

Reference Genome Enabled Variant Discovery in Octoploid Strawberry

Mitchell J. Feldmann¹, Michael A. Hardigan¹, Thomas J. Poorten¹, Charlotte B. Acharya¹, Marivi Colle², Patrick P. Edger²,

¹Department of Plant Sciences, University of California, Davis, Davis, CA,

²Department of Horticulture, Michigan State University, East Lansing, MI

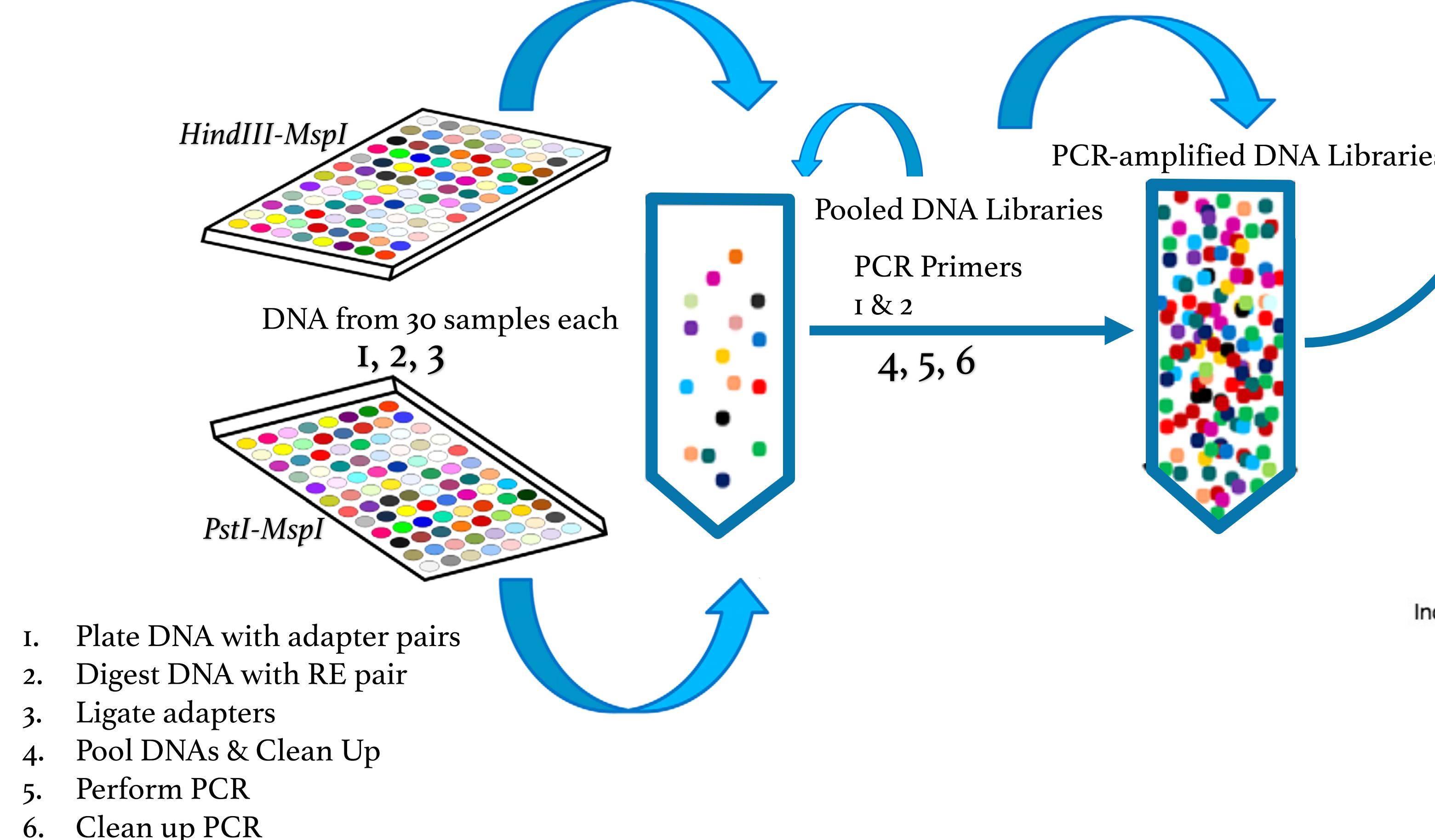
Robert VanBuren², and Steven J. Knapp¹

Introduction

The absence of a reference genome sequence for (*Fragaria × ananassa* Duch.), an allo-octoploid ($2n=8x=56$), previously hindered the application of genotyping-by-sequencing (GBS) and other next generation sequencing (NGS)-based DNA variant discovery. An octoploid reference genome assembly drastically simplifies data analysis and provides a technical framework for the rigorous discovery of markers using short-read sequencing technologies. The ability to determine the sub-genome specificity of short-reads needs to be investigated to enable the exploitation of NGS-based high-throughput genotyping approaches in octoploid strawberry. Here, we employ the double enzyme digest protocol put forward by Poland et al (2012).

