OPEN ACCESS Freely available online

PLOS | ONE

Evolution of Zygotic Linkage Disequilibrium in a Finite Local Population

Xin-Sheng Hu*

Department of Plant Sciences, University of Oxford, Oxford, United Kingdom

Abstract

One crucial feature of zygotic linkage disequilibrium (LD) analysis is its direct use of diploid genotyping data, irrespective of the type of mating system. Previous theories from an evolutionary perspective mainly focus on gametic LD, but the equivalent development for zygotic LD is not available. Here I study the evolution of zygotic LD and the covariances between gametic and zygotic LDs or between distinct zygotic LDs in a finite local population under constant immigration from a continent population. I derive the analytical theory under genetic hitchhiking effects or in a neutral process. Results indicate that zygotic LDs (diploid level) are more informative than gametic LD (haploid level) in indicating the effects of different evolutionary forces. Zygotic LDs may be greater than or comparable to gametic LD under the epistatic selection process, but smaller than gametic LD under the non epistatic selection process. The covariances between gametic and zygotic LDs are strongly affected by the mating system, linkage distance, and genetic drift effects, but weakly affected by seed and pollen flow and natural selection. The covariances between different zygotic LDs are generally robust to the effects of gene flow, selection, and linkage distance, but sensitive to the effects of genetic drift and mating system. Consistent patterns exist for the covariances between the zygotic LDs for the two-locus genotypes with one common genotype at one locus or without any common genotype at each locus. The results highlight that zygotic LDs can be applied to detecting natural population history.

Citation: Hu X-S (2013) Evolution of Zygotic Linkage Disequilibrium in a Finite Local Population. PLoS ONE 8(11): e80538. doi:10.1371/journal.pone.0080538

Editor: Yury E. Khudyakov, Centers for Disease Control and Prevention, United States of America

Received April 16, 2013; **Accepted** October 14, 2013; **Published** November 27, 2013

Copyright: © 2013 Xin-Sheng Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The author has no support or funding to report.

Competing Interests: The author has declared that no competing interests exist.

* E-mail: xin-sheng.hu@plants.ox.ac.uk

Introduction

Zygotic linkage disequilibrium (LD) refers to the difference between the joint genotypic frequency at two loci and the product of genotypic frequencies at each locus [1], [2], [3], [4], [5], [6]. The concept itself is a purely statistical term, and can also be viewed as the covariance of genotypic frequencies, analogous to the covariance of allelic frequencies for the concept of gametic LD [7], [8], [9]. Its biological significance can be viewed when used for detecting the effects of evolutionary forces by comparing its empirical distribution with the predicted distribution once an evolutionary model is specified [9], [10]. The commonality between gametic and zygotic LDs lies in their utility for measuring non-random associations between loci. The crucial difference is that zygotic LD analysis does not require a random mating assumption since it directly uses diploid genotyping data. However, gametic LD calculation inferred from diploid genotypes needs this assumption since haplotypes must be priorly known. Such a difference is significant because the potential false-positive errors could be substantial in inferring haplotypes/linkage phases using the diploid genotyping data sampled from a natural population of a mixed mating system.

Previously relevant theories emphasize the joint frequency of double heterozygotes or double homozygotes in a neutral process, or the joint descent measures for a population with a mixed mating system [3], [4]. [11], [12], [13], [14], [15]. Zygotic LD is implicitly indicated from interpreting character associations in a partial inbreeding system [16], or from explaining an excess of the equilibrium genotypic frequencies at two independent loci in a mixed mating system [17], or from defining the covariance of heterozygosities [18] or the covariance of descent identities [19]. More recent studies concentrate on the statistical issues, including the procedure for testing zygotic LDs [5], [6], [20], [21], [22] and the potential application of zygotic LD to mapping quantitative trait loci (QTL) [23]. Unlike gametic LD that has received considerably theoretical studies from the evolutionary perspective [9], [10], [24], an equivalent theory for zygotic LD has not been fully developed. Although the evolutionary forces acting on the gametic LD may, in principle, also affect the zygotic LD, these effects and the resultant patterns have not been explicitly studied. This void motivates me to study how zygotic and other high-order LDs evolve under the effects of different evolutionary forces.

In flowering plants, three distinct processes in a life cycle are involved in changing zygotic LD and its relationship with gametic LD in a local population. One process is the asymmetric immigration through haploid pollen flow and diploid seed flow from a source population. Pollen flow directly generates gametic LD, but indirectly affects zygotic LD since each pollen grain only carries one gamete in fusion with ovules in the recipient population. Seed flow can generate both zygotic and gametic LDs since each seed carries two gametes into the recipient population simultaneously.

The second process influencing zygotic LD in plants is the mating system [25]. Selfing facilitates both gametic and zygotic LDs, even for the loci with a free recombination [17]; while random mating erodes both gametic and zygotic LDs. This effect

can be unequal on zygotic and gametic LDs since zygotic LD might be more sensitive than gametic LD to the mating system.

The third process influencing zygotic LD in plants is selection in either the gametophyte or the sporophyte stage, or in both stages. Selection against heterozygote or epistatic selection at the sporophyte stage can directly change zygotic LD, but indirectly changes gametic LD [26], such as in natural hybrid zones [27], [28], [29], [30]. This is analogous to the conventional artificial selection that directly exerts effects on zygotic LD but indirectly on gametic LD in plant and animal breeding programs. Selection in the gametophyte stage directly changes gametic LD, but indirectly changes zygotic LD owing to the connection between the gametophyte and sporophyte stages in one life cycle. The natural overloading of pollen on the stigma of a flower implies pollen competition and the occurrence of natural selection [31], [32], [33]. Also, an excess of ovule abortion in many single embryo or polyembryony plants implies the occurrence of natural selection in ovules [34], [35]. Some genes can express at both the gametophyte and sporophyte stages [36], [37], [38], [39], [40], and might experience different extents of selection pressure. Selection can change both gametic and zygotic LDs among these genes. Thus, it is of both theoretical and practical significance to study how the above three distinct processes evolve zygotic LD.

Here I examine how different driving forces (mating system, genetic drift, migration, and natural selection) affect zygotic LD from the evolutionary perspective, complementary to the existing statistical issues. An island-continent model is considered, with an emphasis on the evolution of zygotic LD in the finite island population. I begin by presenting an exact model and use Monte Carlo (MC) simulations to evaluate the evolution of zygotic LD under different evolutionary forces in the population with a mixed mating system. I then derive analytical theories in two specific cases (genetic hitchhiking effects and a neutral process) under random mating, and validate the theories through MC simulations. Through the analytical and simulation results, I explore the evolutionary properties of zygotic LDs and discuss their potential utility.

