

Supporting material: Linkage disequilibrium under genetic hitchhiking in finite populations

P. Pfaffelhuber*, A. Lehnert†, W. Stephan†

February 24, 2008

Numerical comparison of $\mathbb{E}[\widehat{r^2}]$ and $\widehat{\sigma_D^2}$

In Hudson (1985), a heuristics concerning the connection between $\mathbb{E}[\widehat{r^2}]$ and $\widehat{\sigma_D^2}$ was established under neutrality: if we ignore variants which occur in low frequency, both measures should be approximately the same. In Figure 1 this connection is established using numerical results for the case of a population at the end of a selective sweep. We see that $\mathbb{E}[\widehat{r^2}]$ and $\widehat{\sigma_D^2}$ show the same pattern.

Comparison of simulations and analytical results for various α

Complementing Figure 8 of the main text we simulated selective sweeps for several values of the selection strength α . We show comparison results to (10) and (12) in the main text for 500, 1000 and 2000 in Figures 2, 3 (same as Figure 8 in the main text) and 4, respectively. In Subfigures (A), we fix $\rho_{LR}/\alpha = 0.05$ while in Subfigures (B), we always have $\rho_{LR}/\alpha = 0.5$.

*University of Vienna, Department of Mathematics, Nordbergstraße 15, 1090 Vienna, Austria

†Ludwig-Maximilians University, Biocenter, Großhaderner Straße 2, 82152 Planegg, Germany.

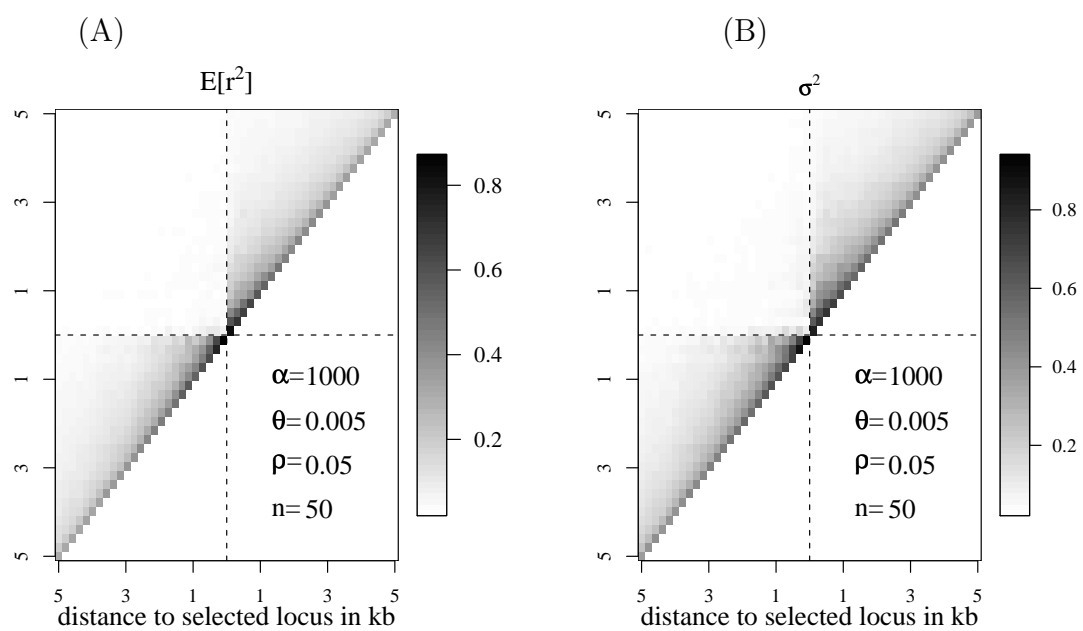


Figure 1: The measures (A) $\mathbb{E}[r^2]$ and (B) σ^2 at the end of a selective sweep in a sample of size $n = 50$. In both figures, variants which occur in less than 5% were ignored.

