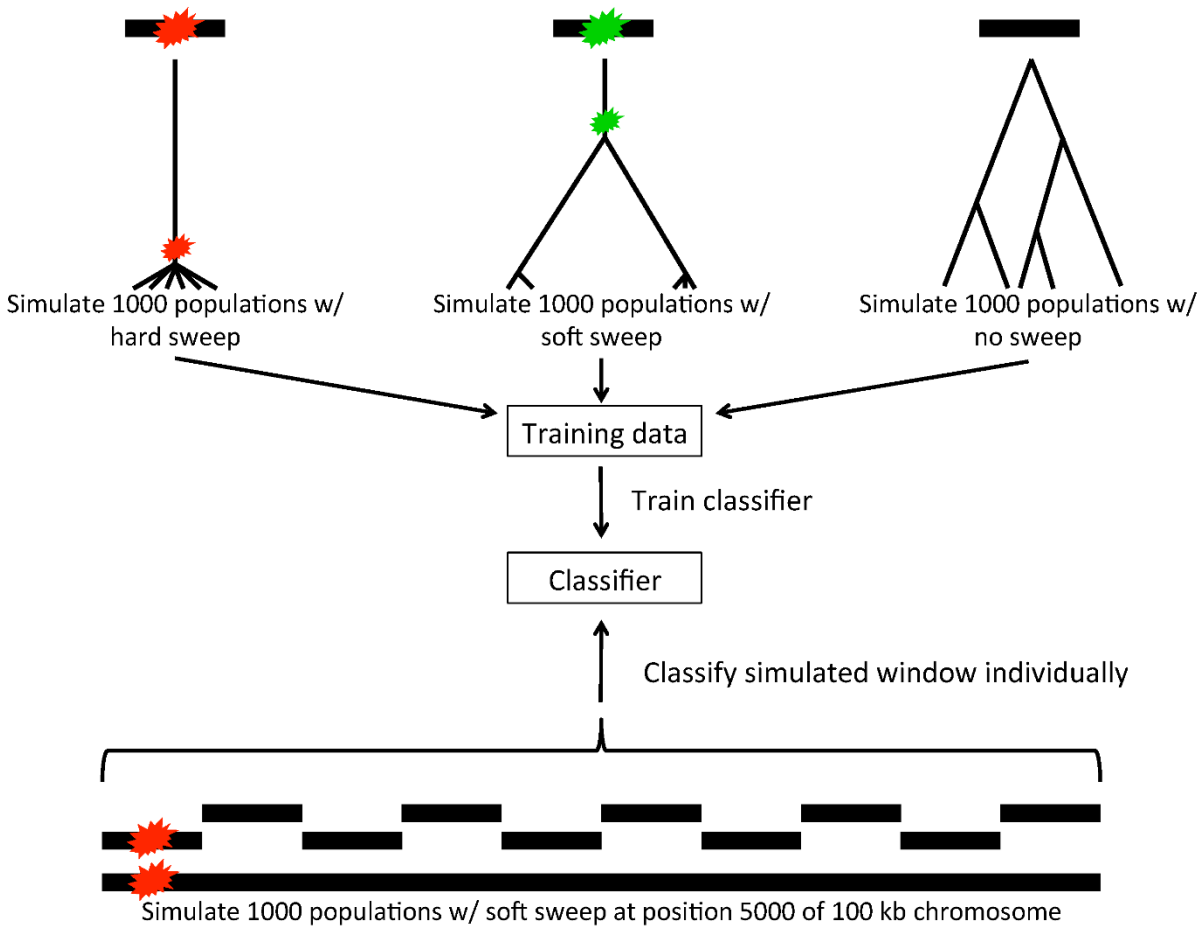# Figure S1

Figure S1   Strategy for classifying genomic windows as hard sweeps, soft sweeps, or evolving neutrally. Diagram of evolutionary scenarios of simulated 10 kb chromosomes used to train the classifier, and 100 kb chromosomes which are segmented into 10 kb windows (the first of which contains a hard sweep) to which the classifier was applied. Example genealogies