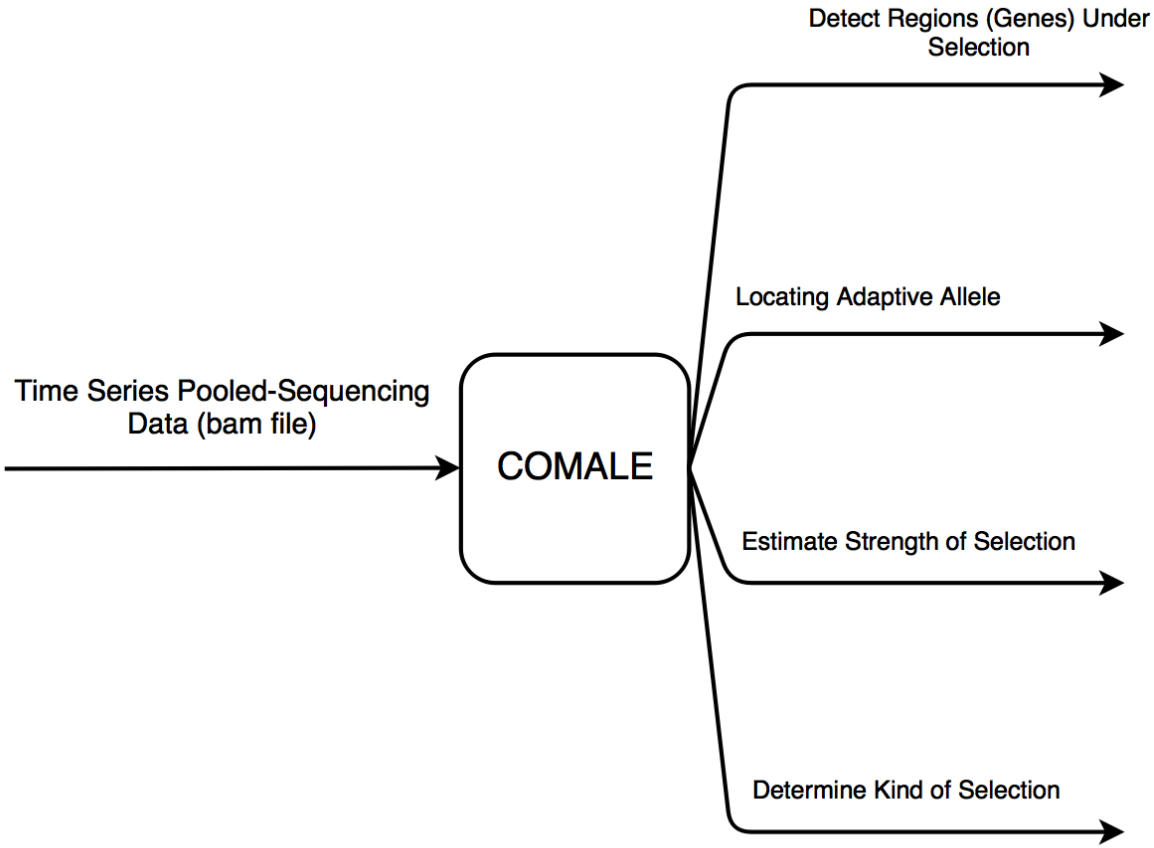
Detecting Selection in Experimental Evolution Experiment

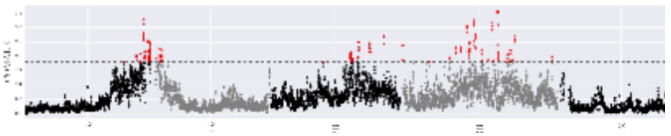
Arya Iranmehr
airanmehr@ucsd.edu

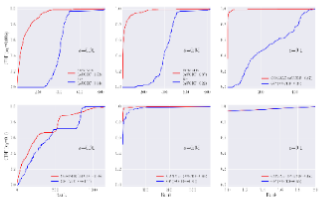
Bafna Lab
University of California, San Diego

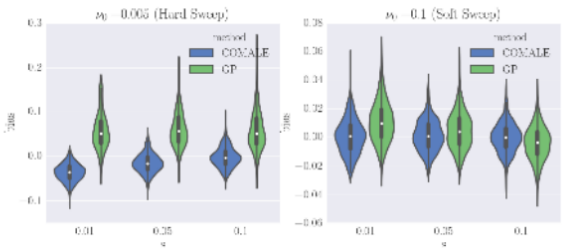
July, 2016

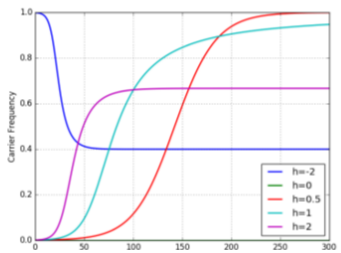
Goals of COMALE: Composite Of MArkovian Likelihoods for Experimental evolution