Results

Exact Model

Consider an island population, with constant immigration rates of pollen (m_P) and seeds (m_S) per generation from a continent population. The continent population is assumed to be sufficiently large in size and stable in genetic composition. Migration from the island to continent population is neglected, and mutation effects are excluded. At the gametophyte stage, pollen and ovules are subject to natural selection before they are combined to produce seeds. The plant life cycle follows a sequence of events: pollen and ovules generation, pollen flow, selection at the gametophyte stage, mixed mating with a selfing rate α ($0 \le \alpha \le 1$), seed flow, selection at the sporophyte stage, and genetic drift. Selection strength may be either strong or weak; or epistatic selection is allowed in either the gametophyte or the sporophyte stage, or in both. The same mating system is assumed in the island and continent populations.

Consider two diallelic nuclear loci, with alleles $A_1$and A_2 at locus A, alleles B_1 and B_2 at locus B, and a recombination rate r between the two loci. These alleles may refer to single nucleotide polymorphic (SNP) markers since tri- or tetra-allelic SNP sites are infrequent in natural populations. In the island population, let $p_{A_iB_k}$(i, k = 1, 2) and D_{AB}be the frequency of gamete A_iB_k and the gametic LD in current adults, respectively; and $p_{A_iB_k}$can be expressed as $p_{A_iB_k} = p_{A_i}p_{B_k} + (-1)^{i+k}D_{AB}$. For the random mating part, the frequency of gamete A_iB_k in pollen or ovules (the next generation) produced by the current adults can be expressed as $p^*_{A_iB_k} = p_{A_i}p_{B_k} + (-1)^{i+k}(1-r)D_{AB}$ [41]. Similarly, let $p_{A_iA_jB_kB_l}$, $p_{A_iA_j}$, and $p_{B_kB_l}$ be the frequencies of genotypes $A_iA_jB_kB_l$, A_iA_j and $B_kB_l(i,j,k,l = 1,2; i \le j; k \le l)$ in the current adult population, respectively. Let $D_{A_iA_jB_kB_l}$ be the zygotic LD between genotypes A_iA_j at locus A and $B_kB_l$at locus B for two-locus genotype $A_iA_jB_kB_l$, i.e. $D_{A_iA_jB_kB_l} = p_{A_iA_jB_kB_k} - p_{A_iA_j}p_{B_kB_l}$. There are eight zygotic LDs in total, but only four of them are independent since the following constraints hold: $\sum_{i,j=1;i\le j}^{2} D_{A_iA_jB_kB_l} = \sum_{k,l=1;k\le l}^{2} D_{A_iA_jB_kB_l} = 0$. This leads to $\sum_{i,j=1}^{2} D_{A_iA_iB_jB_j} = D_{A_1A_2B_1B_2}$ and $\sum_{k,l=1;k\le l}^{2} D_{A_iA_iB_kB_l} = \sum_{i,j=1;i\le j}^{2} D_{A_iA_jB_{B_k}B_k} = -D_{A_1A_2B_1B_2}$ [5], [6].

In the continent population, let $Q_{A_iA_jB_kB_l}, Q_{A_iA_j}$, and $Q_{B_kB_l}$ be the frequencies of genotypes $A_iA_jB_kB_l$, A_iA_j and $B_kB_l(i,j,k,l = 1,2; i \le j; k \le l)$, respectively. Let $Q_{A_iB_k}, \bar{D}_{AB}$, and $\bar{D}_{A_iA_jB_kB_l}$be the frequency of gameteA_iB_k, the gametic LD, and the zygotic LD in the current adults, respectively. Similar constraints for zygotic LDs to the case in the island population hold as well. All zygotic and gametic LDs are assumed to be constant in the continent population.

Let $w_{A_iB_k(P)}$ and $w_{A_iB_k(O)}$ be the fitness of gamete A_iB_k in pollen and ovules, respectively. The average fitness in pollen and ovules, denoted by w_P and w_O, respectively, can be expressed as $w_P = \sum_{i=1}^{2}\sum_{l=1}^{2} w_{A_iB_k(P)}p'^*_{A_iB_k}$ and $w_O = \sum_{i=1}^{2}\sum_{l=1}^{2} w_{A_iB_k(O)}p^*_{A_iB_k}$ where $p'^*_{A_iB_k}(=(1-m_P)p^*_{A_iB_k} + m_PQ^*_{A_iB_K})$ is the gametic frequency after pollen flow. The gametic frequencies in ovules remain unaltered since ovules do not move after pollen flow. Let $w_{A_iA_jB_kB_l}$ be the fitness for $A_iA_jB_kB_l$ $(i,j = 1,2, i \le j; k,l = 1,2, k \le l)$. The average fitness in the sporophyte stage, w, can be calculated by $w = \sum_{i=1}^{2}\sum_{j=1,j\le i}^{2}\sum_{k=1}^{2}\sum_{l=1}^{2} w_{A_iA_jB_kB_l}p^*_{A_iA_jB_kB_l}$ where $p^*_{A_iA_jB_kB_l}$ is the genotypic frequency after seed flow. Following the plant life cycle, the genotypic frequency after selection in the sporophyte stage, denoted by $p^{**}_{A_iA_jB_kB_l}$ $(= w_{A_iA_jB_kB_l}p^*_{A_iA_jB_kB_l}/w)$ for $A_iA_iB_kB_l$ $(i,j,k,l = 1,2; i \le j; k \le l)$, is derived in Appendix S1.

After genetic drift, the numbers of distinct genotypes follow a multinomial distribution. Here, the genetic sampling of N breeding individuals (an effective population size) is analogous in technique to but different in biological meaning from the statistical sampling of N individuals [9]. Allelic and genotypic frequencies fluctuate but eventually reach steady-state distributions under the joint effects of migration, selection, and genetic drift. Gametic and zygotic LDs can eventually reach steady-state distributions as well. Since the probability density functions (pdf) of gametic and zygotic LDs are difficult to derive, their expectations can be indirectly evaluated through multiple independent simulations.

Genetic drift at each generation can cause the associations between gametic and zygotic LDs or between different zygotic LDs due to their sharing of some alleles or genotypes. There are four types of covariances between gametic and zygotic LDs, $\mathrm{cov}(D_{A_1B_1}, D_{A_iA_jB_kB_l})(i,j = 1,2, i < j; k,l = 1,2, k < l)$, and six types of covariances between distinct zygotic LDs. Note that other high-order LDs, such as trigenic and composite LDs [9], are not examined here although they can be calculated with more complicated analyses. Fisher's delta method is used to approximate

the covariances between gametic and zygotic LDs or between different zygotic LDs (high-order LDs) ([9], p118), [42].

