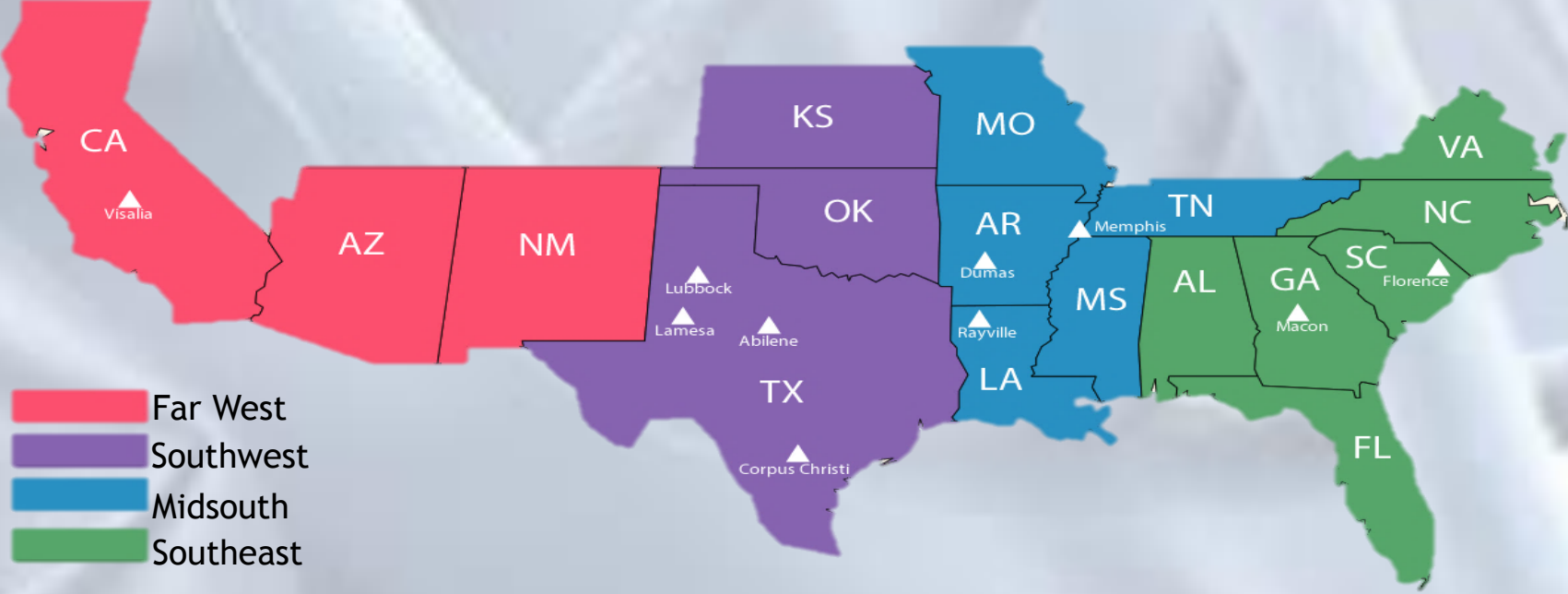
Genetic Diversity of Cotton

- ▶ Most cotton cultivars in the United States can be traced back to an original 1806 introduction from Mexico (Smith and Cothren 1999)
- ▶ Many genetic studies corroborate a narrow genetic diversity (Paterson 2009 and references therein)
- ▶ Low genetic diversity could allow for tapping into standing genetic variation of obsolete cultivars without much yield drag (Tyagi et al. 2014)
- ▶ Can we tap into alleles that may have been overlooked from the USDA Obsolete US Cultivar Collection using genomic selection?
 - ▶ What role does population structure play with fiber quality?
 - ▶ Lui et al. 2015 (PLOS one), Wientjes et al. 2013 (Genetics),exc.

What We are Working With

- ▶ **Genotype information on 258 obsolete US cotton Cultivars**
 - ▶ Selected based on SSR diversity analysis of collection
 - ▶ Hinze et al. 2015 (TAG)
 - ▶ Genotyped using 63K SNP array
 - ▶ Hulse-Kemp et al. 2015 (G3)
 - ▶ Approximately 24,000 high confidence SNPs
- ▶ **Phenotype data on sub-sample of 128 cultivars**
 - ▶ Planted in 2016 using RCBD with 3 replications in 2 locations
 - ▶ HVI & AFIS
 - ▶ Waiting on spin data