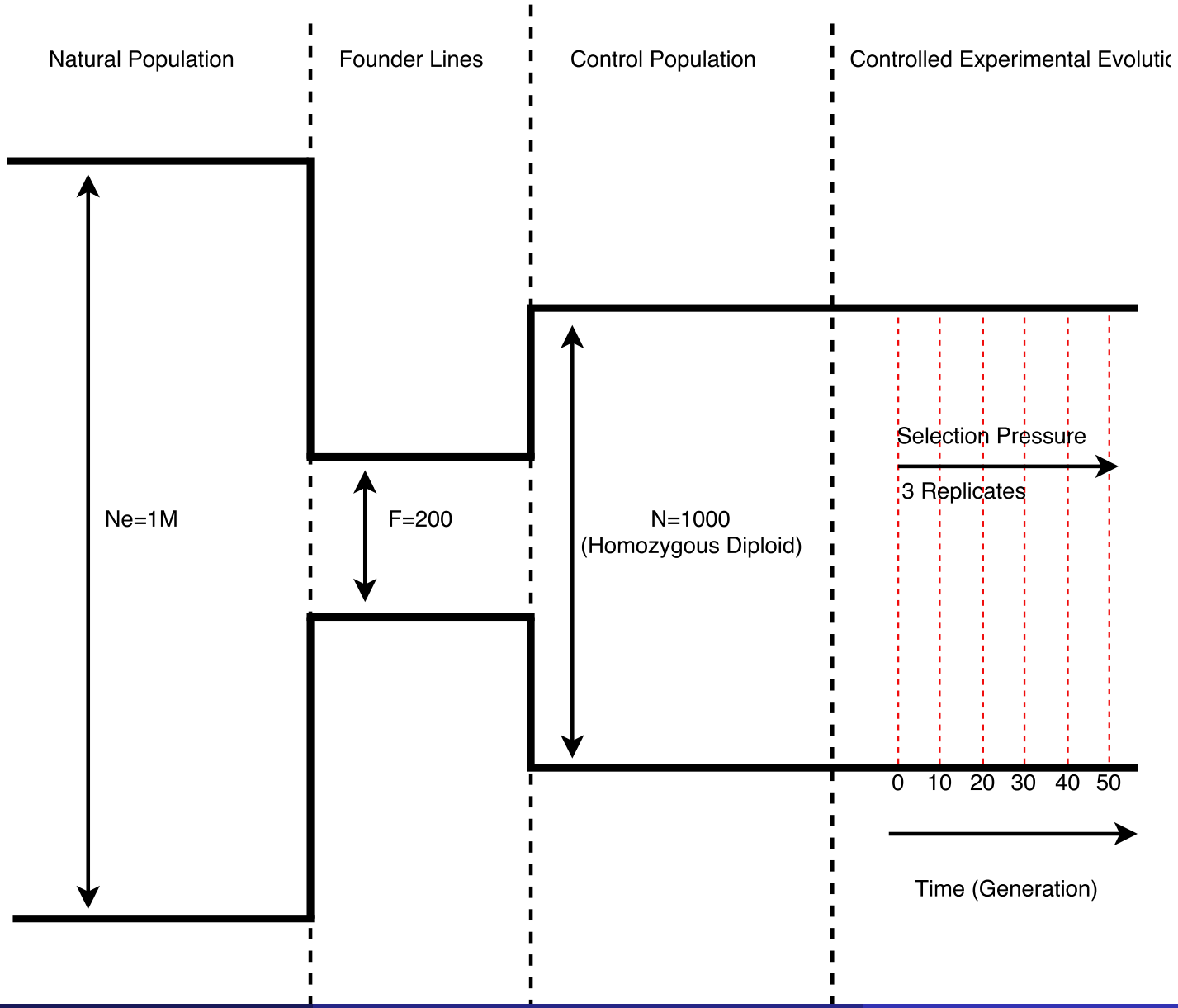








Simulations



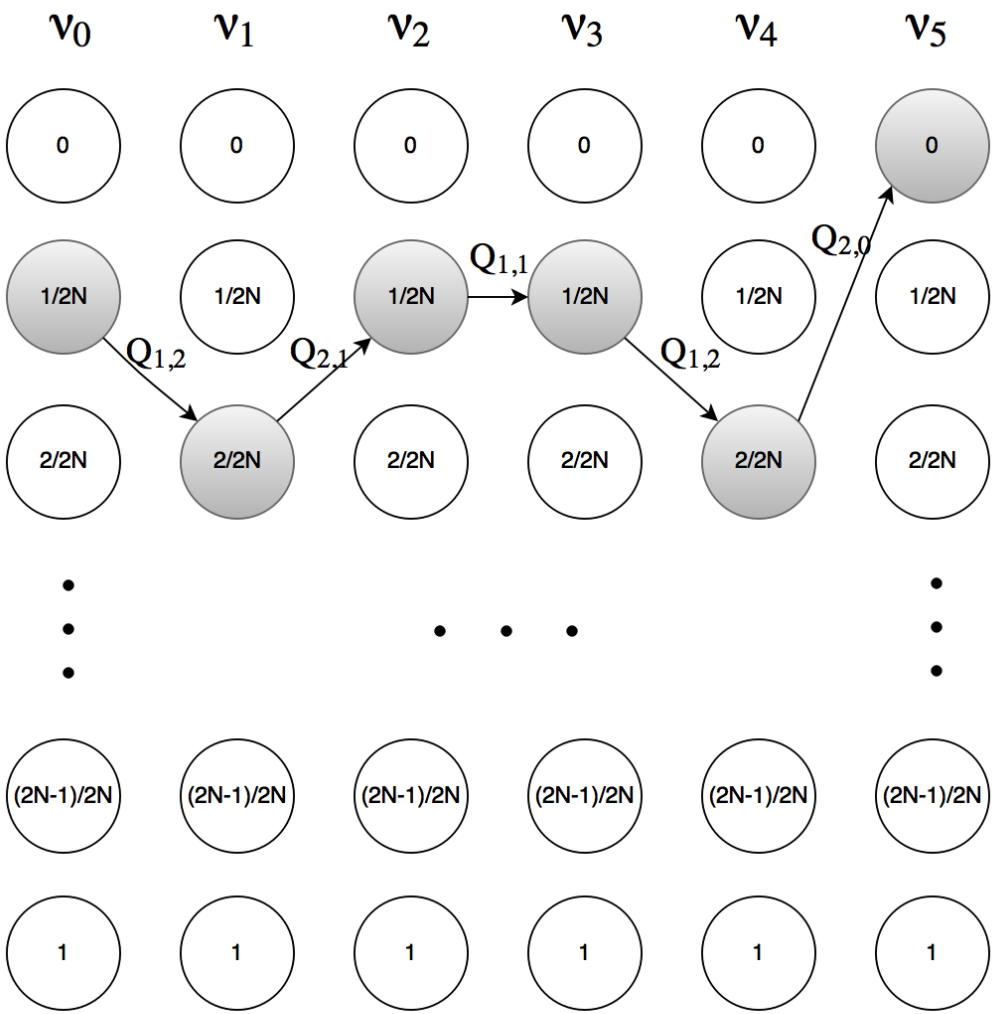
Case I: Exact Allele Frequency

- Given allele frequencies $\{\nu_0, \ldots, \nu_T\}$ at each site compute λ-statistic for each SNP by

$$\lambda = \frac{H_1}{H_0} = \frac{\Pr(\nu_0, \ldots, \nu_T | s = s^*)}{\Pr(\nu_0, \ldots, \nu_T | s = 0)} \tag{1}$$