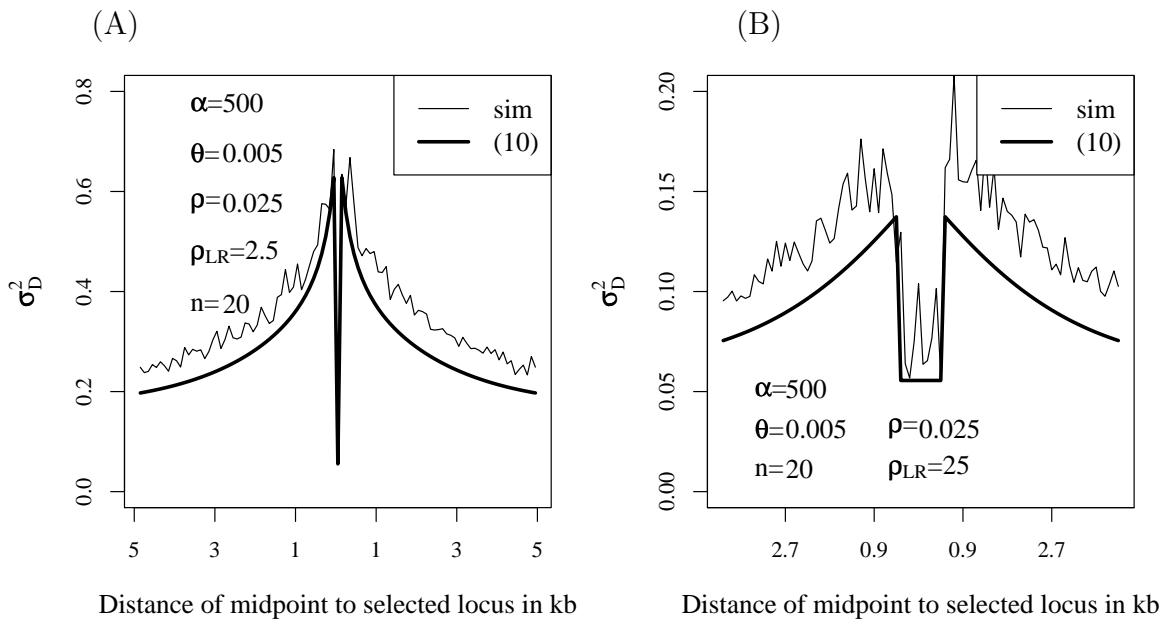


Figure 2: Comparison of simulations and prediction using the star-like approximation from (10) and (12) for $\alpha = 500$. The neutral loci in the simulation fall in windows which are (A) 0.2 kb (B) 2 kb apart. Every curve is based on 10^3 simulations.

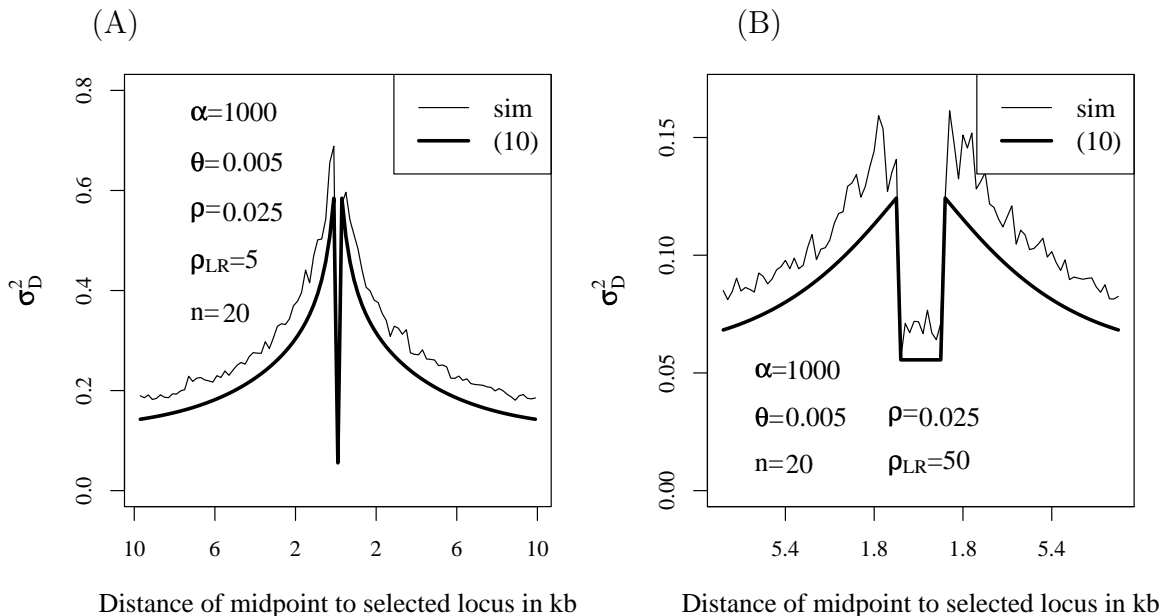


Figure 3: Same as Figure 8 in the main text

Comparison of simulations and analytical results for various sample sizes

Complementing Figure 8 of the main text we simulated selective sweeps for various sample sizes n . We show comparison results to (10) and (12) in the main text for $n = 10, 20, 50$ in Figures 5, 6 (same as Figure 8 in the main text) and 7, respectively.