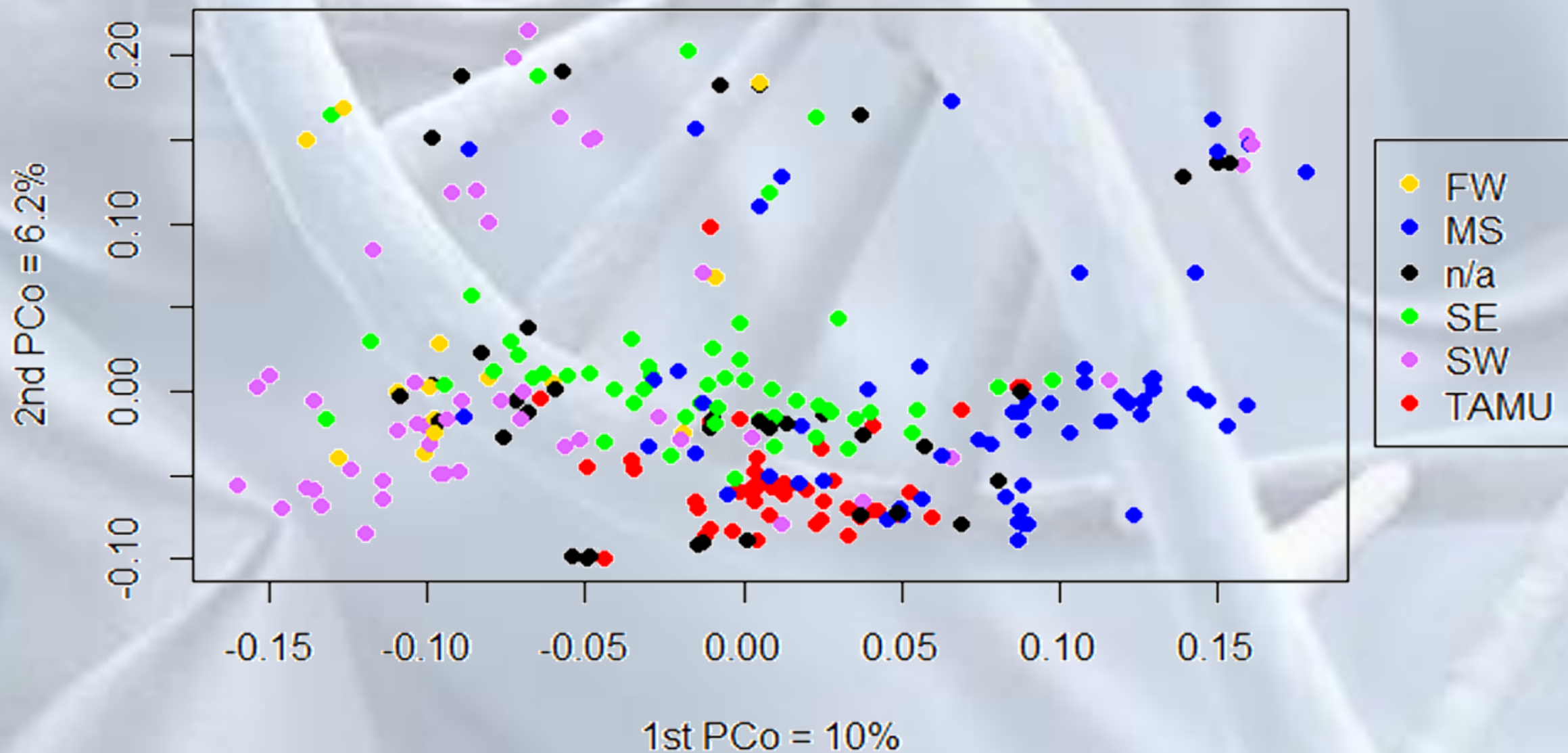
US Cotton Growing Regions



- Far West
- Southwest
- Midsouth
- Southeast

U.S. Dept. of Agriculture Classing office

2 Dimensional PCoA Colored by Location



2020
2019
2018
2017
2016
2015
2014
2013
2012
2011
2010
2009
2008
2007
2006
2005
2004
2003
2002
2001
2000
1999
1998
1997
1996
1995
1994
1993
1992
1991
1990
1989
1988
1987
1986
1985
1984
1983
1982
1981
1980
1979
1978
1977
1976
1975
1974
1973
1972
1971
1970
1969
1968
1967
1966
1965
1964
1963
1962
1961
1960
1959
1958
1957
1956
1955
1954
1953
1952
1951
1950
1949
1948
1947
1946
1945
1944
1943
1942
1941
1940
1939
1938
1937
1936
1935
1934
1933
1932
1931
1930
1929
1928
1927
1926
1925
1924
1923
1922
1921
1920
1919
1918
1917
1916
1915
1914
1913
1912
1911
1910
1909
1908
1907
1906
1905
1904
1903
1902
1901
1900
1899
1898
1897
1896
1895
1894
1893
1892
1891
1890
1889
1888
1887
1886
1885
1884
1883
1882
1881
1880
1879
1878
1877
1876
1875
1874
1873
1872
1871
1870
1869
1868
1867
1866
1865
1864
1863
1862
1861
1860
1859
1858