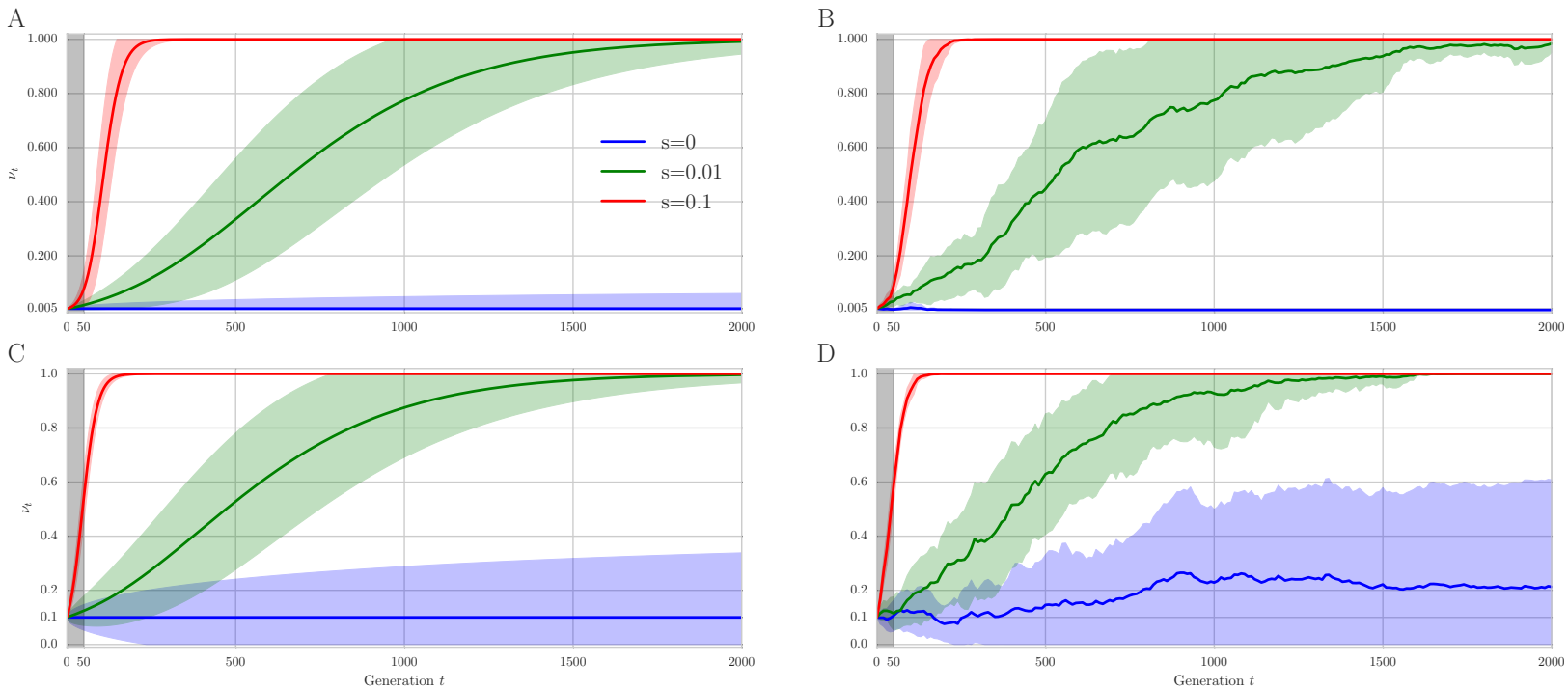
where s^* is the maximum likelihood estimate of s (strength of selection).