of each evolutionary scenario are shown, as well as the time at which mutations that result in a sweep (hard or soft) occur. Mutations that begin sweeping to fixation immediately upon occurrence are denoted by a red explosion, while mutations that are initially fitness-neutral but later sweep to fixation are denoted by a green explosion.
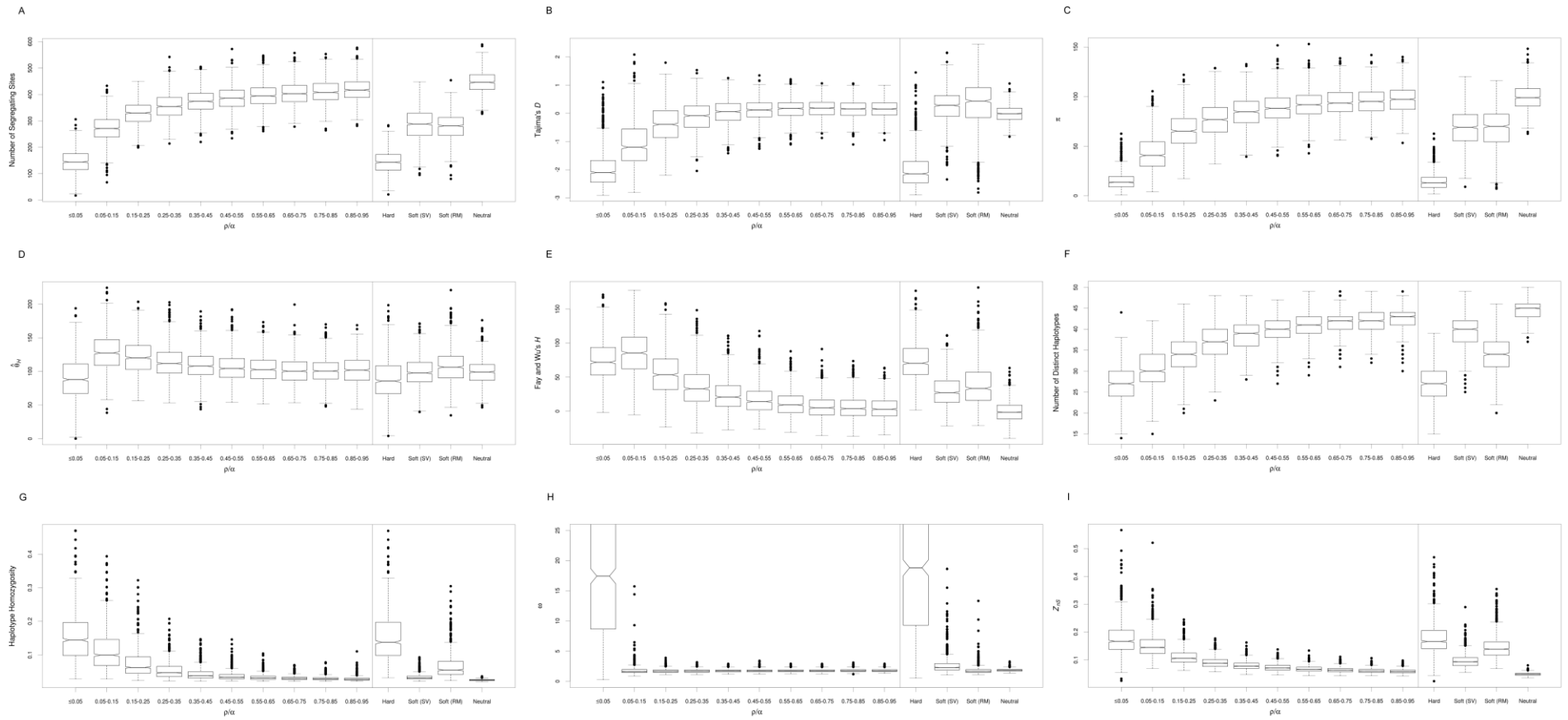
D. R. Schrider *et al.*