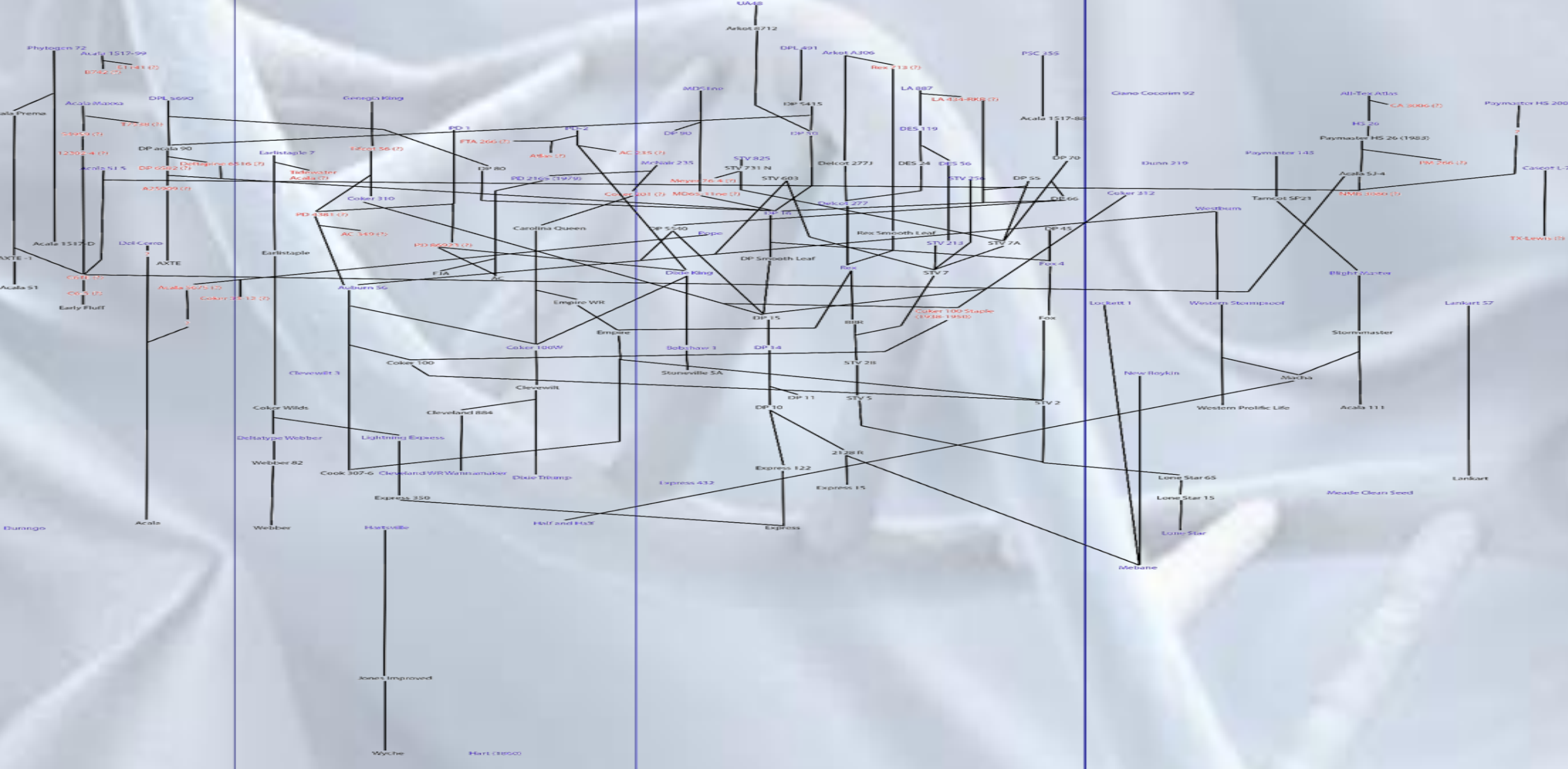
Far West

South East

Mid-South

South West

DP 1558NR B2XF DP 1555 B2XF DG 3355 B2XF ST 5115 GLT NG 3405 B2XF ST 6182 GLT PHY 222 WRF FM 1900 GLT FM 2007 GLT



