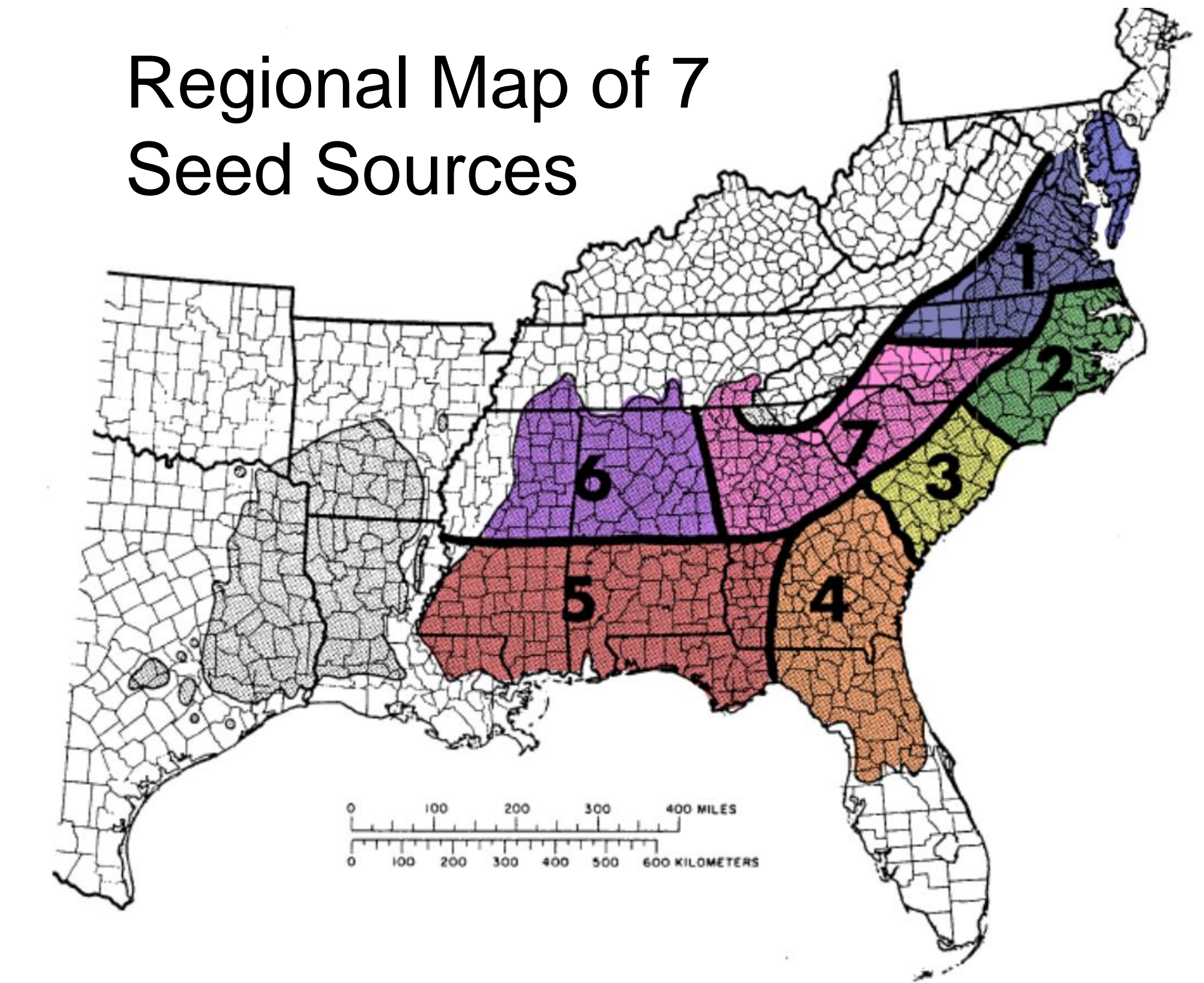


- 90% of samples had detectable SNPs
- An average of 62,089 SNPs/individual

Regional Map of 7 Seed Sources



F<sub>st</sub> Calculations Among Seed Sources

5	4	1	6	2	7	
0.005322	0.008592	0.005528	0.006278	0.005886	0.005235	3
	0.007754	0.00709	0.005853	0.007407	0.006695	5
		0.011488	0.00654	0.013245	0.01098	4
			0.008102	0.004743	0.004796	1
				0.009311	0.008044	6
					0.00491	2

- The highest amount of differentiation occurs between population 4 and 2
- The lowest amount of differentiation occurs between population 1 and 2
- There appears to be a substantial amount of missing data