For example, the covariance between $D_{A_1B_1}$ and $D_{A_1A_1B_1B_1}$ or between $D_{A_1A_1B_1B_1}$ and $D_{A_1A_2B_1B_2}$ is derived as:

$$\begin{aligned}
&\mathrm{cov}(D_{A_1B_1}, D_{A_1A_1B_1B_1}) = \\
&\frac{1}{N}\Big((1-p^{**}_{A_1A_1}-p^{**}_{B_1B_1})(1-p^{*}_{A_1}-p^{*}_{B_1})p^{**}_{A_1A_1B_1B_1} \\
&-p^{**}_{B_1B_1}(1-2p^{**}_{B_1}-p^{**}_{A_1})p^{**}_{A_1A_1B_1B_2}/2+p^{**}_{B_1}p^{**}_{B_1B_1}p^{**}_{A_1A_1B_2B_2} \\
&-p^{**}_{A_1A_1}(1-p^{**}_{B_1}-2p^{**}_{A_1})p^{**}_{A_1A_2B_1B_1}+p^{**}_{A_1}p^{**}_{A_1A_1}p^{**}_{A_2A_2B_1B_1}, \\
&-(p^{**}_{A_1B_1}-2p^{**}_{A_1}p^{**}_{B_1})(p^{**}_{A_1A_1B_1B_1}-2p^{**}_{A_1A_1}p^{**}_{B_1B_1})\Big)
\end{aligned} \tag{1}$$

$$\begin{aligned}
&\mathrm{cov}(D_{A_1A_1B_1B_1}, D_{A_1A_2B_1B_2}) = \\
&\frac{1}{N}\Big(p^{**}_{A_1A_1}p^{**}_{B_1B_2}p^{**}_{A_1A_2B_1B_1}+p^{**}_{A_1A_2}p^{**}_{B_1B_1}p^{**}_{A_1A_1B_1B_2} \\
&-(p^{**}_{A_1A_2B_1B_2}-2p^{**}_{A_1A_2}p^{**}_{B_1B_2})(p^{**}_{A_1A_1B_1B_1}-2p^{**}_{A_1A_1}p^{**}_{B_1B_1})\Big),
\end{aligned} \tag{2}$$

where $p^{**}_{\cdot}$ is the allelic or genotypic frequency after selection in the sporophyte stage but before genetic drift. Other high-order covariances can be derived in a similar way. These covariances are needed for calculating the expectations of zygotic LDs described in the section of Analytical Theory [10]. Note that the above covariances produced by genetic drift are conditional on the allelic and genotypic frequencies $p^{**}_{\cdot}$ before genetic drift, i.e. the expectation on the basis of genetic drift (see E_δ in Appendix S4). These high-order covariances can reach steady-state distributions, and their means, e.g., $E(\mathrm{cov}(D_{A_1A_1B_1B_1}, D_{A_1A_2B_1B_2}))$, can be calculated in theory according to their joint probability density distribution (the expectation E is based on the pdf, $\Phi_{\cdot}$ described in the section of Analytical Theory). Similarly, multiple independent simulations can be used to evaluate the expectations of these high-order LDs.

Note that the above general model can reduce to specific models with different numbers of evolutionary forces (e.g., the model with a single evolutionary force). Also, I only concentrate on the covariances between allelic frequencies, or between genotypic frequencies, or between gametic and zygotic LDs, or between different zygotic LDs. The expectations of their normalized values, like the square of normalized gametic LD, r^2_D [8] or Lewontin's D' [43], are difficult to derive under genetic drift effects [44], and hence are not explored further.

Monte Carlo Simulations

MC simulations are used to examine how different evolutionary forces change zygotic LDs and other types of covariances in the plant species of a mixed mating system. Suppose that the island population initially has the same genetic composition as the continent population. For simplicity, notation for the alleles and subscripts in the above exact model is changed as A for A_1, a for A_2, B for B_1, and b for B_2. Simulations are conducted according to the plant life cycle. Given a set of parameters, including the genotypic frequencies in migrants and in the initial island population, the selection coefficients and the effective population size, the genotypic frequencies before genetic drift are calculated from Eqs. (A1) ~ (A5) in Appendix S1. For the genetic drift, random samples are generated using the genotypic frequencies that follow a multinomial distribution. Random numbers with uniform distribution within (0, 1) for sampling purpose are generated using the routine of Press et al. ([45], pp. 210–211). Ten thousand independent simulation runs are conducted for each case. The replicates are used to estimate means and standard deviations of zygotic LDs and other covariances.

Mating System. To examine the effects of mating system, I fix all other parameters except the selfing rate α. Here, gametic and zygotic LDs are not further decomposed into the components of identity (inbreeding in recent ancestry) and non-identity disequilibria [3], [4], and hence the interaction between selfing and genetic drift is unnecessarily specified. Simulations confirm that gametic and zygotic LDs and other covariances can reach steady-state distributions. Note that the parameter settings in all numerical examples are arbitrary as long as these parameters are biologically meaningful. Figure 1a (a coupling linkage phase, $D_{AB}>0$) shows that the steady-state gametic and zygotic LDs have different patterns although they exhibit non monotonic trends with the selfing rate. Their standard deviations also exhibit non monotonic trends with α (Figure 1b). Thus, gametic and zygotic LDs are not a linear function of α, similar to the result in a cytonuclear system [46]. An overlap between the steady-state $E(D_{AaBB})$ and $E(D_{AABb})$ is expected when the two loci initially have the same settings in selection coefficients and genotypic frequencies. There are the same patterns between the steady-state $E(cov(D_{AB}, D_{AABb}))$ and $E(cov(D_{AB}, D_{AaBB}))$, or between the steady-state $E(cov(D_{AABB}, D_{AaBB}))$ and $E(cov(D_{AABB}, D_{AaBB}))$, or between the steady-state $E(cov(D_{AaBB}, D_{AaBb}))$ and $E(cov(D_{AABb}, D_{AaBb}))$ in this example. Selfing increases homozygosity but reduces heterozygosity, resulting in different patterns among gametic and zygotic LDs. The steady-state $E(D_{AB})$ may be smaller than the steady-state expectations of some zygotic LDs in a predominant selfing species (e.g., $E(D_{AABB})$ in Figure 1a).

The steady-state covariances between gametic and zygotic LDs (Figure 1c) or between distinct zygotic LDs (Figure 1e) exhibit a monotonic pattern with α. Selfing facilitates the covariances between gametic and zygotic LDs for the genotypes with homozygotes at one locus (the steady-state $E(cov(D_{AB}, D_{AABb}))$ and $E(cov(D_{AB}, D_{AaBB}))$) or at two loci (the steady-state $E(cov(D_{AB}, D_{AABB}))$, but reduces the steady-state $E(cov(D_{AB}, D_{AaBb}))$. Their standard deviations exhibit different patterns with the selfing rate (Figure 1d). Selfing also facilitates the covariances of zygotic LDs between the genotypes sharing one homozygote (the steady-state $E(cov(D_{AABB}, D_{AaBB}))$ and $E(cov(D_{AABB}, D_{AABb}))$) or sharing one heterozygote (the steady-state $E(cov(D_{AaBB}, D_{AaBb}))$ and $E(cov(D_{AABb}, D_{AaBb}))$), but reduces the covariances of zygotic LDs between the genotypes without any common genotypes (the steady-state $E(cov(D_{AABB}, D_{AaBb}))$ and $E(cov(D_{AaBB}, D_{AABb}))$). The standard deviations for these high-order LDs are stable with the selfing rate except their slight increases at the complete selfing (no effects from pollen flow at $\alpha = 1$; Figure 1f).