Figure S2   Distributions of additional summary statistics in neutral, selected, and linked regions.
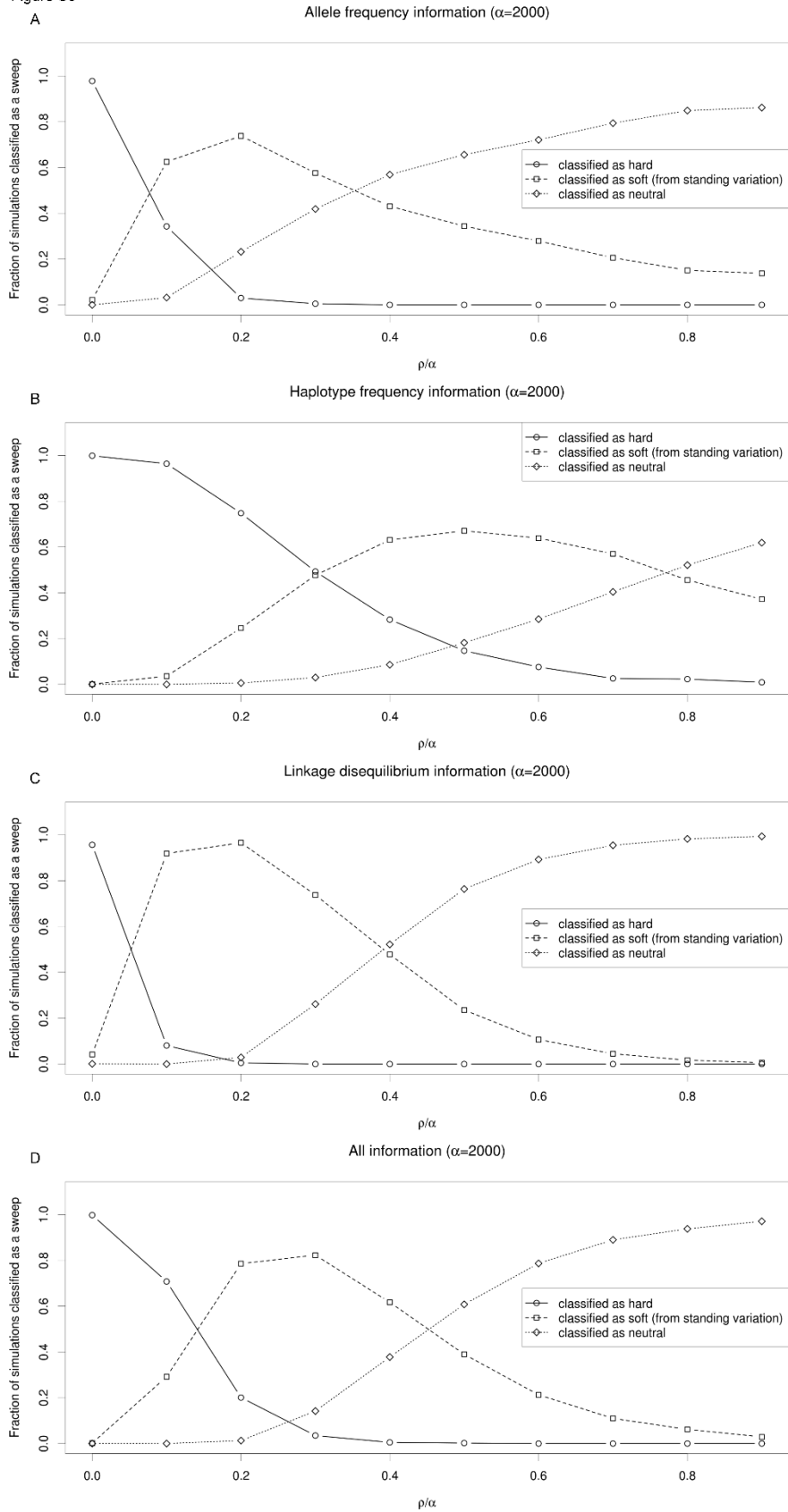
Figure S3