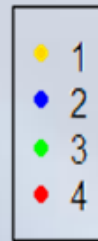
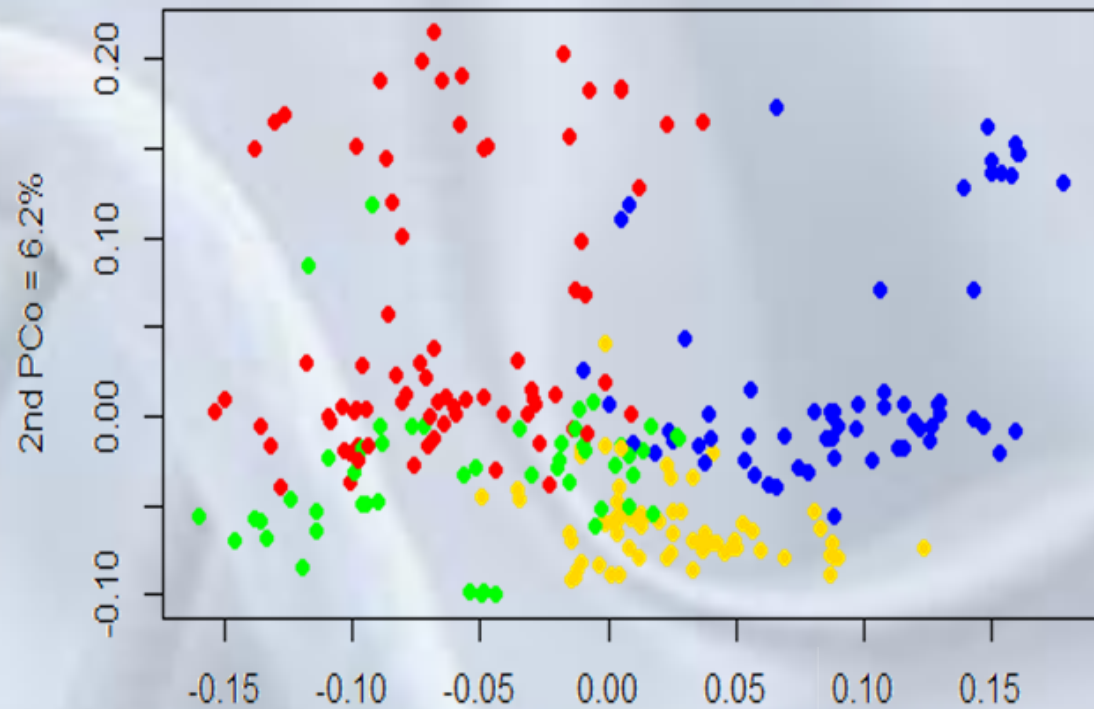
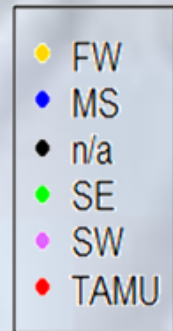
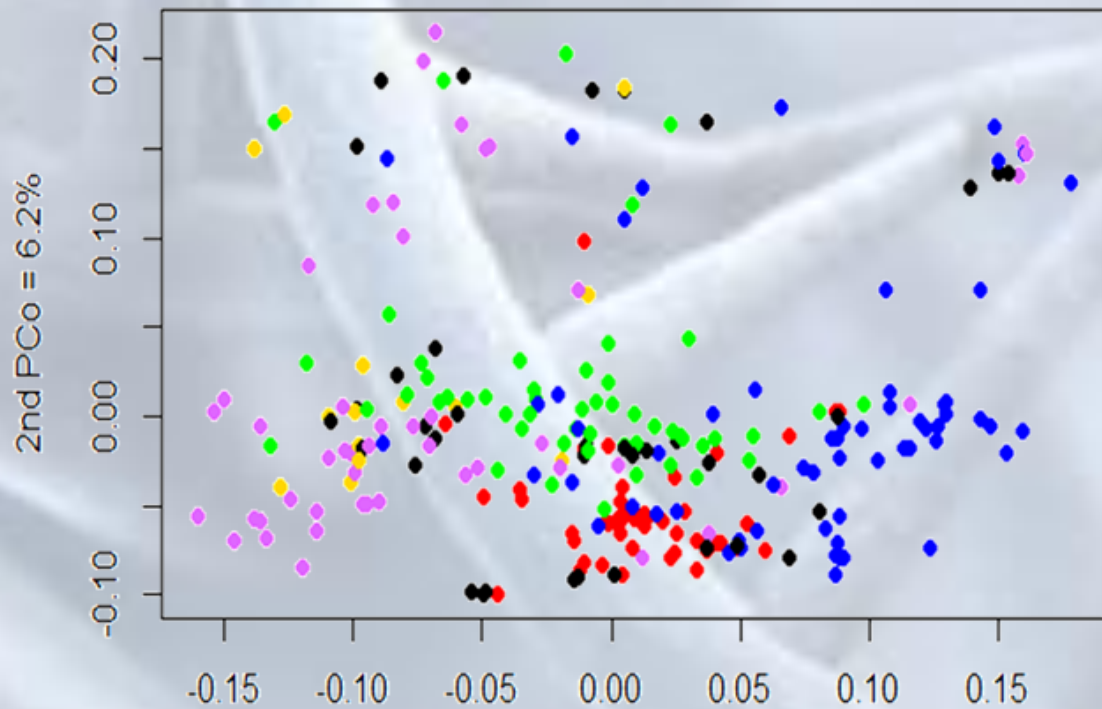
Quantifying Population Structure