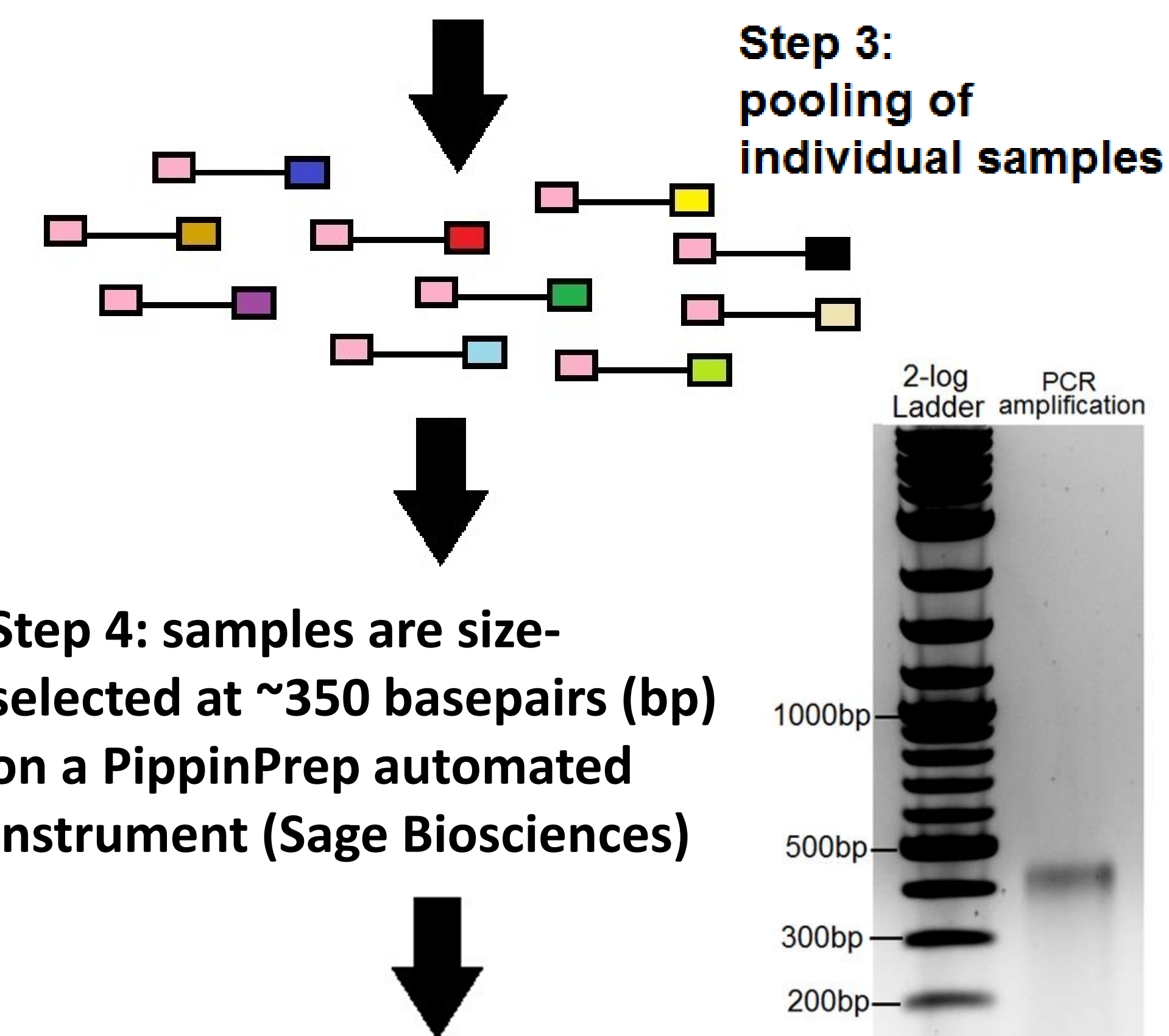
## Further Analysis

- Conducting analysis using the newly published loblolly pine genome to perform genotype calling and make a relationship matrix to determine previously unknown pollen parents.
- Conducting association testing with haplotype and phenotypic data to determine if loci are in LD with genes of adaptive importance