Methods: Library Preparation & Bioinformatics

A.



Objectives

1. Develop a custom bioinformatics pipeline for GBS-facilitated DNA variant discovery in octoploid strawberry.
2. Utilizing a custom bioinformatics pipeline (8X-GBS), quantify the genomic distribution and sub-genome specificity of DNA variants mapped to an octoploid reference genome from sequenced genomic libraries prepared with either *HindIII-MspI* and *PstI-MspI*.

B.

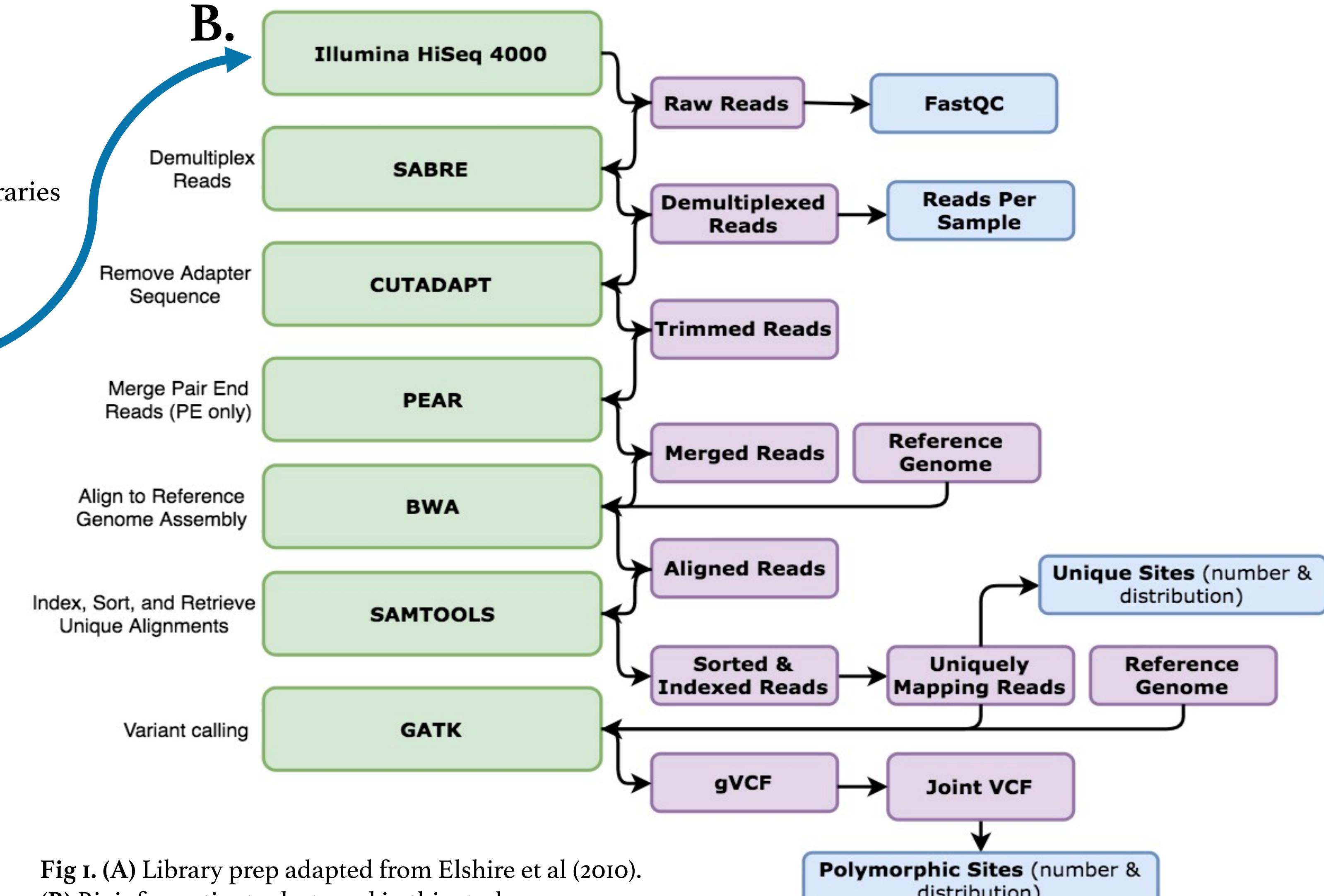


Fig 1. (A) Library prep adapted from Elshire et al (2010). (B) Bioinformatics toolset used in this study.