The steady-state $E(D_{AB})$ and $E(D_{AABB})$ exhibit different patterns with the selfing rate between the repulsion ($D_{AB}<0$) and coupling ($D_{AB}>0$) linkage phases although they have similar patterns in each linkage phase. The steady-state $E(D_{AaBb})$ and $E(D_{AaBB})$ (or $E(D_{AABb})$) display similar patterns with α in each linkage-phase. Patterns are also similar between two linkage-phase cases for the steady-state $E(cov(D_{AB}, D_{AABB}))$, but not for other three covariances between gametic and zygotic LDs. Selfing reduces the absolute steady-state $E(\mathrm{cov}(D_{AB}, D_{AaBb}))$ in each linkage phase. All covariances between different zygotic LDs have the same responding patterns to the selfing rate α in each linkage phase (data not shown here).

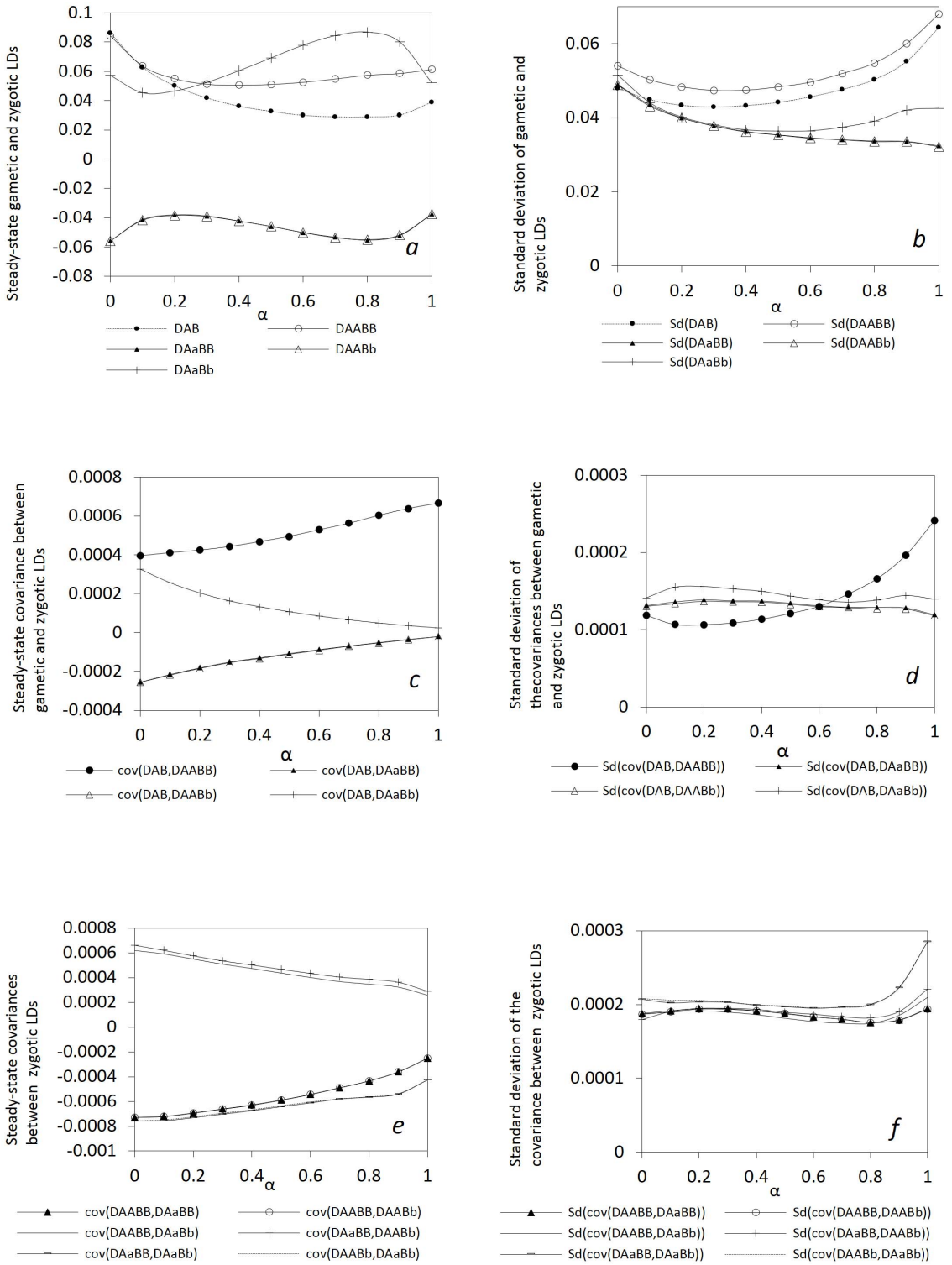

Figure 1. Effects of selfing on the steady-state gametic and zygotic LDs and other types of covariances. Average steady-state gametic and zygotic LDs (a) and their standard deviations (b); average steady-state covariances between gametic and zygotic LDs (c) and their standard deviations (d); and average steady-state covariances between distinct zygotic LDs (e) and their standard deviations (f). Results are obtained from 10000 independent simulation runs. Parameter settings are the recombination rate = 5%, the immigration rate of pollen $m_P = 0.08$ and seeds $m_S = 0.04$, the effective population size = 50, the fitness in the gametophyte stage (pollen and ovules) $w_{AB} = 1$, $w_{Ab} = 0.98$, $w_{aB} = 0.98$, $w_{ab} = 0.96$, and the fitness in the sporophyte stage $w_{AABB} = 1$, $w_{AABb} = w_{AaBB} = 0.98$, $w_{AAbb} = w_{AaBb} = w_{aaBB} = 0.96$, $w_{Aabb} = w_{aaBb} = 0.94$, and $w_{aabb} = 0.92$. The genotypic frequencies in the continent and initial island populations are 0.1225 for *AABB*, 0.105 for *AABb*, 0.0225 for *AAbb*, 0.105 for *AaBB*, 0.245 for *AB/ab*, 0.045 for *ab/aB*, 0.105 for *Aabb*, 0.0225 for *aaBB*, 0.105 for *aaBb*, and 0.1225 for *aabb*.
doi:10.1371/journal.pone.0080538.g001

The above examples indicate that plants with distinct mating systems have different zygotic LDs and other covariances in a local population even under the same impacts of immigration. Both zygotic and gametic LDs are sensitive to the pattern of mating system. Predominant outcrossing species have weaker covariances between zygotic and gametic LDs, but stronger covariances between distinct zygotic LDs than do the predominant selfing species.