The decay of linkage disequilibrium after the sweep

After the selective sweep, linkage disequilibrium decays rapidly and reaches neutral levels. In Figure 8 we present simulation results complementing Figure 9 of the main text. We see that linkage disequilibrium decays fastest directly after the selective sweep.

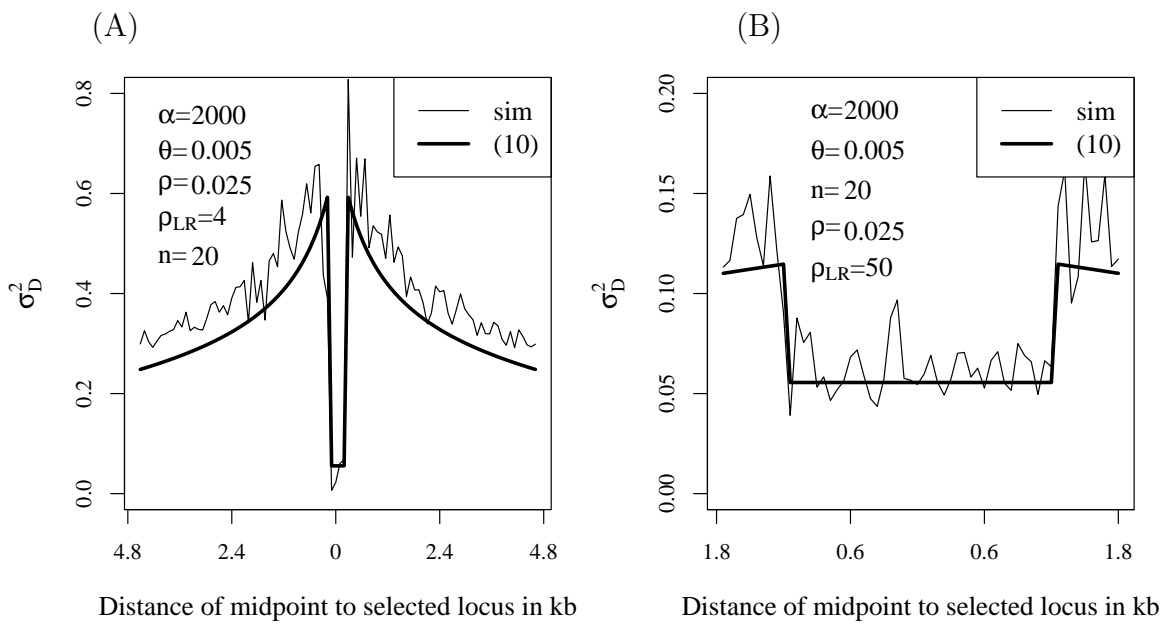


Figure 4: Comparison of simulations and prediction using the star-like approximation from (10) and (12) for $\alpha = 2000$. The neutral loci in the simulation fall in windows which are (A) 0.4 kb (B) 4 kb apart. Every curve is based on 10^3 simulations.