Probability of a Sequence: Wright-Fisher Markov Chain

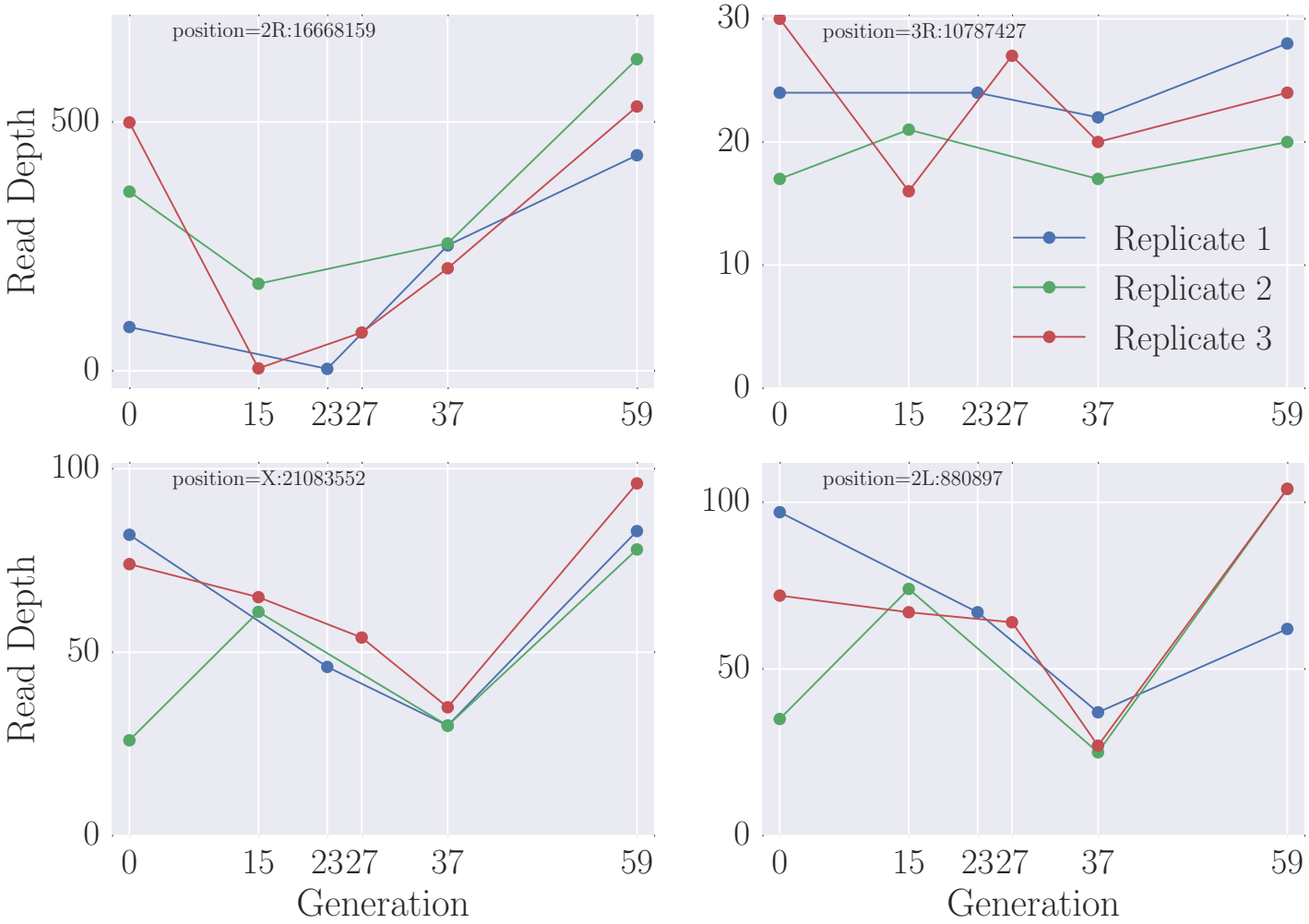


$$P(\nu_0, \ldots, \nu_5) = Q_{1,2}\ Q_{2,1}\ Q_{1,1} Q_{1,2}\ Q_{2,0}$$

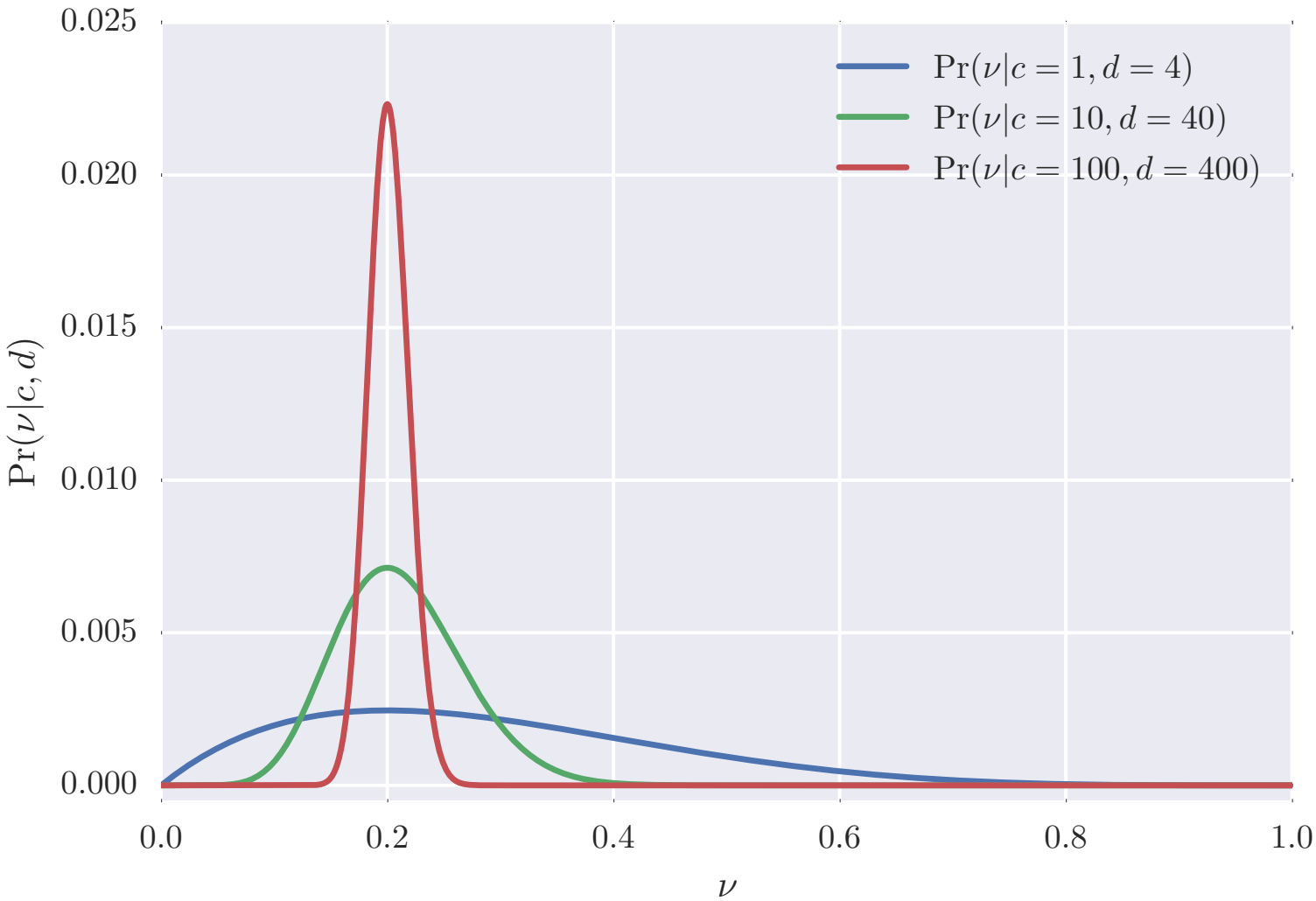
Model vs Observation