Figure S3   **Classification of regions flanking hard sweeps with α=2000.** This is the same plot as Figure 1 but with both *α* and *ρ* set to 2000.
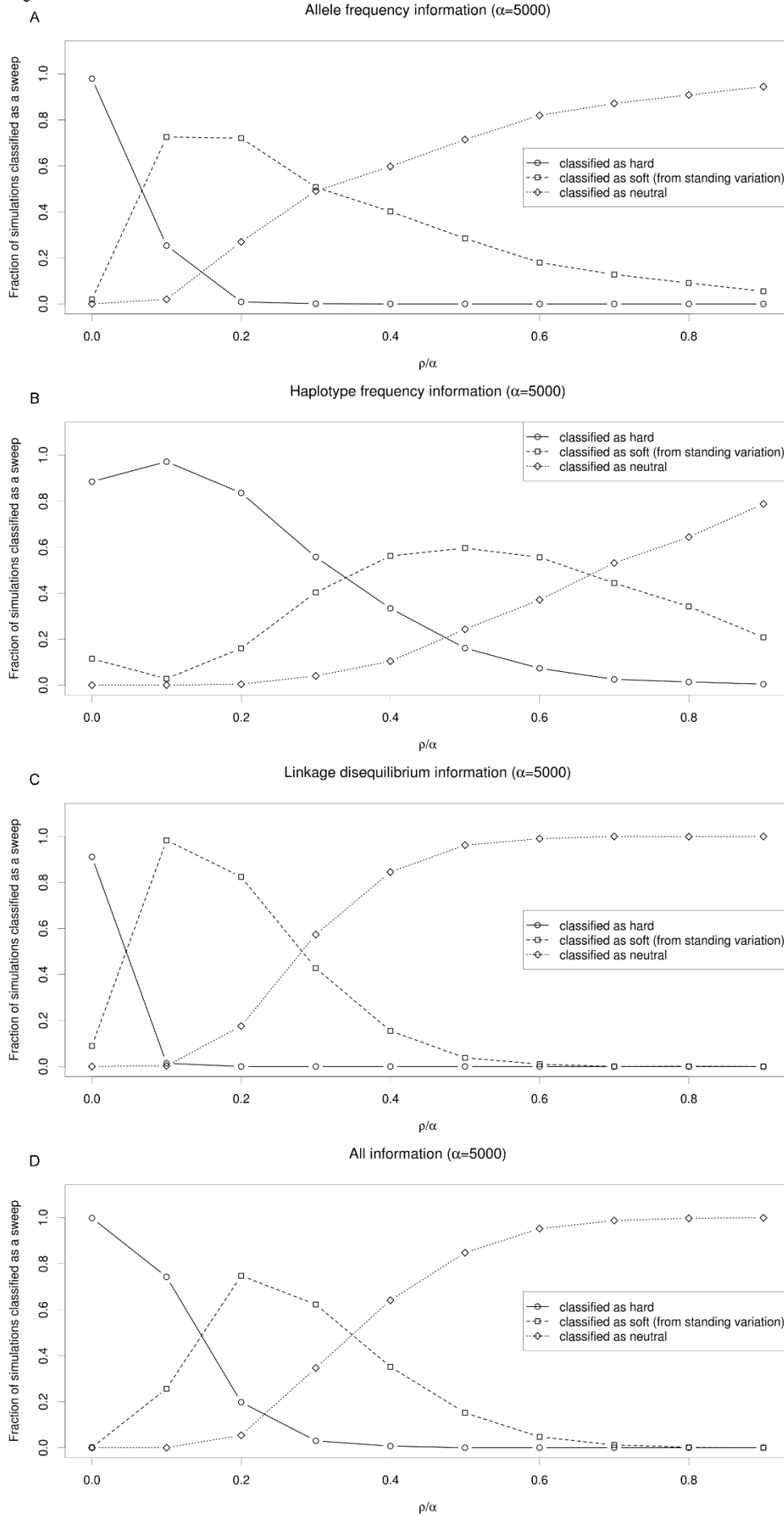
D. R. Schrider *et al.*

Figure S4

A

Allele frequency information (α=5000)

B

Haplotype frequency information (α=5000)

C

Linkage disequilibrium information (α=5000)

D

All information (α=5000)

Figure S4   Classification of regions flanking hard sweeps with *α*=5000. This is the same plot as Figure 1 but with both α and ρ set to 5000.

