Exact Distribution of Linkage Disequilibrium in the Presence of Mutation, Selection or Minor Allele Frequency Filtering

Jiayi Qu, Stephen D Kachman, Dorian Garrick, Rohan L Fernando, and Hao Cheng

Background

Linkage disequilibrium (LD), often expressed in terms of the squared correlation (r^2) between allelic values at two loci, is an important concept in many branches of genetics and genomics.

Genetic drift and recombination have opposite effects on LD, and thus r^2 will keep changing over generations until the effects of these two forces are counterbalanced. Several approximations have been used to determine the expected value of r^2 at equilibrium in the presence or absence of mutation. One of the most famous approximations is derived by Sved (1971) shown as:

$$r_E^2 = \frac{1}{1 + 4N_e c}$$

Goal

In this paper, we proposed a probability-based approach to compute the exact distribution of allele frequencies at two loci in a finite population at any generation t conditional on the distribution at generation $t-1$. As r^2 is a function of this distribution of allele frequencies, this approach can be used to examine the distribution of r^2over generations as it approaches equilibrium. The exact distribution of LD from our method is used to describe, quantify and compare LD at different equilibria, including equilibrium in the absence or presence of mutation, selection, and filtering by minor allele frequency. We also proposed a deterministic formula for expected LD in the presence of mutation at equilibrium based on the exact distribution of LD.

Method

For a diallelic two-loci system, four possible haplotypes (i.e., A_1B_1, A_1B_2, A_2B_1, and A_2B_2) are possible to form. In a population of size N_e, the frequency counts of these four haplotypes can take on $k = \frac{(2N_e+3)!}{3!\,(2N_e)!}$ possible values, which corresponds to k possible allele frequencies at two loci. Given the probabilities of the k possible frequency counts at generation t ($\mathbf{P}_t$), the probabilities in generation $t+1$ can be computed in general as

$$\boldsymbol{P}_{t+1} = \boldsymbol{A}\boldsymbol{P}_t$$

where $\mathbf{A}$ is a $k \times k$ transition matrix that is derived by considering various circumstances of recombination, mutation and selection. Given the computed probabilities of the frequency counts, we are able to compute the corresponding distribution of allele frequencies and r^2.

In the present study, starting with an allele frequency of 0.5 at each locus and linkage equilibrium between the two loci, the expected value of r^2 was computed over generations given some values of the mutation rate, recombination rate, selection coefficient and effective population size. Mutation rate of $u = 0$ is used to represent the absence of mutation, while $u = 1\times10^{-9}$ is used to represent the existence of mutation. Similarly, a selection coefficient of $s = 0$ is used to represent the absence of selection, while $s = 0.1$ or $s = 0.01$ is used to represent the existence of selection.

1. The expected value of LD at equilibrium may decrease in the presence of selection.

2. Caution is needed when LD between a causal variant and a marker is inferred after filtering out marker loci with low MAF.

3. "Fake" equilibrium may appear in the presence of mutation.

4. Deterministic formulas for expected LD in the presence of mutation have been proposed.

$$r_E^2 = \frac{1}{1.17 + 11.1\,N_e c} \quad (c = 0.04)$$

SCAN ME

Take a picture to download the full paper

UCDAVIS Department of Animal Science Quantitative Genetics Lab

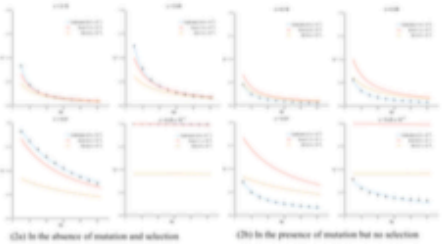

Figure 2. Comparison of Sved's and Hill's approximation to exact distribution of r_E^2 (scatter points) derived from transition-matrix approach. Mean square errors are shown in the parentheses. "Calibrated" denotes the non-linear regression formula derived from our transition-matrix approach.

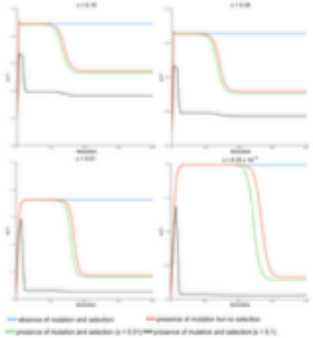

Figure 3. Expected value of LD ($E(r^2)$) over generations under four different conditions at different recombination rates (c) for a population of effective size N_e= 50.

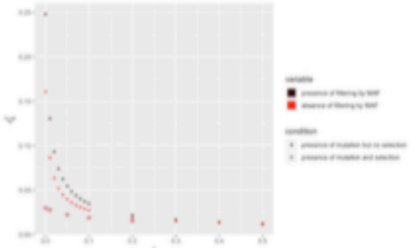

Figure 4. Relationship between the expectation of LD at equilibrium (r_E^2) and recombination rate (c) in the absence or presence of selection or filtering by MAF at locus B for a population of effective size N_e= 50 with mutation rate u= 1×10^{-9}. Only locus A is under selection with s = 0.1.

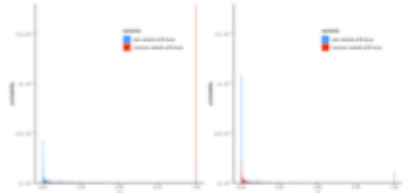

Figure 5. Distribution of LD (r^2) with recombination rate (c) of 6.25×10^{-3} at equilibrium in the presence of mutation but no selection or in the presence of mutation and selection in a population of effective size Ne = 50. Only locus A is under selection.