Depth Heterogeneity



Uncertainty in allele frequency



Case II: Unknown Allele Frequency (Pooled-Seq data)

- Given $\{x_0, \ldots, x_T\}$ at each site where $x_t = (c_t, d_t)$ is a tuple of derived allele read count and the read depth, we are interested in computing λ-statistic for each SNP by

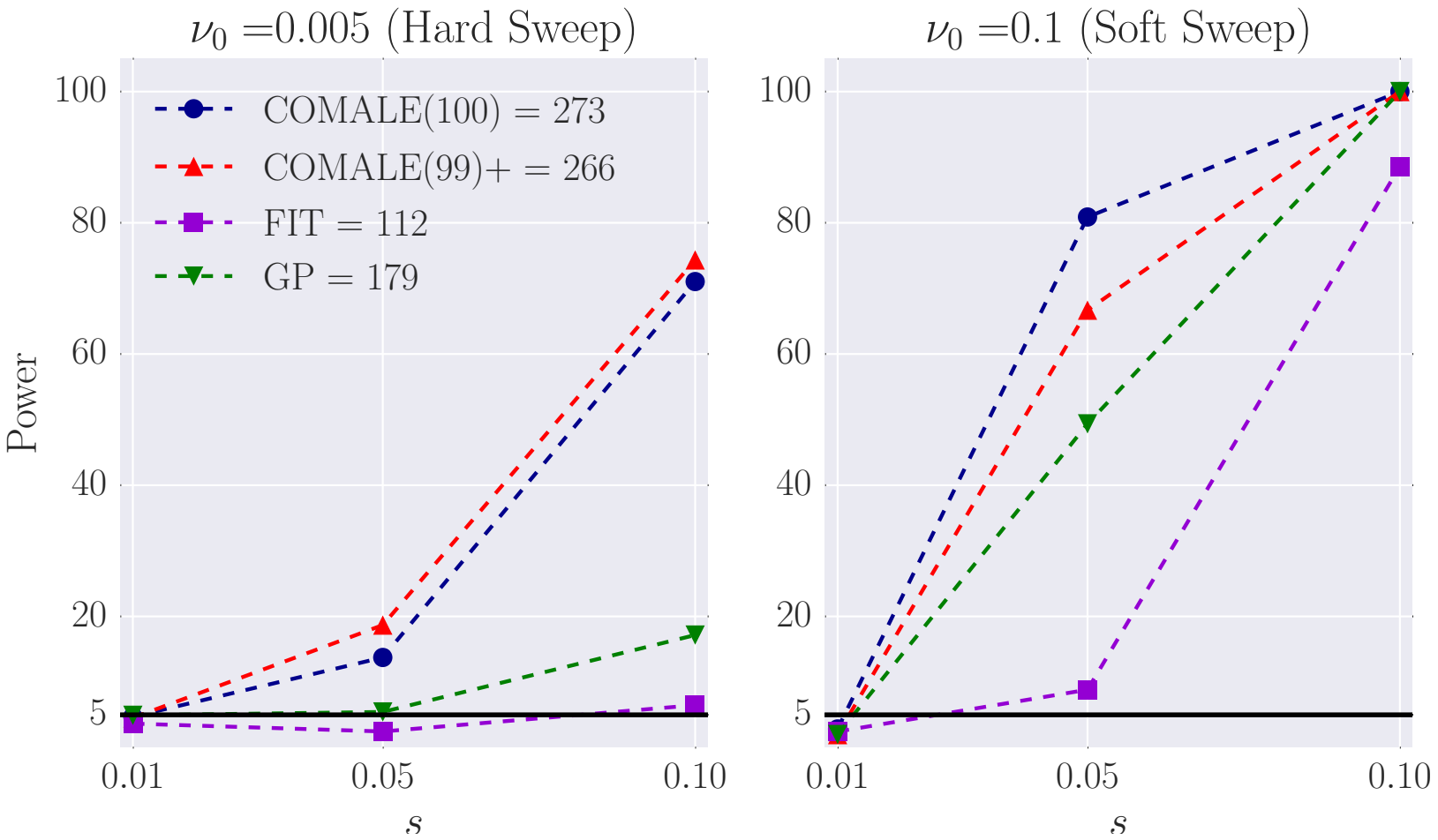
$$\lambda = \frac{H_1}{H_0} = \frac{\Pr(x_0, \ldots, x_T | s = s^*)}{\Pr(x_0, \ldots, x_T | s = 0)} \tag{2}$$