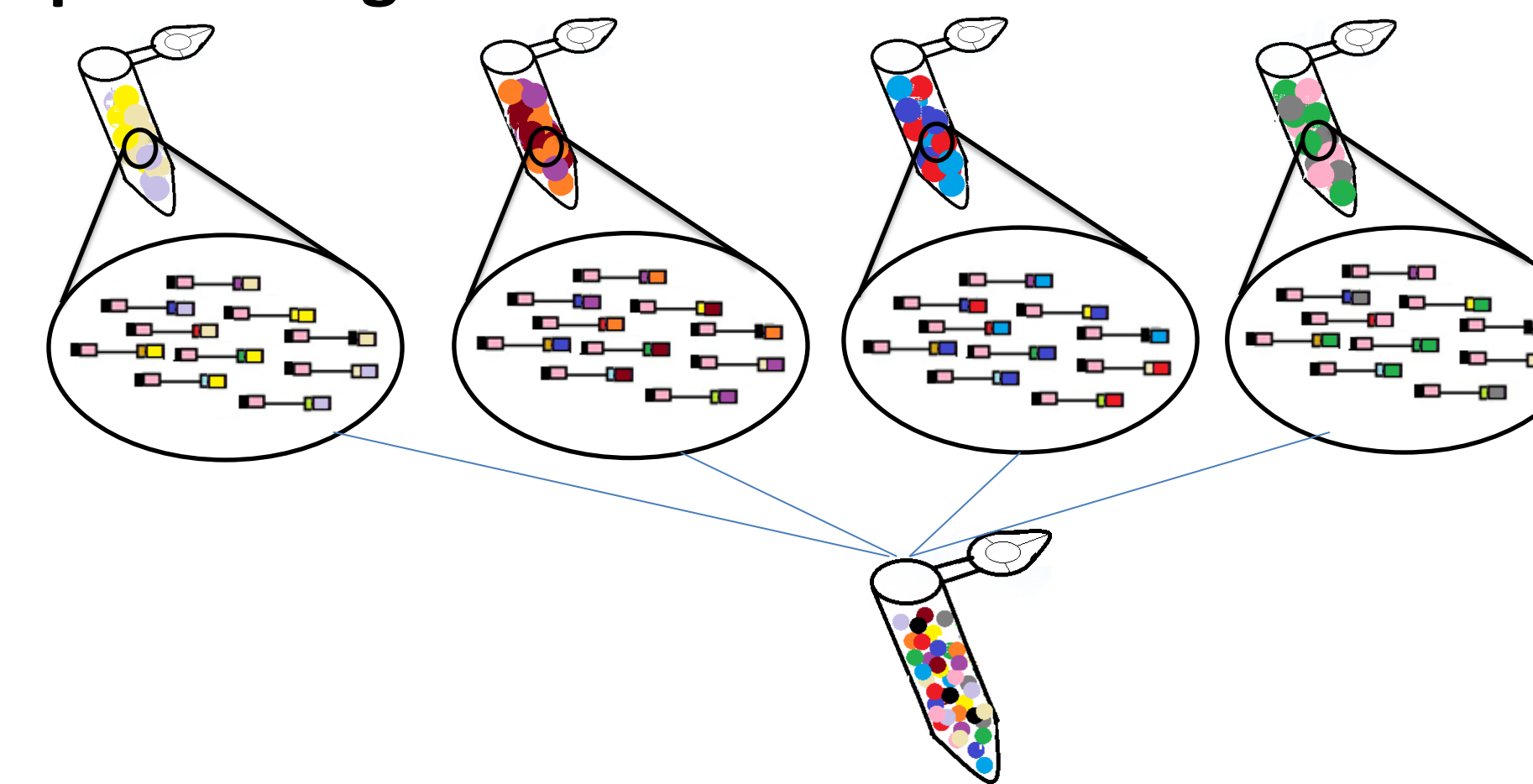
## Acknowledgments

- We thank Ed Buckler, Jeff Glaubitz, Terry Casstevens, Fei Lu, and James Harriman for making the GBS analysis pipeline software available, and for their guidance in data analysis.
- We thank the members of the NCSU Cooperative Tree Improvement Program, as well as past and current faculty and staff of the Program, for their work to establish and maintain the Plantation Selection Seed Source Study, and for making that study available for use in this project.
- We thank Will Kohlway for his capable technical assistance in getting the GBS protocol working to produce data from pine.

## GBS continued



**Step 5:**  
The resulting DNA is then used as template for 8 parallel PCR reactions, using 3 indexing primers per plate of samples allowing 4 plates to be pooled together.



- Sequencing on Illumina HiSeq2000.
- Spread one pool over 4 lanes (~ 1 plate/lane)
- Analysis using STACKS software.