Seed and Pollen Flow. To examine the effects of pollen (or seed) flow, I fix all other parameters except the migration rate of pollen (or seeds). Figure 2a shows that the steady-state $E(D_{AB})$ slightly increases with m_P in a coupling linkage phase ($D_{AB}>0$). The steady-state $E(D_{AABB})$ and $E(D_{AaBb})$ slightly decrease with m_P, while the steady-state $E(D_{AaBB})$ and $E(D_{AABb})$ (negative) slightly increase with m_P(Figure 2a). The steady-state $E(cov(D_{AB}, D_{AABb}))$, $E(cov(D_{AB}, D_{AaBB}))$, and $E(cov(D_{AB}, D_{AaBb}))$ slightly increase with m_P while the steady-state $E(cov(D_{AB}, D_{AABB}))$ decreases with m_P(Figure 2c). All covariances between different zygotic LDs slightly decrease with m_P(Figure 2e). All standard deviations slightly decrease with m_P(Figures 2b, d, and f).

Seed flow has greater effects than pollen flow on zygotic LDs and other covariances (Figure 3; the same parameter settings as in Figure 2 except the different migration rates of seeds and pollen). The steady-state $E(D_{AB})$ changes faster with m_S than any steady-state zygotic LDs. Generally, the steady-state zygotic LDs and the covariances between different zygotic LDs or between gametic and zygotic LDs do not monotonically change with m_S. The steady-state $E(D_{AABB})$ and $E(D_{AaBb})$ slightly increase as m_S approaches the value of selection coefficient, and then slightly decrease afterwards (Figure 3a). Similar patterns exist for the change of the steady-state $E(cov(D_{AB}, D_{AABB}))$, $E(cov(D_{AB}, D_{AaBb}))$, $E(cov(D_{AABB}, D_{AaBb}))$, and $E(cov(D_{AaBB}, D_{AABb}))$ with m_S(Figures 3c and e). To the contrary, the steady-state $E(D_{AaBB})$ and $E(D_{AABb})$ slightly decrease as m_S approaches the value of selection coefficient, and then slightly increase afterwards. Similar patterns exist for the change of the steady-state $E(cov(D_{AB}, D_{AaBB}))$, $E(cov(D_{AB}, D_{AABb}))$, $E(cov(D_{AABB}, D_{AaBB}))$, $E(cov(D_{AABB}, D_{AABb}))$, $E(cov(D_{AaBB}, D_{AaBb}))$, and $E(cov(D_{AABb}, D_{AaBb}))$ (Figures 3b and c). The same pattern exists for the covariances between zygotic LDs for the genotypes with a common genotype at one locus, or for the genotypes without any common genotype at each locus. All standard deviations gradually decrease with m_S (Figures 3b, d, and f).

These examples indicate that a local plant population generally exhibits robust responses to the impacts of immigration of pollen and seeds in terms of zygotic LDs, or the covariances between gametic and zygotic LDs, or the covariances between distinct zygotic LDs. Seed and pollen flow have small effects on high-order LDs in a local population.

Genetic Drift. To examine the effects of genetic drift, I fix all other parameters except the effective population size (N). Figure 4 shows the results for a predominant outcrossing species ($\alpha=5\%$). The steady-state $E(D_{AB})$ and $E(D_{AABB})$ slightly increase with N. The steady-state $E(D_{AaBB})$, $E(D_{AABb})$, and $E(D_{AaBb})$ (genotypes with heterozygote at one locus or two loci) slightly decrease as the effective population size increases (Figure 4a). The steady-state $E(cov(D_{AB}, D_{AABB}))$ and $E(cov(D_{AB}, D_{AaBb}))$ gradually reduce to zero as N increases. To the contrary, the steady-state $E(cov(D_{AB}, D_{AABb}))$ and $E(cov(D_{AB}, D_{AaBB}))$ gradually increase to zero as N increases (Figure 4c). The steady-state $E(cov(D_{AABB}, D_{AaBb}))$ and $E(cov(D_{AaBB}, D_{AABb}))$ gradually reduce to zero with N, while other steady-state $E(cov(D_{AABB}, D_{AaBB}))$, $E(cov(D_{AABB}, D_{AABb}))$, $E(cov(D_{AaBB}, D_{AaBb}))$, and $E(cov(D_{AABb}, D_{AaBb}))$ gradually increase to zero with N (Figure 4e). It is clear that covariances between gametic and zygotic LDs or between different zygotic LDs are more sensitive than gametic and zygotic LDs to the genetic drift effects. All standard deviations gradually decrease with N (Figures 4b, d, and f).

The examples indicate that a small local population can affect zygotic LDs, and has large effects on the covariances between gametic and zygotic LDs or between distinct zygotic LDs. These high-order covariances are more informative than gametic LD in signaling the effects of population demographic dynamics.

Selection. To assess the effects of linear-additive selection, I examine three selection schemes: gametic selection only, zygotic selection only, and both gametic and zygotic selection. Table 1 shows a comparison in the steady-state zygotic and gametic LDs and other types of covariances. The steady-state $E(D_{AB})$ slightly decreases while the absolute steady-state zygotic LDs and other types of covariances increase from the case of gametic selection only to the case of zygotic selection only, and to the case of joint selection. The examples indicate that cumulative selection can enhance zygotic LDs and other covariances in the linear additive-viability model (Table 1).

To assess the effects of epistatic selection, I use Dobzhansky-Muller's incompatibility model [27], [28], [47] as an example to demonstrate how epistatic selection in the sporophyte stage affects gametic and zygotic LDs. Three cases with different extents of epistatic selection are examined. Selection in the gametophyte stage is excluded in each case. In Case I, the genotypic fitness is set as $w_{AABB}=w_{aabb}=1$, $w_{AaBB}=w_{aaBb}=0.99$, $w_{AABb}=w_{Aabb}=0.99$, $w_{aaBB}=0.98$, $w_{AAbb}=0.98$, and $w_{AaBb}=0.98$. In Case II, the genotypic fitness is set as $w_{AABB}=w_{aabb}=1$, $w_{AaBB}=w_{aaBb}=0.99$, $w_{AABb}=w_{Aabb}=0.5$, $w_{aaBB}=0.98$, $w_{AAbb}=0.5$, and $w_{AaBb}=0.5$. In Case III, the genotypic fitness is set as $w_{AABB}=w_{aabb}=1$, $w_{AaBB}=w_{aaBb}=0.99$, $w_{AABb}=w_{Aabb}=0.1$, $w_{aaBB}=0.98$, $w_{AAbb}=0.1$, and $w_{AaBb}=0.1$. These three cases are the same as matrices (13), (14), and (15) of Gavrilets [48], respectively. In these settings, alleles A and b have a progressively negative interaction on fitness (incompatible background interactions) from Cases I to III.