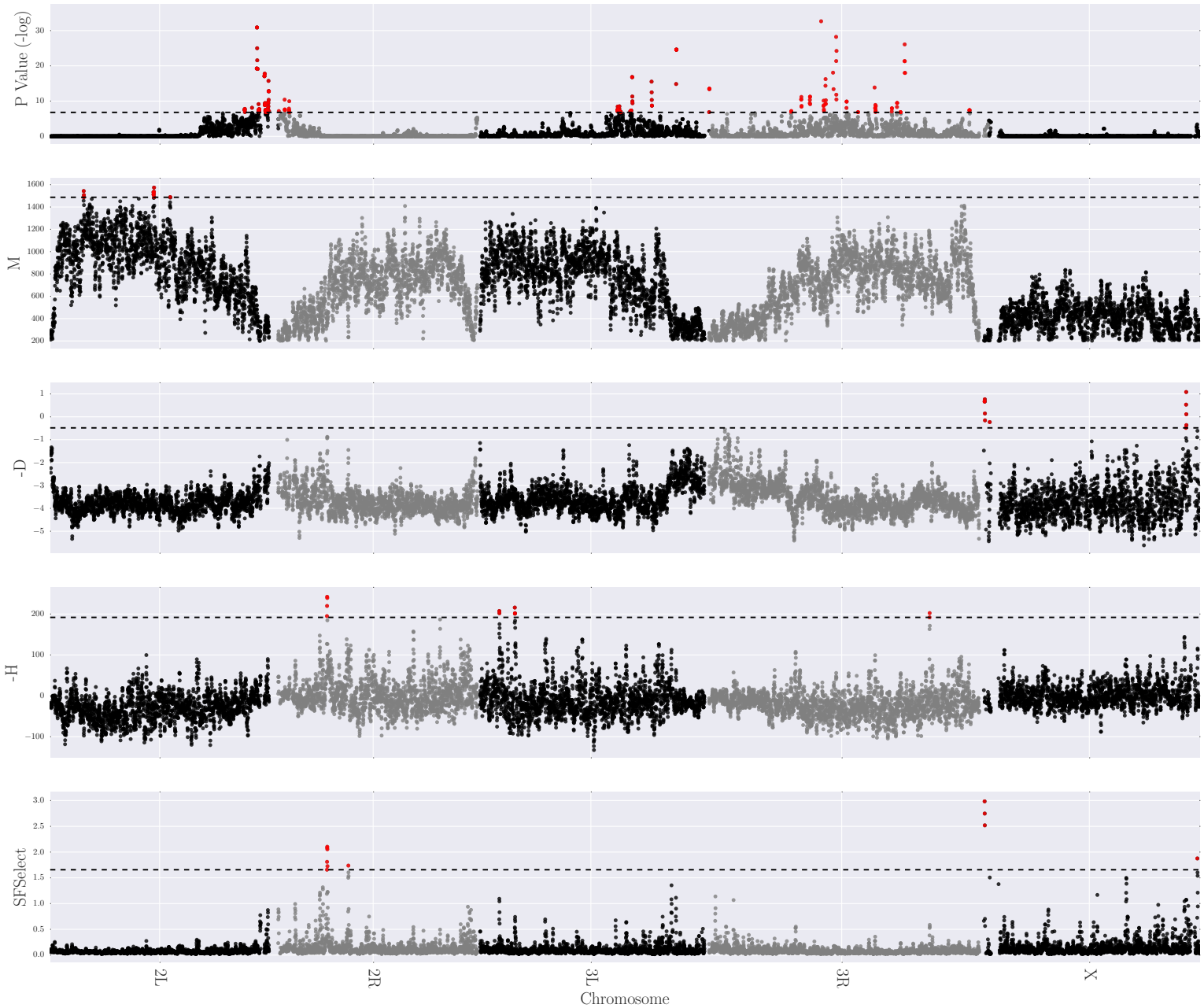
- Hidden Markov Model computes null and alternative likelihoods.