## Cost Analysis

	Extraction	Library Prep	Sequencing	Grand total
	Reagents + plastics	Reagents + plastics	BGI	
per plate	\$ 49.37	\$ 107.77	\$ 1,425.00	\$ 1,582.14
per sample	\$ 0.51	\$ 1.12	\$ 14.84*	\$ 18.12

\*Actual sequencing yielded ~16.94 when adjusting for lost samples

## Sample Processing

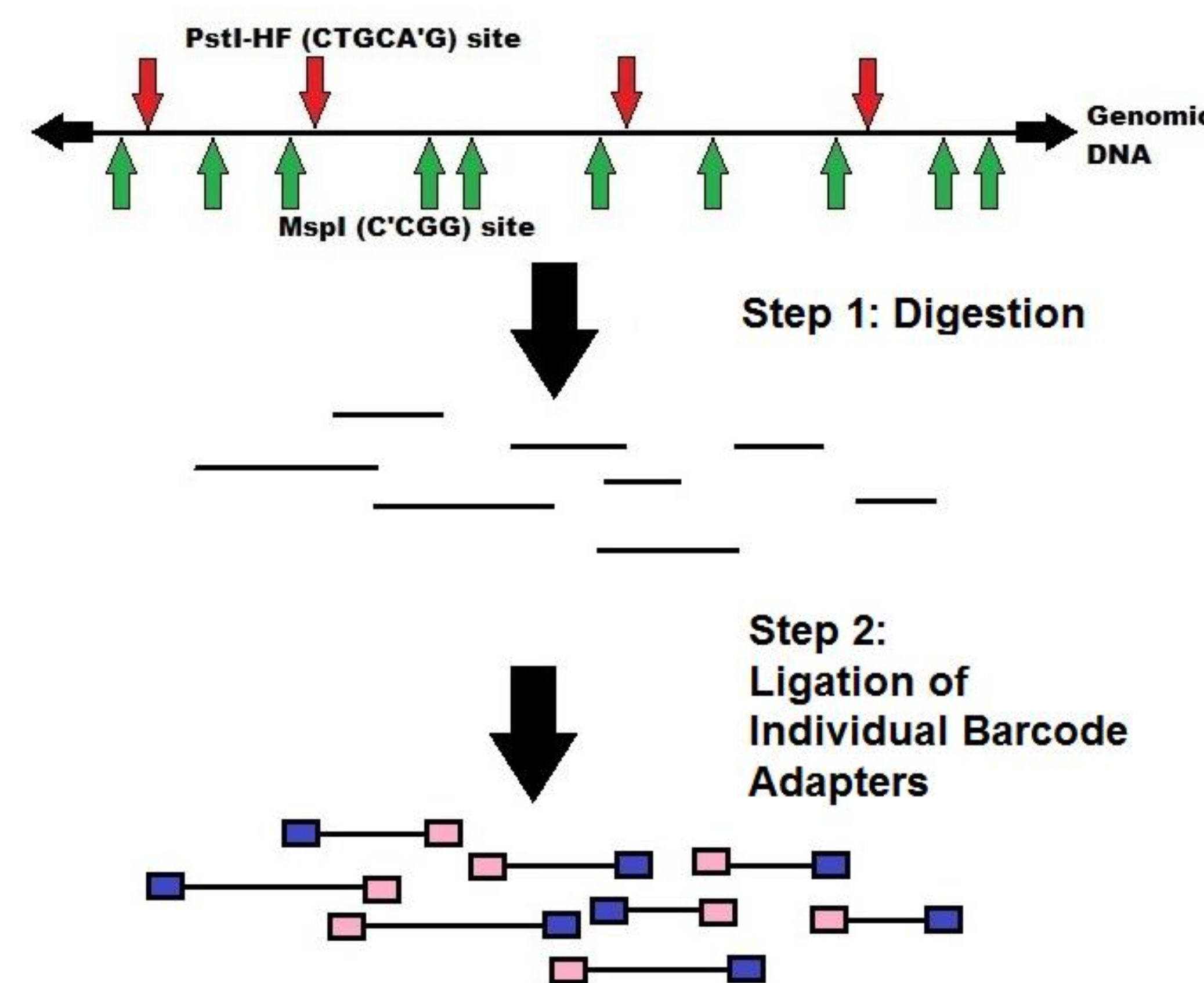
- DNA extraction was adapted from a published protocol (Ivanova et al, 2008).
- Phloem disk halves were rolled flat and placed in 96-well plates for extraction



- 1471 samples had DNA yields high enough to be sequenced.

## GBS

- First described for maize and barley by Elshire et al. (2011), and for wheat and oats by Poland et al. (2012).
- Peterson et al. (2012) included size-selection step and use of an index primer set to increase the number of individual samples.