▶ **fastStructure**

- ▶ Raj et al. 2014 (Genetics)
- ▶ K=1 through K=10
 - ▶ Not clear at distinguishing the best model
- ▶ Combined with PCoA to see what makes the most sense

2 Dimensional PCoA Colored by Location

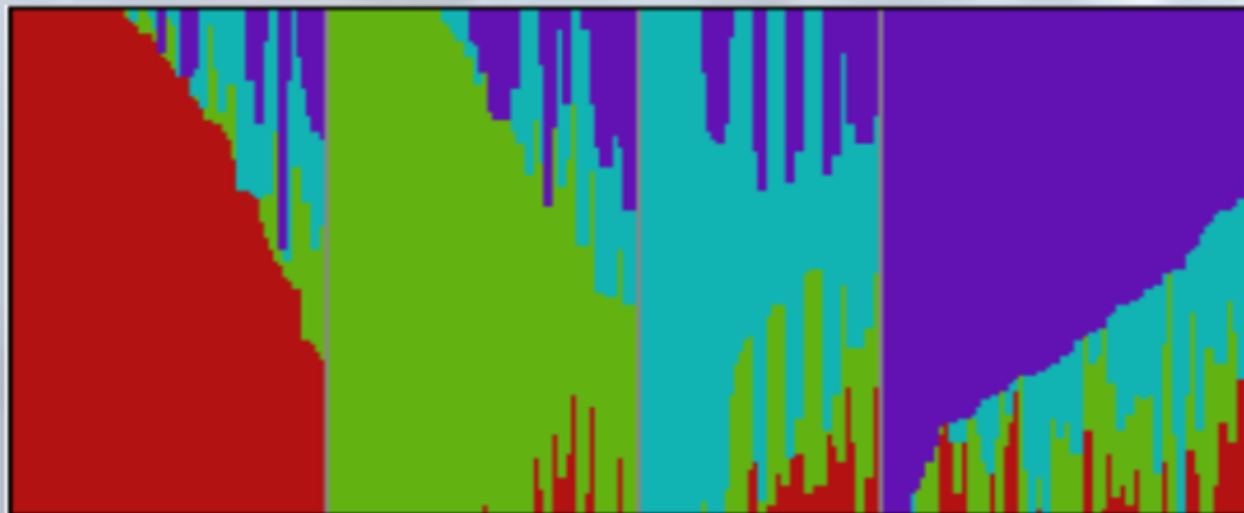
2 Dimensional PCoA Colored by K = 4



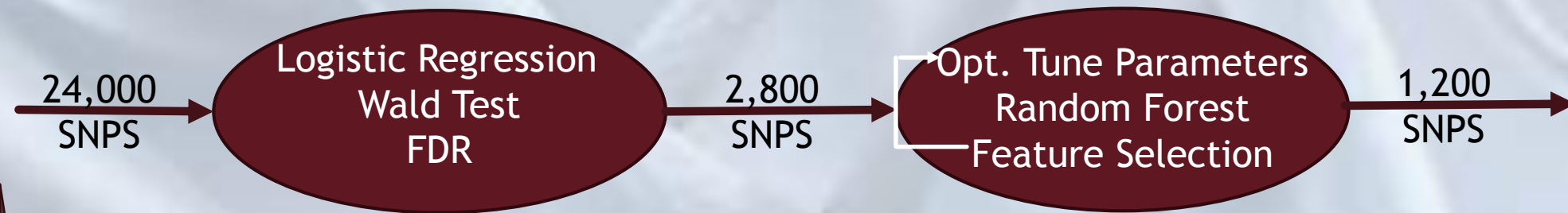
Distruct Plot

1st PCo = 10%

1st PCo = 10%