Results

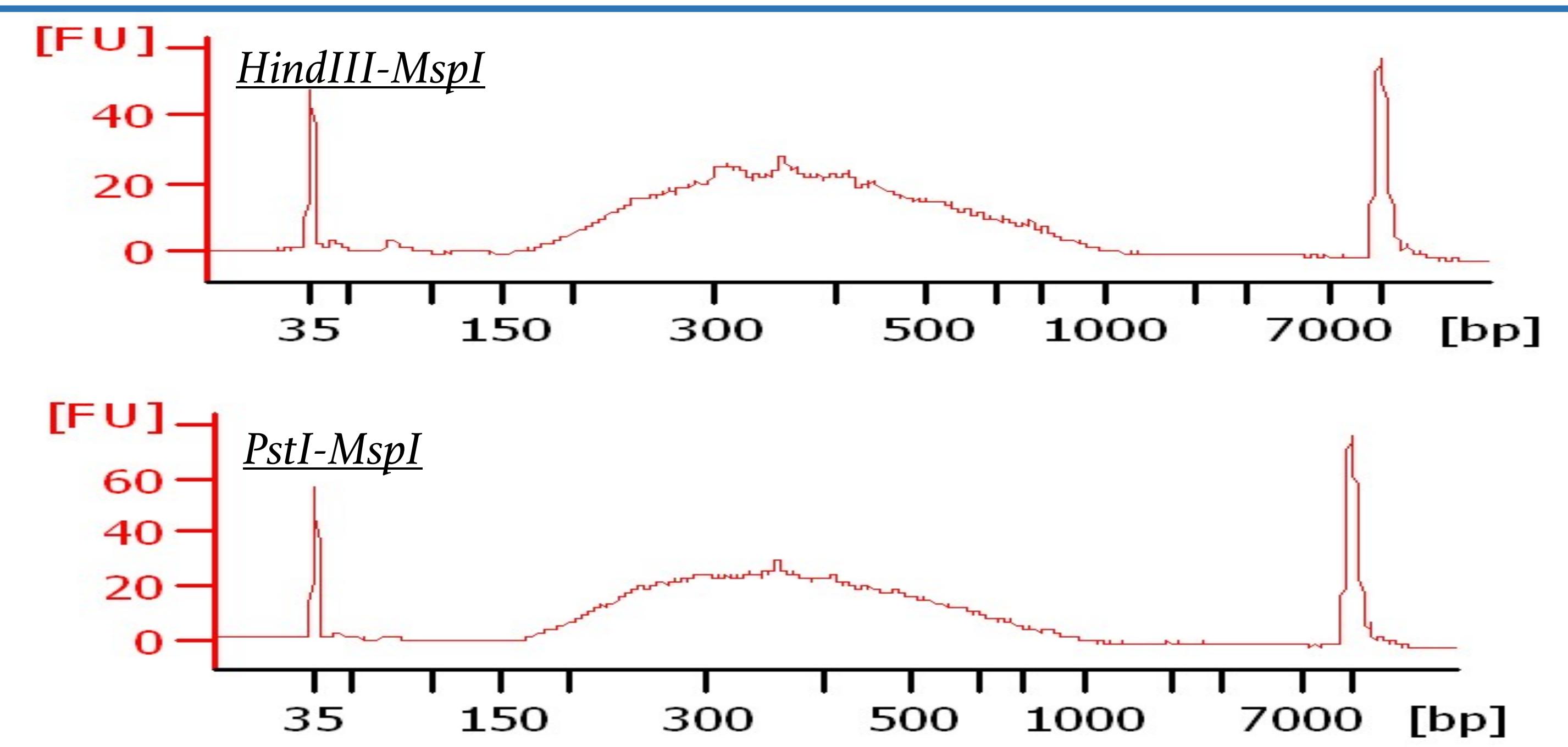


Fig 2. Fragment length distribution for individual libraries; *HindIII-MspI* & *PstI-MspI*

Category	<i>HindIII - MspI</i>	<i>PstI - MspI</i>
Total Reads	3,906,096 per sample	3,798,060 per sample
Mapped Reads	3,753,634 (96.1% of total)	3,730,571 (98.2% of total)
Uniquely Mapped Reads (MAPQ≥30)	2,023,367 (51.8% of total)	1,859,605 (48.9% of total)
Unique Sites	2,362,556	1,591,764
Raw SNPs	988,879	436,903
Filtered SNPs	415,048 (41.9% of raw)	176,626 (40.4% of raw)
Average Depth per Site	294X	507X

Table 1. Read count, variant count, and sequencing depth.