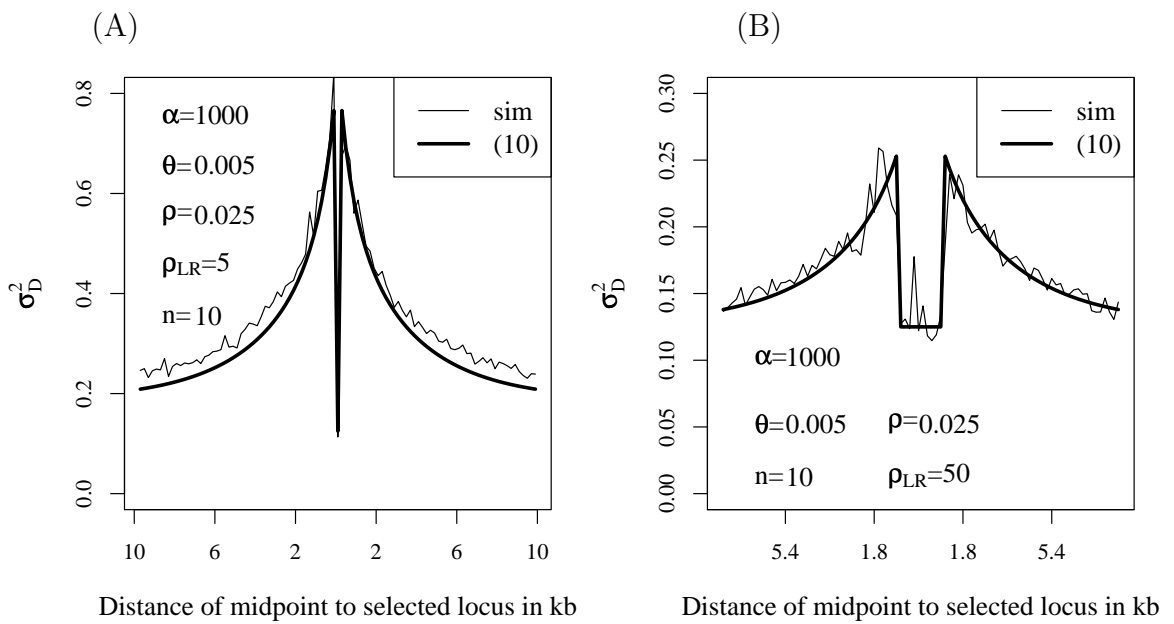


Figure 5: Comparison of simulations and prediction using the star-like approximation from (10) and (12) for $n = 10$. The neutral loci in the simulation fall in windows which are (A) 0.2 kb (B) 2 kb apart. Every curve is based on 10^3 simulations.

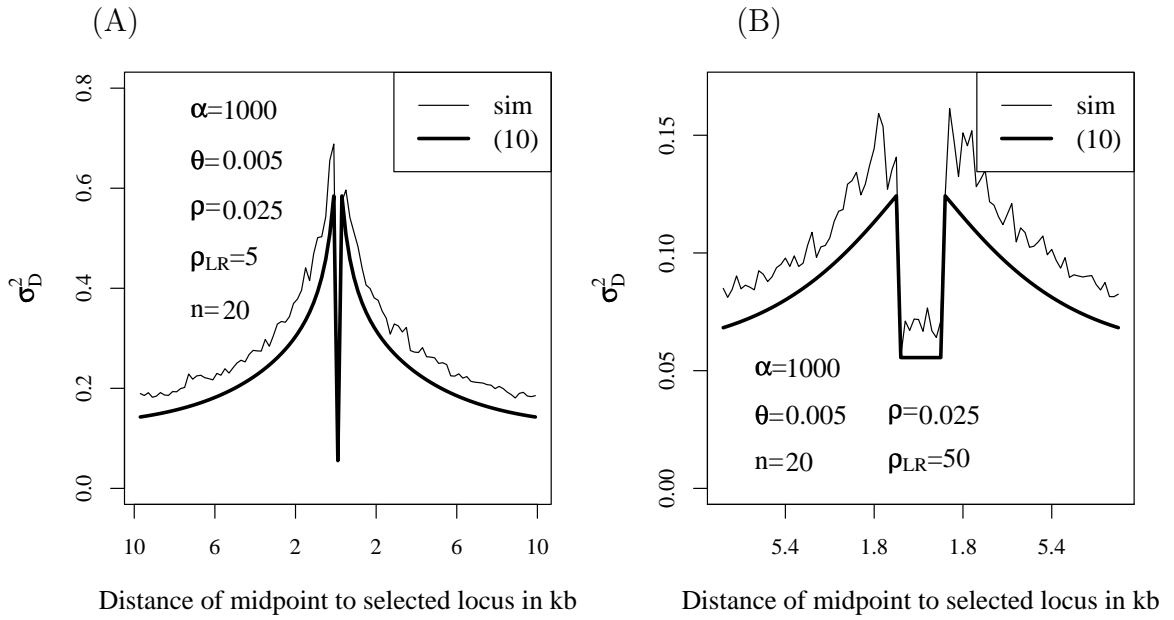


Figure 6: Same as Figure 8 in the main text, i.e., same as Figure 5 for $n = 20$.