What Markers Contribute the Most to Population Structure

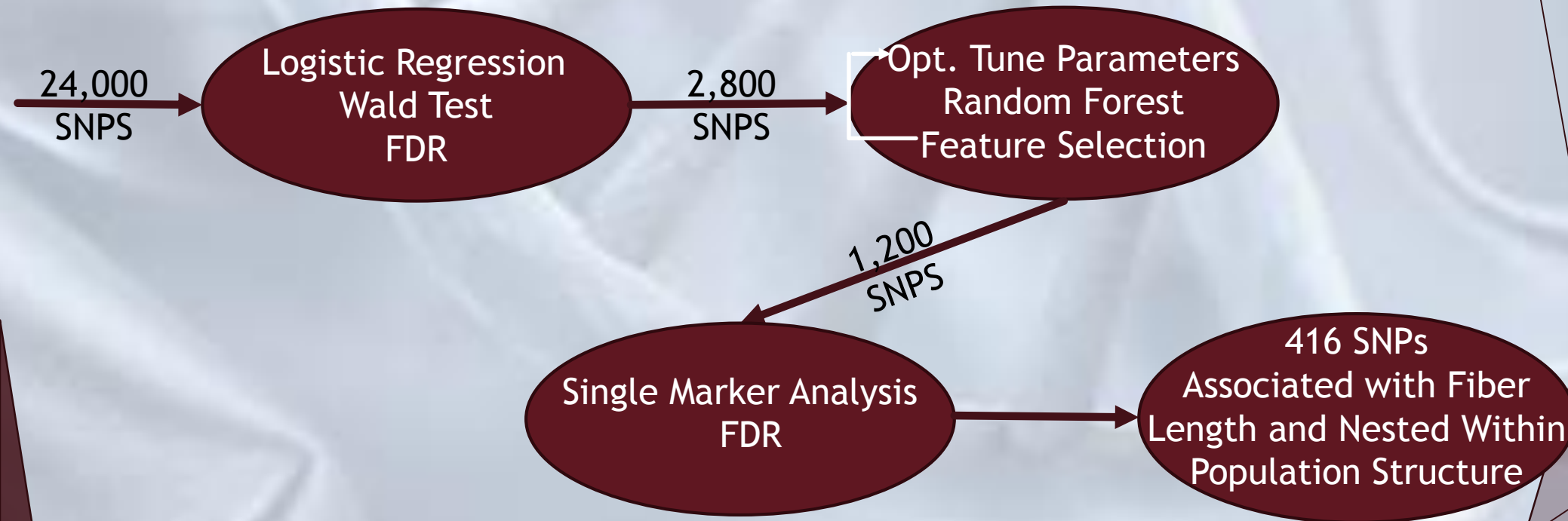


Prediction Error Comparison for Full and Reduced Random Forest Model

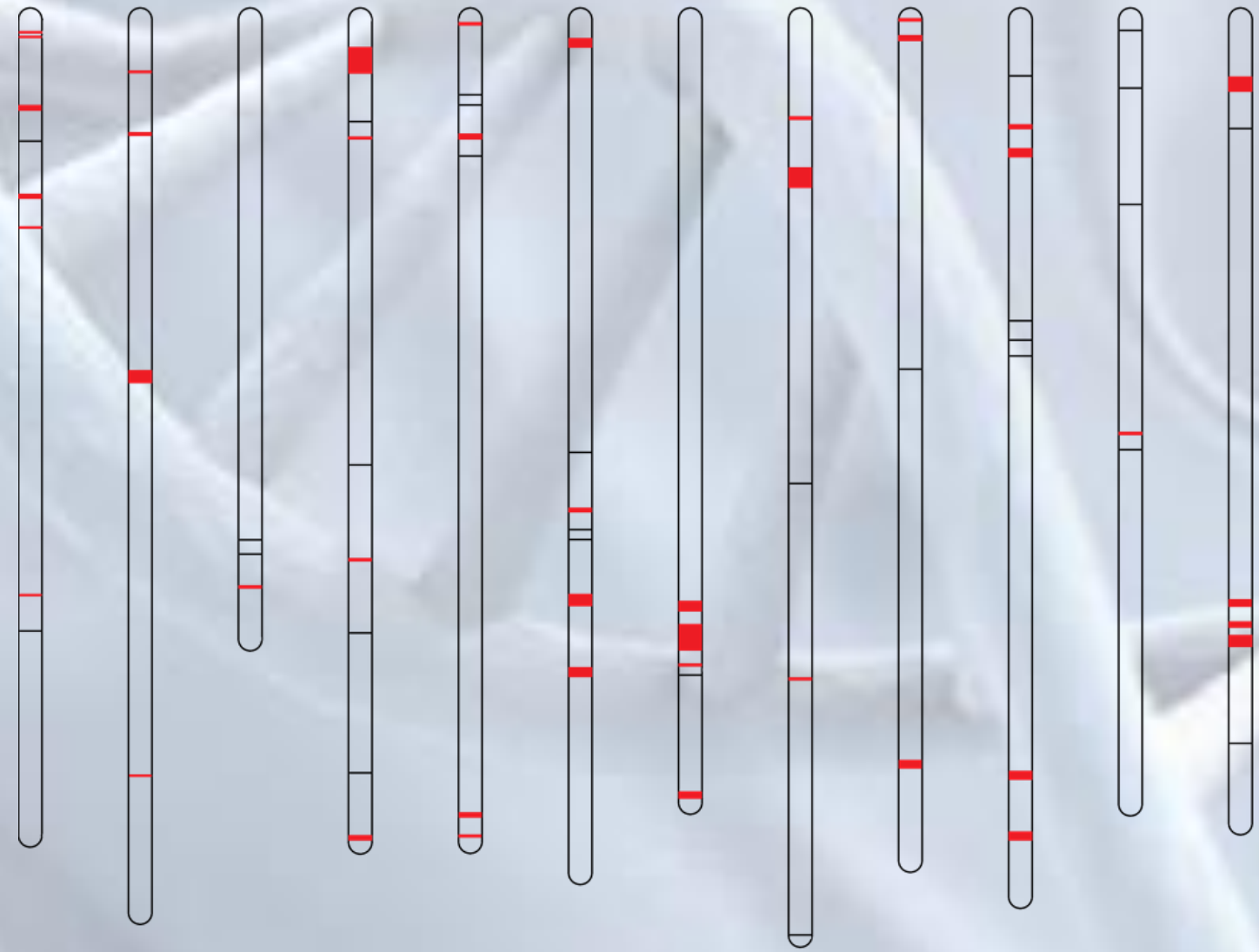
Model	1 st Quartile	2 nd Quartile	Median	Mean	3 rd Quartile	4 th Quartile
Full	0.07364	0.09302	0.1008	0.09767	0.1037	0.1124
Reduced	0.0814	0.09012	0.0969	0.09574	0.09981	0.1085