Results indicate that epistatic selection can change the relative gametic and zygotic LDs (Table 2). The steady-state frequency of allele B increases while the steady-state frequency of allele A decreases from Cases I to III. The steady-state $E(D_{AaBB})$ and absolute steady-state $E(D_{AaBb})$ become greater than the steady-state $E(D_{AB})$ in Case III. The steady-state $E(D_{AB})$, $E(D_{AABB})$, and $E(D_{AaBb})$ decrease while the steady-state $E(D_{AaBB})$ and $E(D_{AABb})$ increase from Cases I to III. Epistatic selection also changes the covariances between gametic and zygotic LDs or between different zygotic LDs. The steady-state $E(cov(D_{AB}, D_{AABB}))$, $E(cov(D_{AB}, D_{AABb}))$, and $E(cov(D_{AB}, D_{AaBb}))$ decrease while the steady-state $E(cov(D_{AB}, D_{AaBB}))$ increases from Cases I to III. The steady-state $E(cov(D_{AABB}, D_{AABb}))$, $E(cov(D_{AABB}, D_{AaBb}))$, and $E(cov(D_{AaBB}, D_{AABb}))$ decrease while the steady-state $E(cov(D_{AABB}, D_{AaBB}))$, $E(cov(D_{AaBB}, D_{AaBb}))$, and $E(cov(D_{AABb}, D_{AaBb}))$ increase from Cases I to III.

The above examples indicate that zygotic and gametic LDs have different responding patterns to natural selection. The cumulative selection can enhance zygotic LDs and other covariances in the additive-viability selection model. One striking result is that epistatic selection at the diploid level can produce zygotic LDs that are greater than or comparable to gametic LD. This pattern can be used to detect the epistatic selection process in natural populations.

Analytical Theory

To further understand the evolution of zygotic LDs, I derive the analytical theory in a linear-additive-viability model with weak selection and random mating ($\alpha=0$). The gametic fitness in pollen and ovules is decomposed as $w_{A_iB_k(P)}=1+s_{A_i(P)}+s_{B_k(P)}$ and $w_{A_iB_k(O)}=1+s_{A_i(O)}+s_{B_k(O)}$ where $s_{A_i(P)}$ and $s_{A_i(O)}$ are the

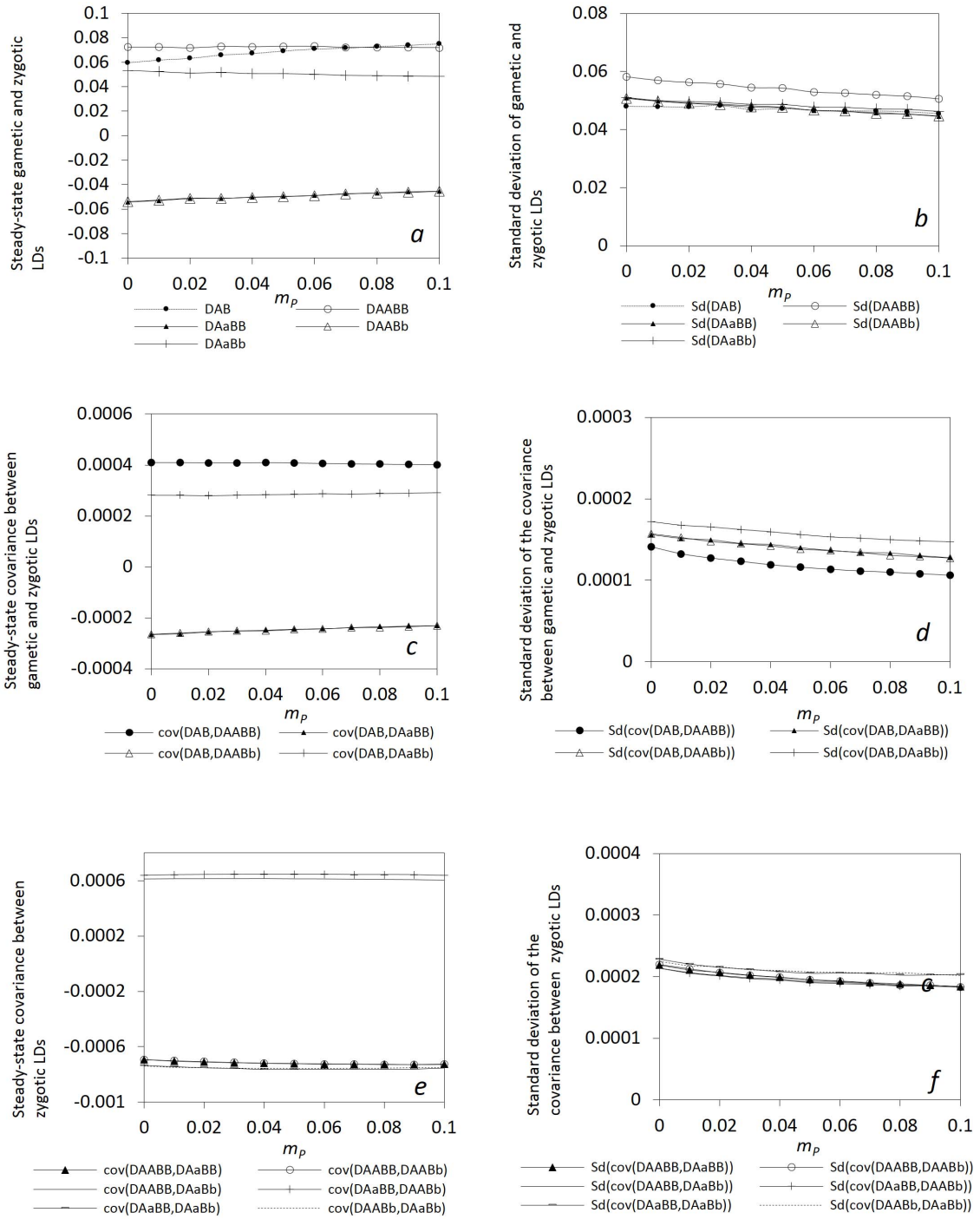

Figure 2. Effects of pollen flow on the steady-state gametic and zygotic LDs and other types of covariances. Average steady-state gametic and zygotic LDs (a) and their standard deviations (b); average steady-state covariances between gametic and zygotic LDs (c) and their standard deviations (d); and average steady-state covariances between distinct zygotic LDs (e) and their standard deviations (f). Results are obtained from 10000 independent simulation runs. Parameter settings are the selfing rate = 5%, the recombination rate = 5%, the effective population size = 50, the immigration rate of seeds $m_S = 0.04$, and the fitness in the gametophyte stage (pollen and ovules) $w_{AB} = 1$, $w_{Ab} = 0.98$, $w_{aB} = 0.98$, $w_{ab} = 0.96$, and the fitness in the sporophyte stage $w_{AABB} = 1$, $w_{AABb} = w_{AaBB} = 0.98$, $w_{AAbb} = w_{AaBb} = w_{aaBB} = 0.96$, $w_{Aabb} = w_{aaBb} = 0.94$, and $w_{aabb} = 0.92$. The genotypic frequencies in the continent and initial island populations are 0.1225 for *AABB*, 0.105 for *AABb*, 0.0225 for *AAbb*, 0.105 for *AaBB*, 0.245 for *AB/ab*, 0.045 for *ab/aB*, 0.105 for *Aabb*, 0.0225 for *aaBB*, 0.105 for *aaBb*, and 0.1225 for *aabb*.
doi:10.1371/journal.pone.0080538.g002

selection coefficients for allele A_i in pollen and ovules, respectively; $s_{B_k(P)}$ and $s_{B_k(O)}$ are the selection coefficients for allele B_k in pollen and ovules, respectively. The genotypic fitness in the sporophyte stage is expressed as $w_{A_iA_jB_kB_l} = 1 + s_{A_iA_j} + s_{B_kB_l}$ where $s_{A_iA_j}$ and $s_{B_kB_l}$ are the selection coefficients for genotypes A_iA_j and B_kB_l, respectively.