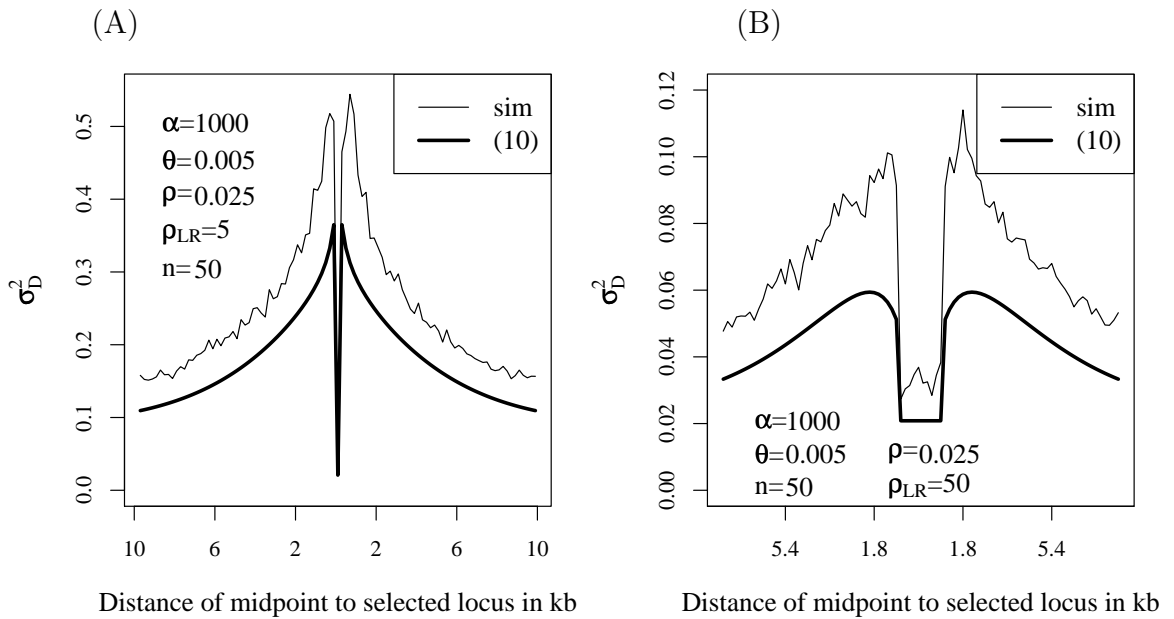


Figure 7: Same as Figure 5 for $n = 50$.

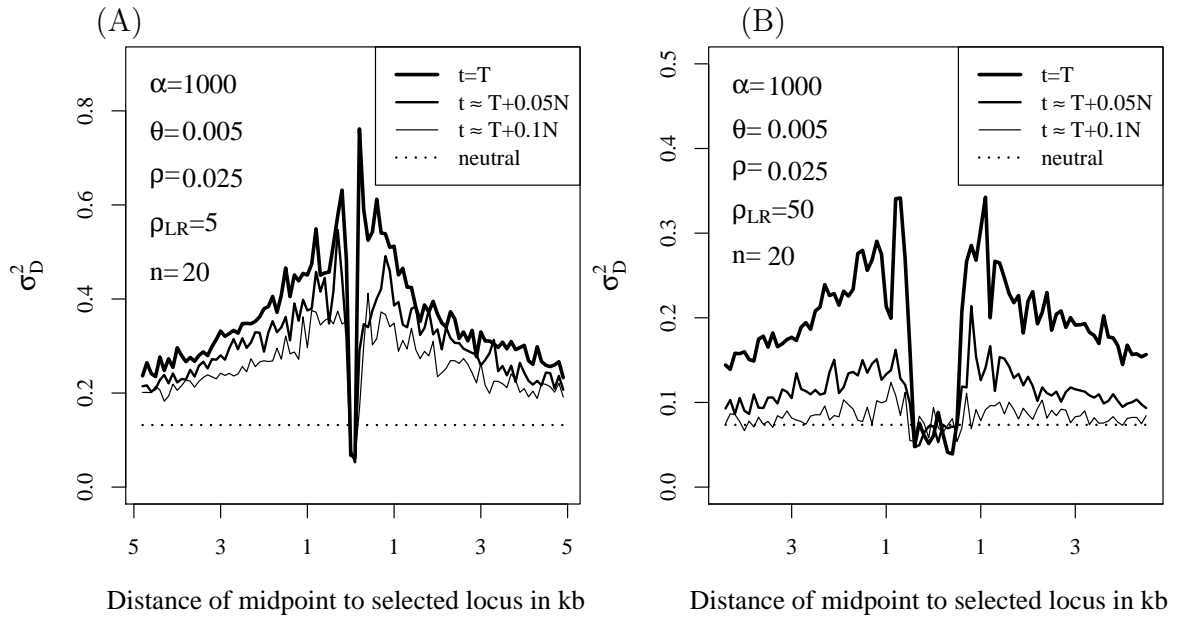


Figure 8: The pattern of linkage disequilibrium at different time-points after the selective sweep. The $t = T$ -curve gives standardized linkage disequilibrium at the time of fixation of the beneficial allele. The $t = T + 0.05N$ - and $t = T + 0.1N$ -curves describe the pattern at the time $0.05N$ and $0.1N$ generations afterwards, respectively. Both neutral loci are (A) 0.2 kb (B) 2 kb apart. Every curve is based on 10^3 simulations of the 10-kb fragment.

References

HUDSON, R. R., 1985 The sampling distribution of linkage disequilibrium under an infinite allele model without selection. *Genetics* **109**: 611–631.