- ▶ 10 Replications of K = 10 Folds Cross Validation

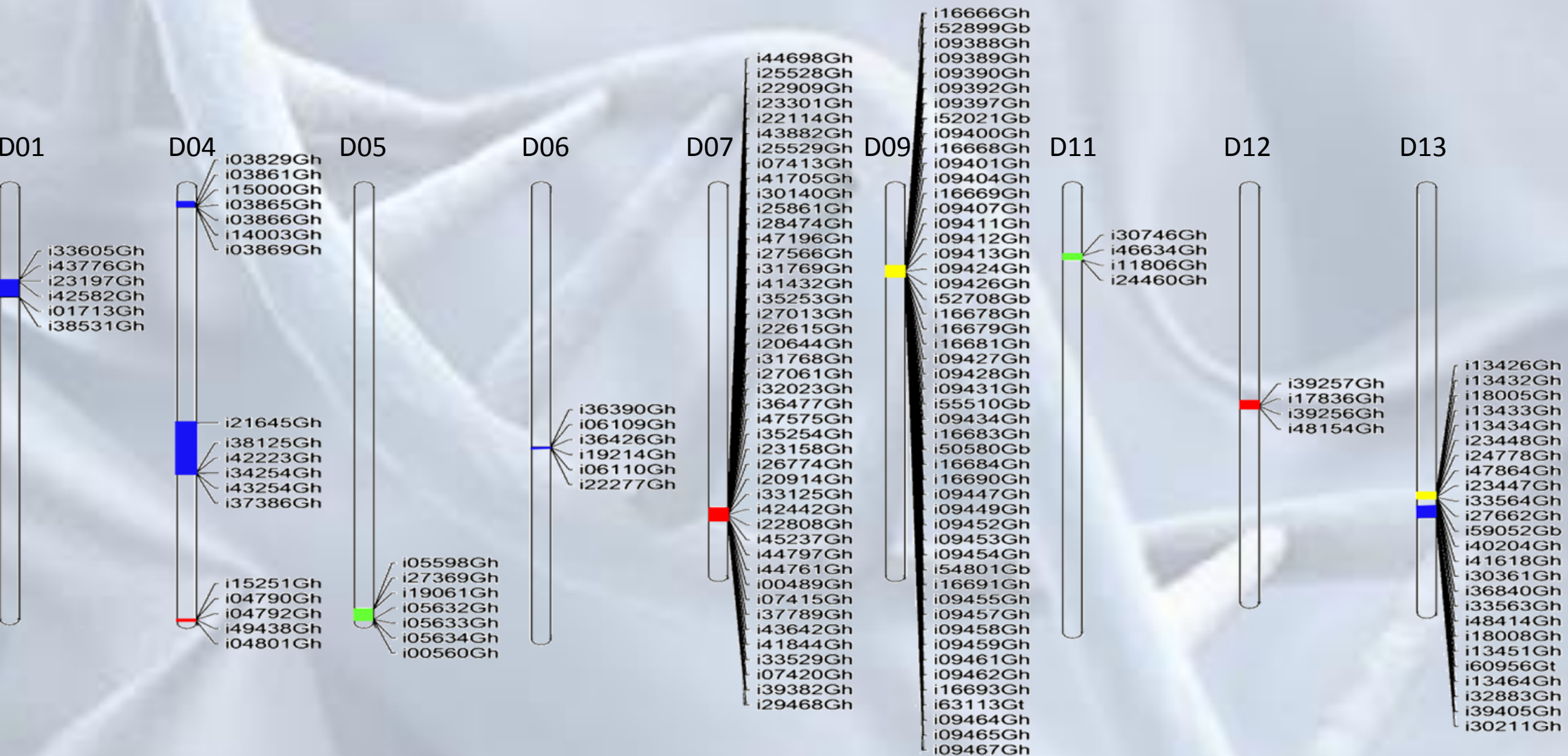
What Markers Contribute the Most to Population Structure