D. R. Schrider *et al.*

Figure S5

A
Allele frequency information (α=1000)

B
Haplotype frequency information (α=1000)

C
Linkage disequilibrium information (α=1000)
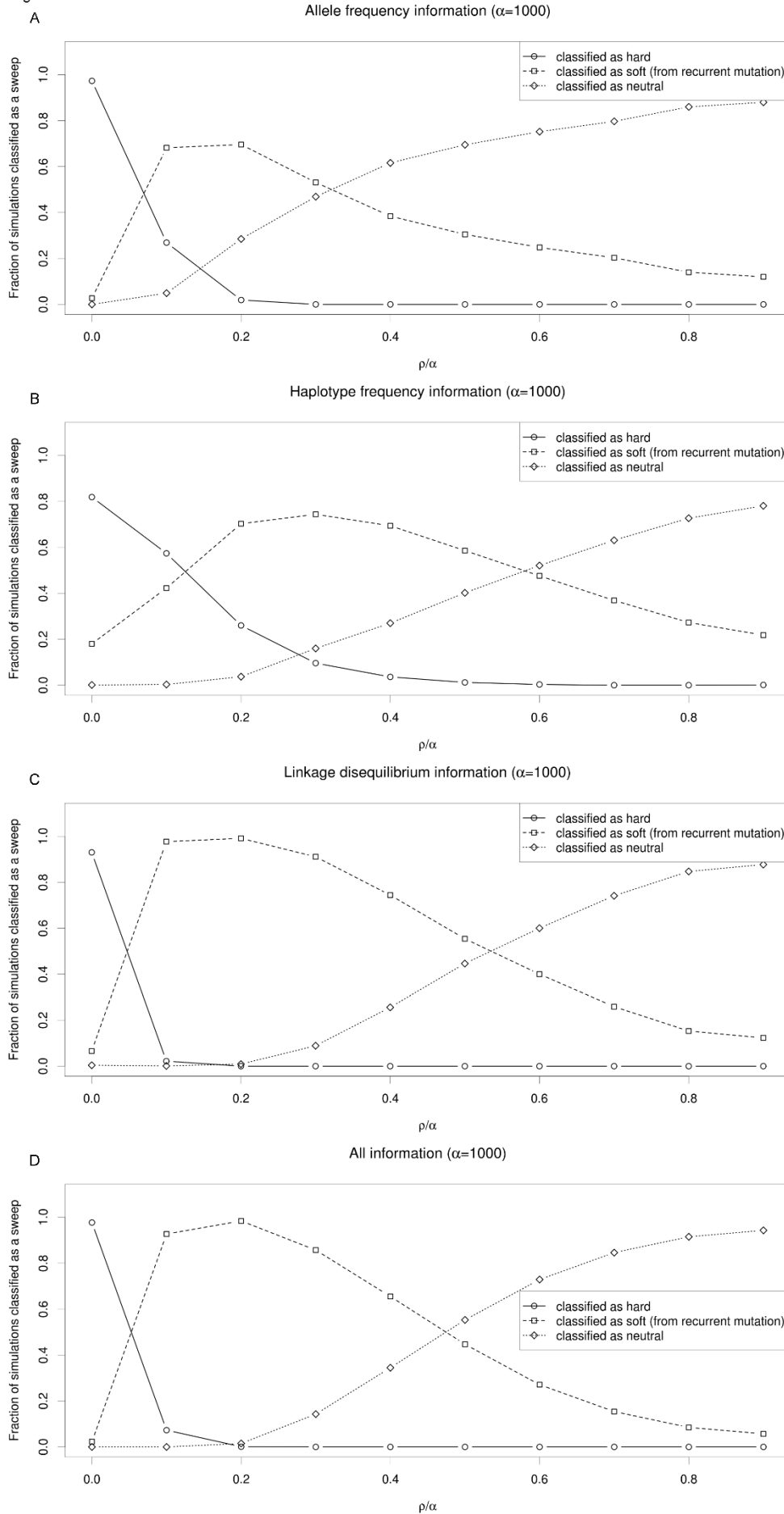