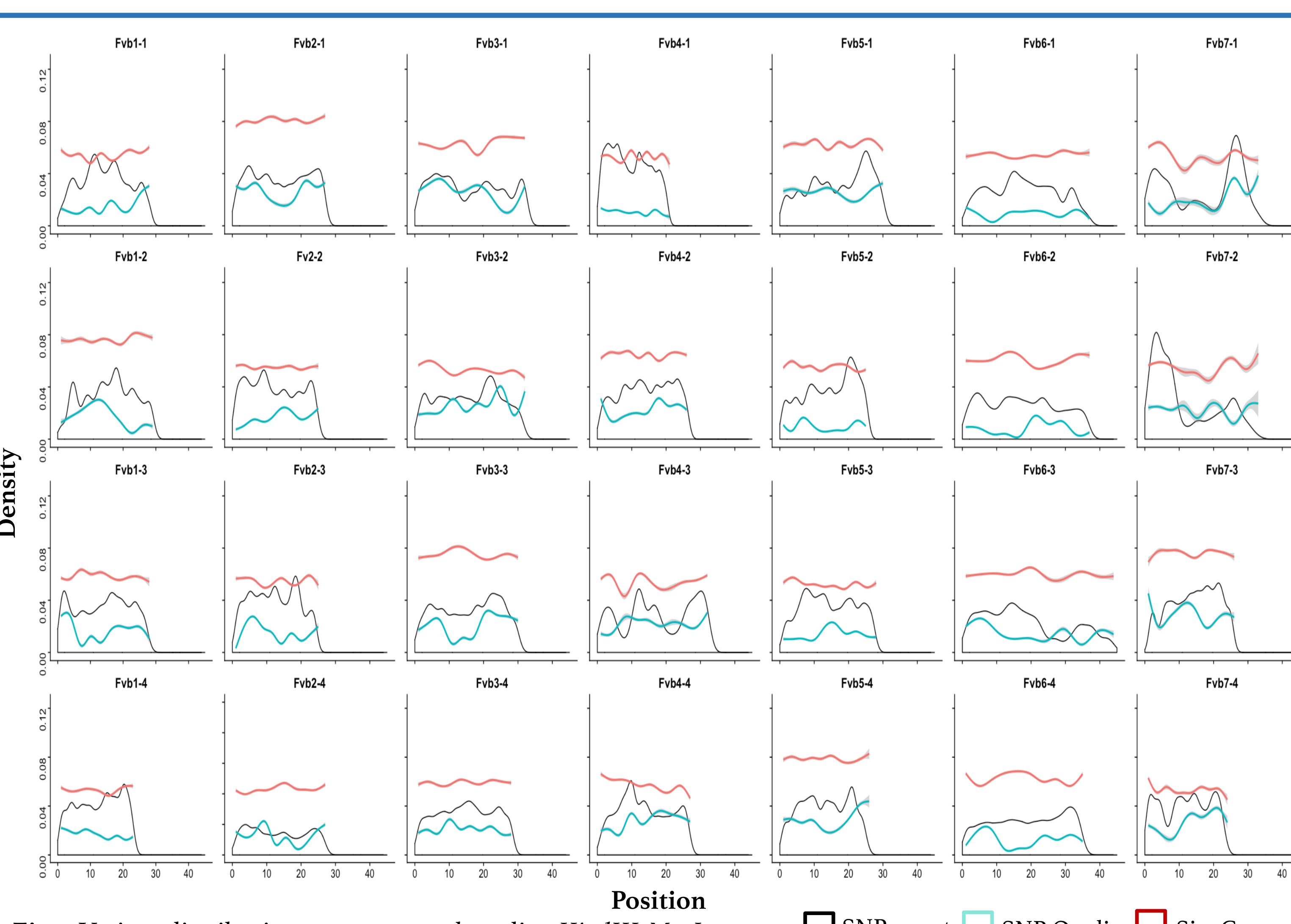


Fig 3. Variant distribution, coverage, and quality: *HindIII-MspI*

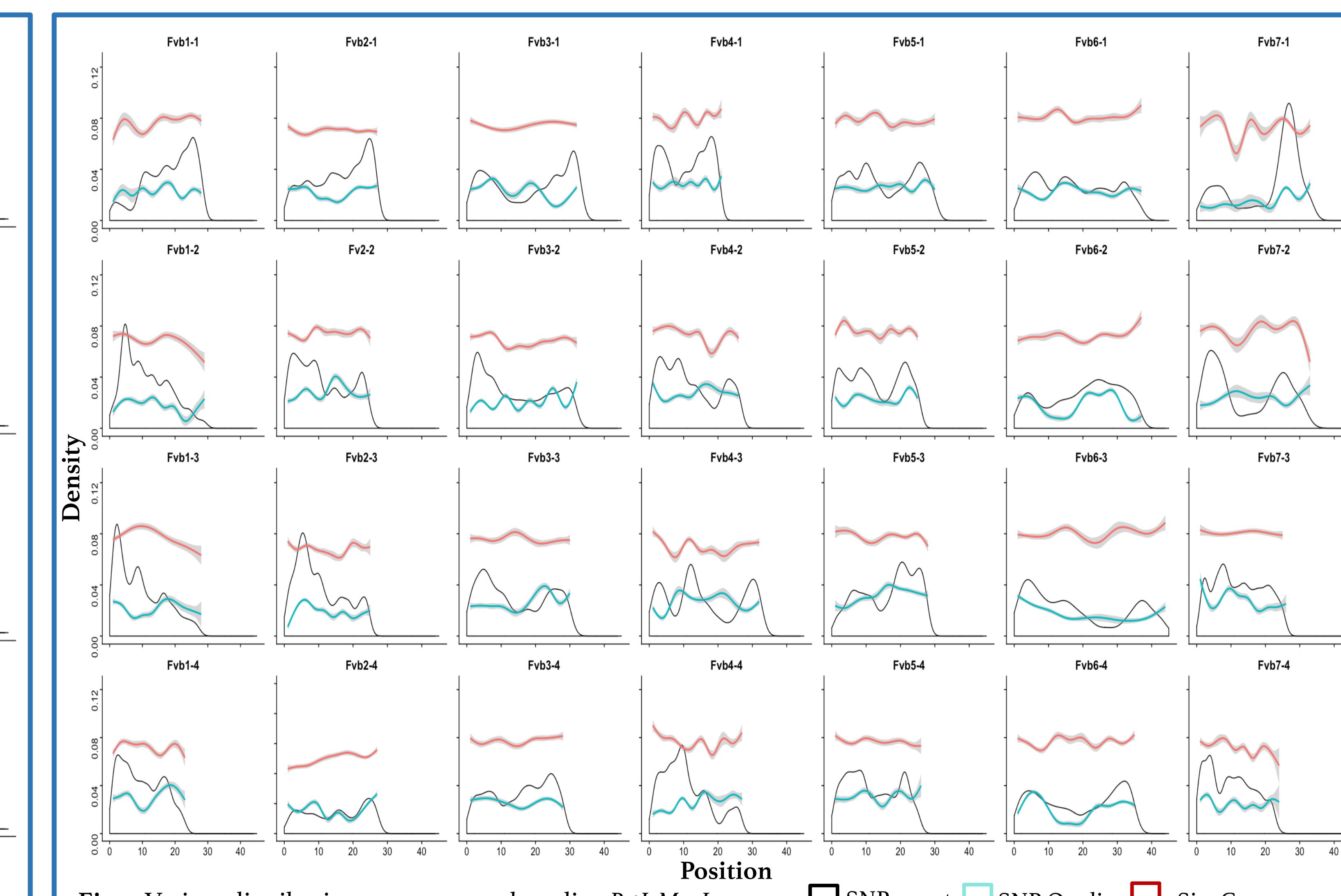


Fig 4. Variant distribution, coverage, and quality: *PstI-MspI*

Conclusions

1. Both libraries had equivalent fragment length distributions, total reads, and uniquely mapped reads per individual. However, SNP coverage per site from *HindIII-MspI* appears to be much more even across all chromosomes.
2. 8X-GBS pipeline resulted in ~175k & 415k genome-wide SNPs with sub-genome specificity for *PstI-MspI* and *HindIII-MspI*, respectively.
3. For variant discovery against in octoploid strawberry in this study, *HindIII-MspI* is the preferred enzymatic pair.