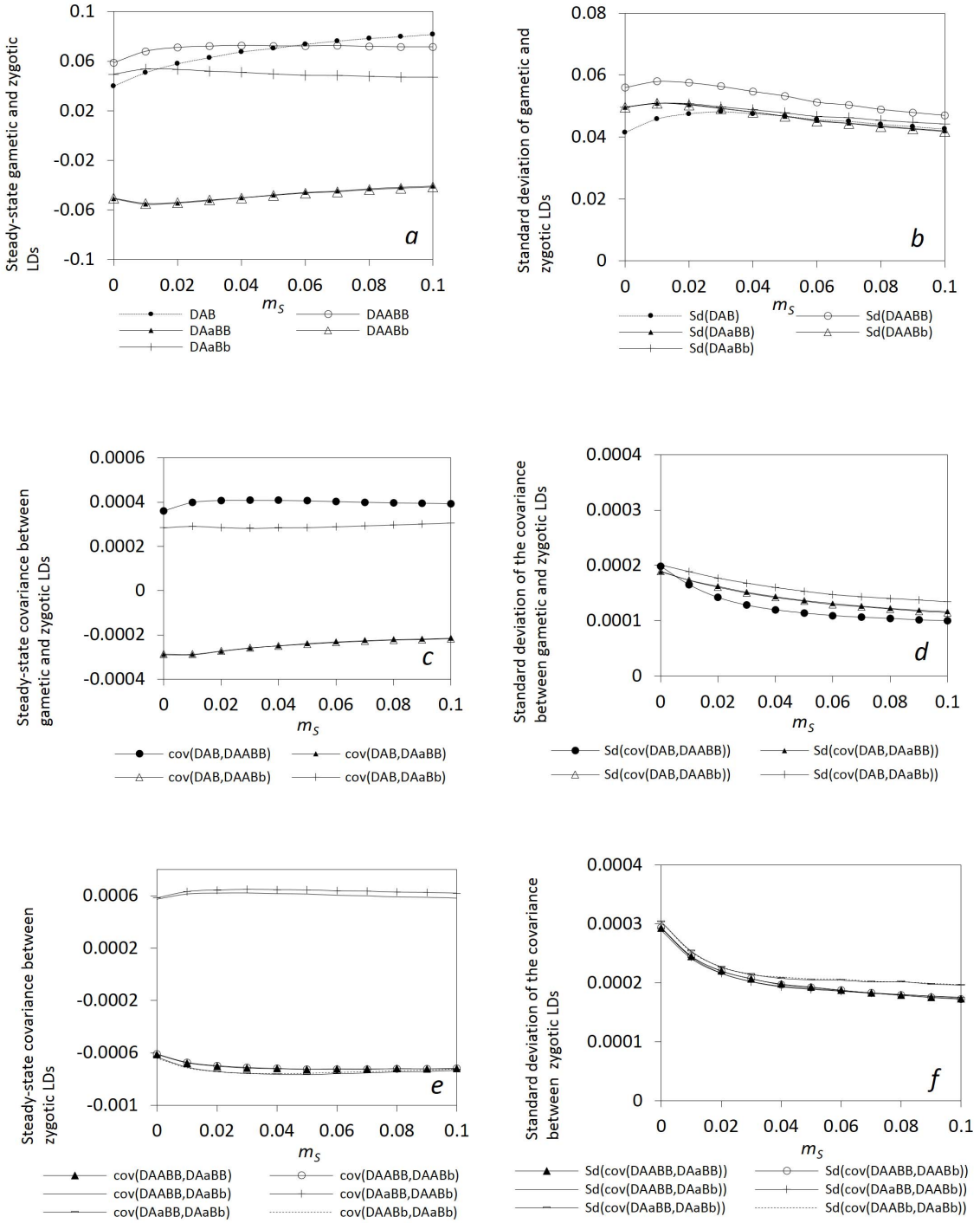

Figure 3. Effects of seed flow on the steady-state gametic and zygotic LDs and other types of covariances. Average steady-state gametic and zygotic LDs (a) and their standard deviations (b); average steady-state covariances between gametic and zygotic LDs (c) and their standard deviations (d); and average steady-state covariances between distinct zygotic LDs (e) and their standard deviations (f). Results are obtained from 10000 independent simulation runs. Parameter settings are the selfing rate $=5\%$, the effective population size $=50$, the immigration rate of pollen $m_P=0.04$, and the fitness in the gametophyte stage (pollen and ovules) $w_{AB}=1$, $w_{Ab}=0.98$, $w_{aB}=0.98$, $w_{ab}=0.96$, and the fitness in the sporophyte stage $w_{AABB}=1$, $w_{AABb}=w_{AaBB}=0.98$, $w_{AAbb}=w_{AaBb}=w_{aaBB}=0.96$, $w_{Aabb}=w_{aaBb}=0.94$, and $w_{aabb}=0.92$. The genotypic frequencies in the continent and initial island populations are 0.1225 for *AABB*, 0.105 for *AABb*, 0.0225 for *AAbb*, 0.105 for *AaBB*, 0.245 for *AB/ab*, 0.045 for *ab/aB*, 0.105 for *Aabb*, 0.0225 for *aaBB*, 0.105 for *aaBb*, and 0.1225 for *aabb*.
doi:10.1371/journal.pone.0080538.g003

With the weak selection, all items containing the second or higher orders of selection coefficients are neglected. The immigration rates of seeds and pollen are assumed to be small. The items containing the second or higher orders of the migration rate (m_P^2, m_S^2, or $m_S m_P$, or higher orders) or the products of the migration rate with selection coefficients (sm_P or sm_S) are