D
All information (α=1000)

Figure S5 Classification of regions flanking hard sweeps with *α*=1000, but considering recurrent mutation. This is the same plot as Figure 1 but the model of soft sweeps considered is one where the adaptive mutation can reoccur on multiple genetic backgrounds during the sweep.
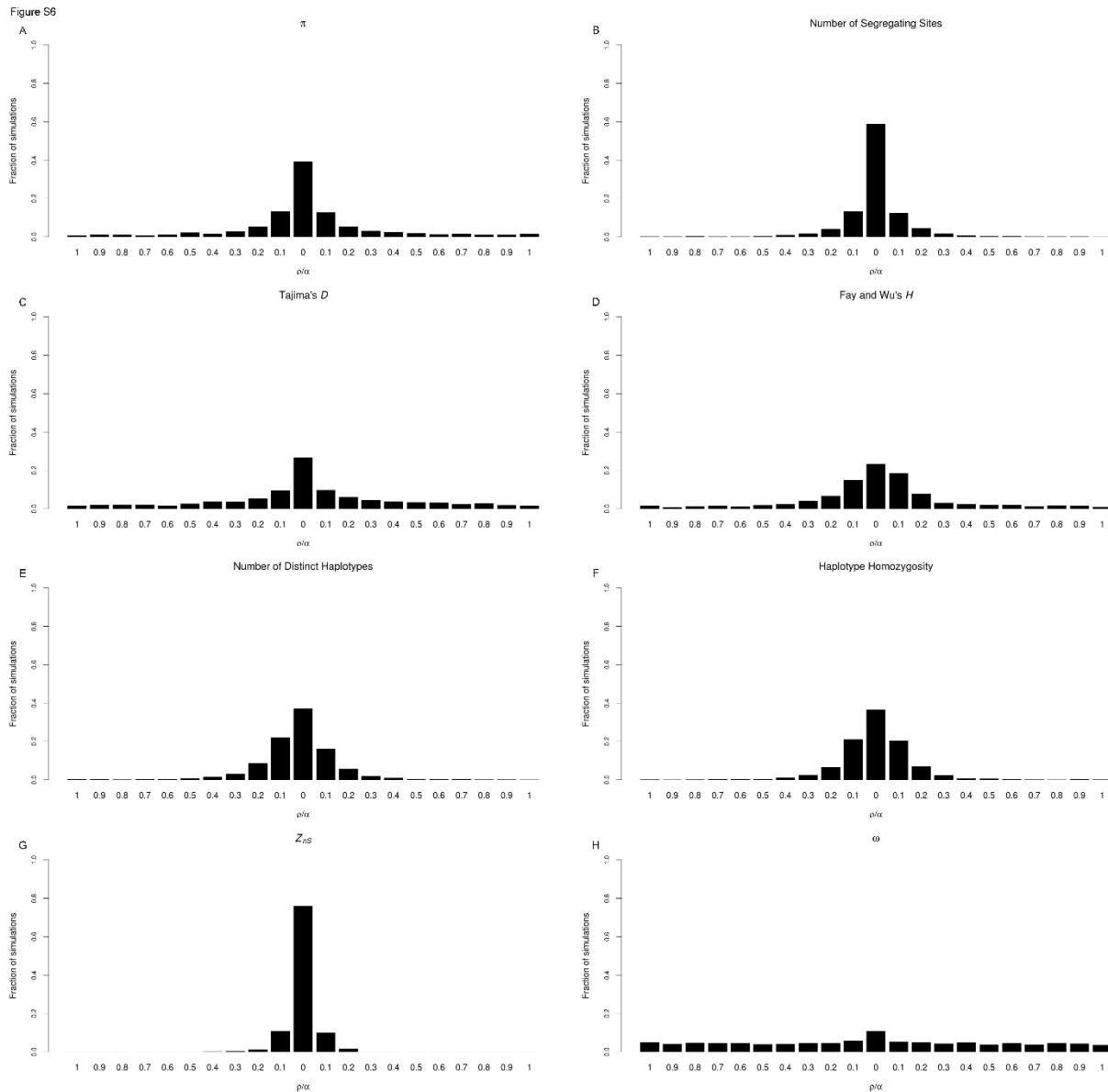
D. R. Schrider *et al.*