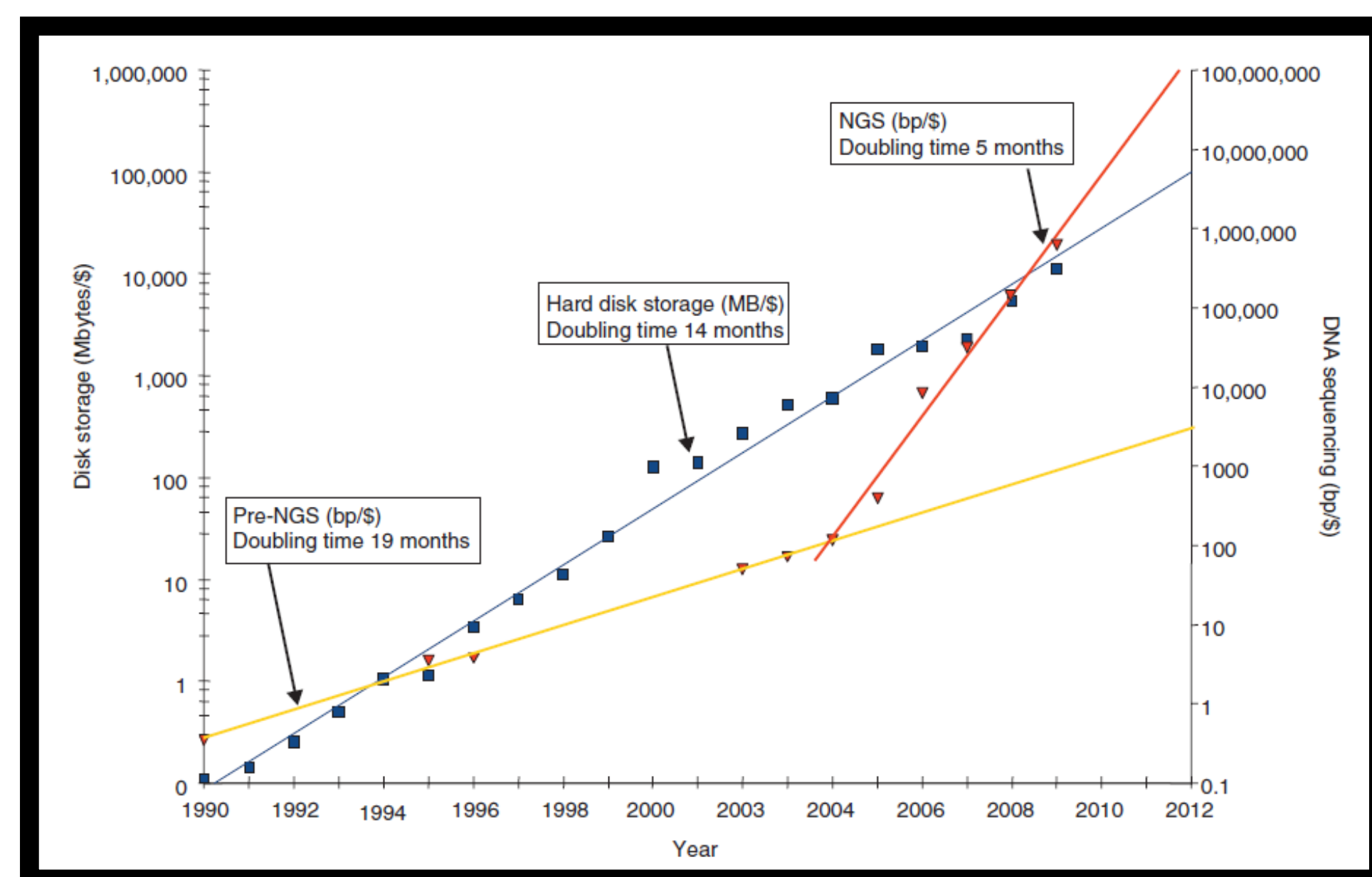


## References

Elshire RJ, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:e19379.  
 Ivanova et al. 2008 Semi-automated, membrane-based protocol for DNA isolation from plants. Plant Mol. Biol. Rep 26:186-198.  
 Peterson BK, et al. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7(5): e37135.  
 Poland JA, et al. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253.  
 Stein LD. 2010. The case for cloud computing in genome informatics. Genome Biol 11(5):207.

## Introduction

- Existing genotyping methods are too expensive for breeding use
- Data yield per dollar from DNA sequencing is increasing exponentially (Stein, 2010-see figure below).
- DNA sequencing may soon provide a cost-effective genotyping method



## Research Objectives

- Develop cost-effective genotyping for pine
- Genotype samples from progeny of parents from 7 seed sources of loblolly pine, growing in common location
- Test for interactions of genetic variants

## Sample Collection

- 1600 tissue samples collected in Screven County, Georgia



## References