COMALE statistic and P-Value for a region

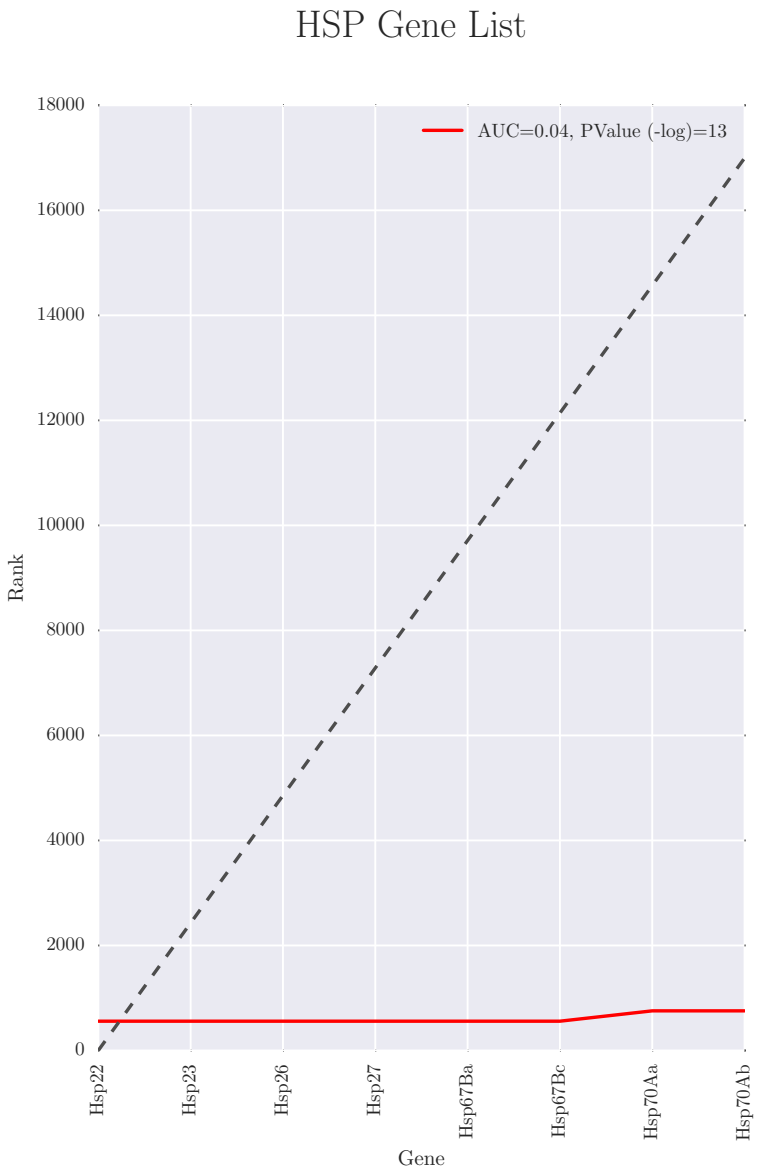
- Having computed $\mathcal{L} = \{\lambda_1, \ldots, \lambda_M\}$ for each SNP of a region, COMALE is the average of top-10 elements.
- P-Value of the COMALE statistic is computed under empirical null distribution, i.e., negative control.

Detection Power




P Value (-log)
M
-D
-H
SFSelect
2L
2R
3L
3R
X
Chromosome

Heat Protein Shock genes



Thanks!