D01 D02 D03 D04 D05 D06 D07 D09 D10 D11 D12 D13



Multiple Markers Clustered Together



Sub 1

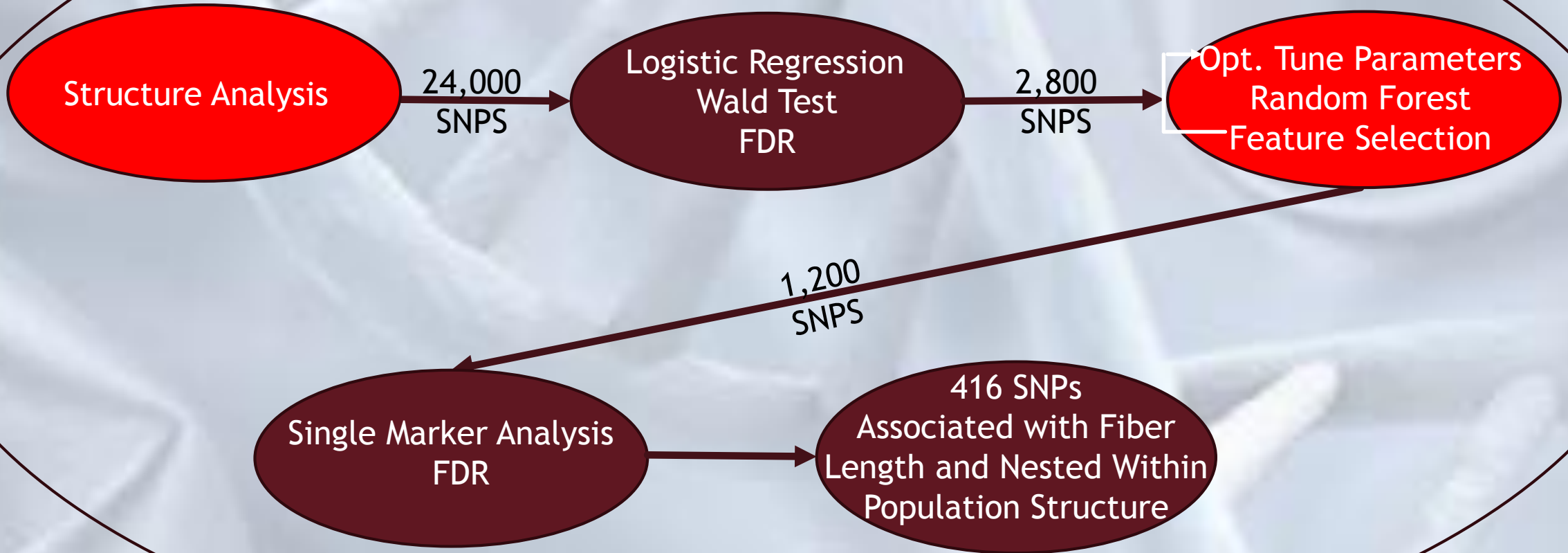
Sub 2

Sub 3

Sub 4

Potential Problem of Overfitting

Cross Validation Loop



Summary

- ▶ This analysis does appear to Identify nesting fiber quality alleles in population structure.
- ▶ This information could prove useful in developing training populations for genomic selection, developing GWAS population, and in general breeding decision making.
- ▶ Potential Problem of Overfitting

Thanks

▶ Cotton Incorporated

▶ Dr. Don Jones

▶ FBRI Fiber Initiative-Texas Tech University

▶ Dr. Eric Hequet, Zach Hinds

▶ Cotton Improvement Lab- Texas A&M University

▶ Dr. C.Wayne Smith, Dr. Steve Hague Mrs. Dawn Deno,
Graduate Students, Undergraduate Workers

▶ TIGSS

▶ Dr. Andrew Hillhouse, Kelli Kochan

▶ Important Contributors

▶ Dr. Alan Dabney, Dr. Lori Hinze, Dr. David Stelly



™

COTTON INCORPORATED

Questions