Figure S6 Signals of selection from various summary statistics in windows containing or flanking a soft sweep from recurrent mutation. For each summary statistic, we examined each individual simulation and located the window exhibiting the most extreme value (in the direction suggestive of a soft sweep). This figure shows the histogram of these locations for each statistic. The total genetic distance of each simulated chromosome (ρ) was 2100. The chromosome was subdivided into 21 equally sized windows (ρ=100) with a soft selective sweep (with α=1000; mutation rate to the adaptive allele ranging from 1 to 2.5) occurring in the central window.
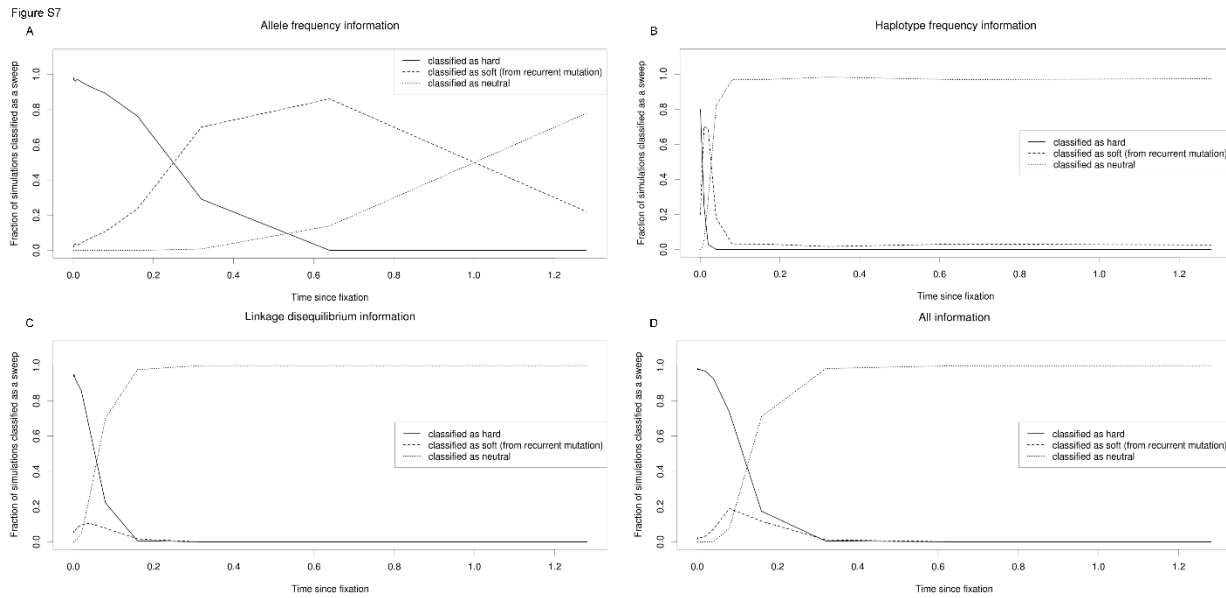
Figure S7 The relationship between age of the sweep and misclassification rate when modeling soft sweeps from recurrent adaptive mutation.

(A) The fraction of simulated windows containing a hard sweep (α=1000) classified as hard, soft (from recurrent mutation), or neutral by an SVM leveraging allele frequency information is shown according to the time in the past at which the sweep completed (in units of 2*N* generations). The most recent sweep examined in this plot completed 0.000625×2*N* generations ago, and we examined older sweeps by continually doubling the time since fixation, stopping at a sweep time of 1.28×2*N* generations in the past. (B) Same as panel A, but using a classifier leveraging haplotype information. (C) Results from an SVM leveraging LD information. (D) An SVM leveraging our full set of summary statistics (Methods).
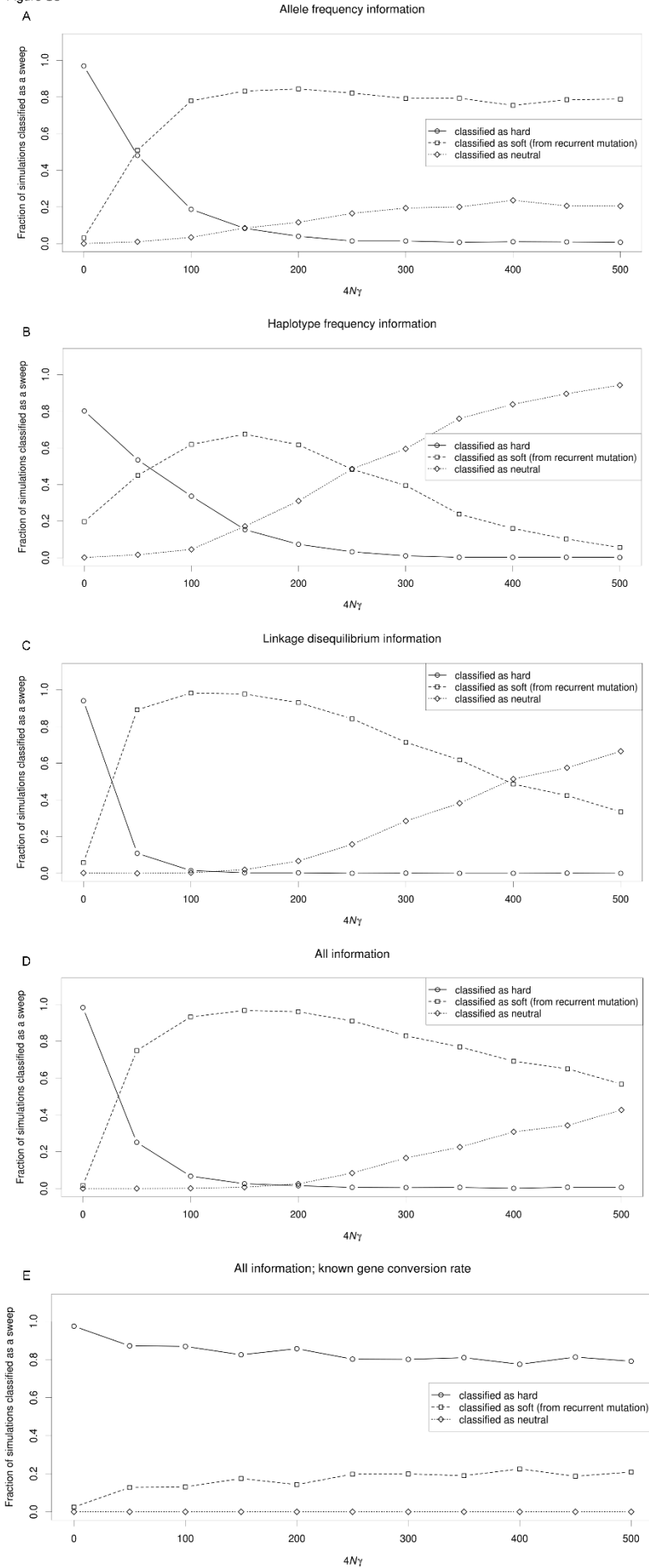
D. R. Schrider *et al.*

Figure S8

Figure S8   The relationship between gene conversion rate and sweep misclassification rate when modeling soft sweeps from recurrent adaptive mutation. (A) The relationship between the locus-wide gene-conversion rate and the fraction of simulated windows containing a hard sweep (α=1000) classified as hard, soft (from recurrent mutation), or neutral. (B) Same as panel A, but using a classifier leveraging haplotype information. (C) Results from an SVM leveraging LD information. (D) An SVM leveraging our full set of summary statistics (Methods). (E) An SVM leveraging our full set of summary statistics and trained from simulated regions experiencing the correct gene conversion rate.

Tables S1-S2

Available for download as Excel files at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.115.174912/-/DC1

Table S1. Summary of simulation datasets used in this study.

Table S2. Classification accuracy of each SVM as assessed on an independent test set.