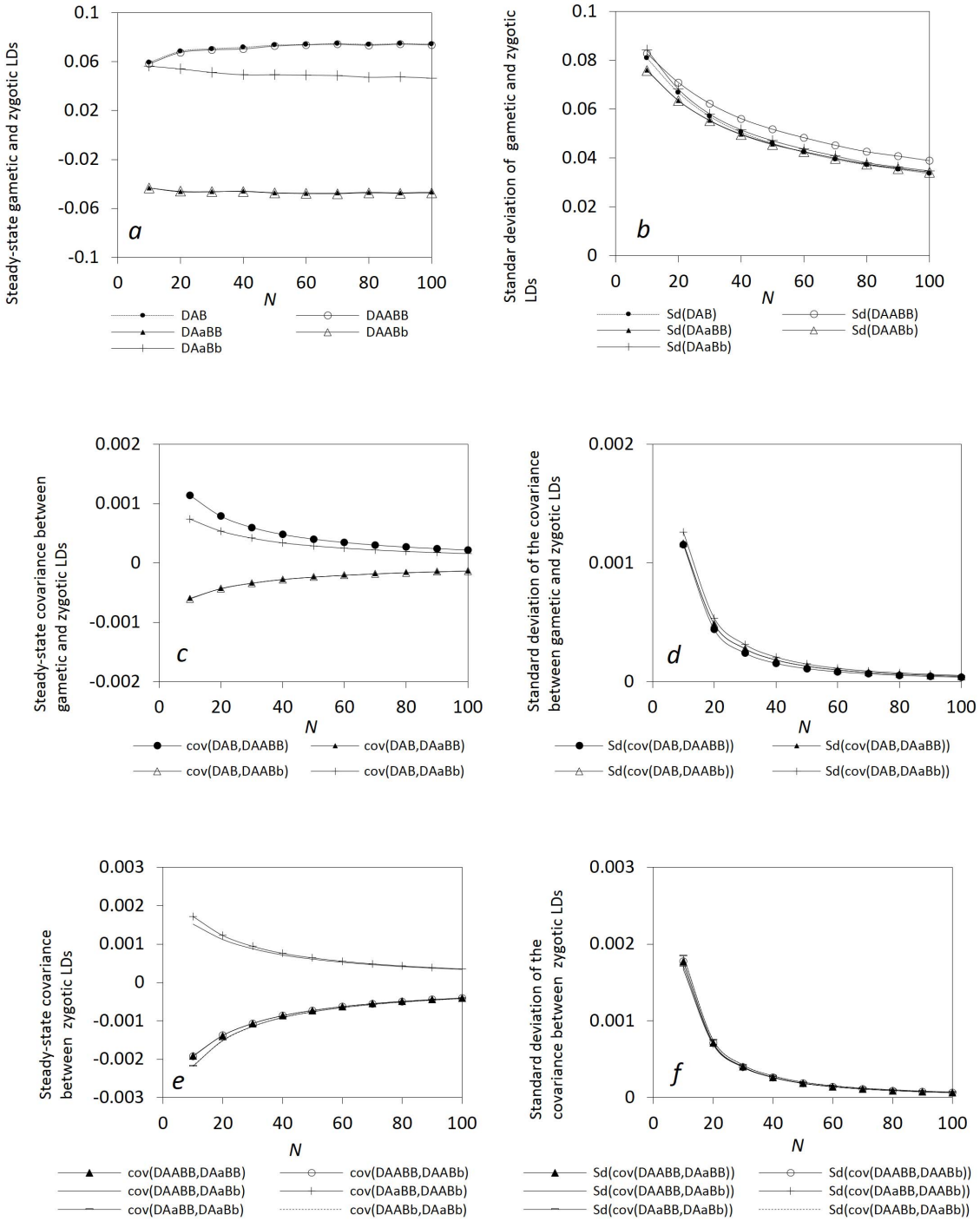

Figure 4. Genetic drift effects on the steady-state gametic and zygotic LDs and other types of covariances. Average steady-state gametic and zygotic LDs (a) and their standard deviations (b); average steady-state covariances between gametic and zygotic LDs (c) and their standard deviations (d); and average steady-state covariances between distinct zygotic LDs (e) and their standard deviations (f). Results are obtained from 10000 independent simulation runs. Parameter settings are the selfing rate $=5\%$, the immigration rate of seeds $m_S=0.04$ and pollen $m_P=0.08$, and the fitness in the gametophyte stage (pollen and ovules) $w_{AB}=1$, $w_{Ab}=0.98$, $w_{aB}=0.98$, $w_{ab}=0.96$, and the fitness in the sporophyte stage $w_{AABB}=1$, $w_{AABb}=w_{AaBB}=0.98$, $w_{AAbb}=w_{AaBb}=w_{aaBB}=0.96$, $w_{Aabb}=w_{aaBb}=0.94$, and $w_{aabb}=0.92$. The genotypic frequencies in the continent and initial island populations are 0.1225 for *AABB*, 0.105 for *AABb*, 0.0225 for *AAbb*, 0.105 for *AaBB*, 0.245 for *AB/ab*, 0.045 for *ab/aB*, 0.105 for *Aabb*, 0.0225 for *aaBB*, 0.105 for *aaBb*, and 0.1225 for *aabb*.
doi:10.1371/journal.pone.0080538.g004

neglected. Again, notation for the alleles and subscripts in the exact model is changed as A for A_1, a for A_2, B for B_1, and b for B_2. Selection coefficients are set as $s_{A_1(P)}=s_{A_1(O)}=0$, $s_{B_1(P)}=s_{B_1(O)}=0$, $s_{A_2(P)}=-s_{aP}$, $s_{A_2(O)}=-s_{aO}$, $s_{B_2(P)}=-s_{bP}$, and $s_{B_2(O)}=-s_{bO}$. Alleles a and b are maladaptive in the island population. Let h_a and h_b are the degrees of dominance at loci A

Table 1. Effects of selection in the gametophyte stage, the sporophyte stage, and in both stages on the steady-state gametic and zygotic LDs and other high-order covariances*

	Gametic selection	Zygotic selection	Gametic and zygotic selection
p_A	0.5780± 0.1221	0.5893±0. 1217	0.6521±0.1135
p_B	0.5822±0.1224	0.5919±0.1209	0.6559±0.1131
D_{AB}	0.0755±0.0497	0.0746±0.0495	0.0730±0.0463
D_{AABB}	0.0619±0.0495	0.0629±0.0499	0.0722±0.0519
D_{AaBB}	−0.0338±0.0427	−0.0355±0.0433	−0.0468±0.0456
D_{AaABb}	−0.0342±0.0428	−0.0358±0.0431	−0.0471±0.0456
D_{AaBb}	0.0437±0.0468	0.0438±0.0467	0.0492±0.0471
$\mathrm{cov}(D_{AB},D_{AABB})$	$3.619\times10^{-4}\pm1.101\times10^{-4}$	$3.679\times10^{-4}\pm1.107\times10^{-4}$	$4.035\times10^{-4}\pm1.099\times10^{-4}$
$\mathrm{cov}(D_{AB},D_{AaBB})$	$-1.813\times10^{-4}\pm1.341\times10^{-4}$	$-1.887\times10^{-4}\pm1.350\times10^{-4}$	$-2.348\times10^{-4}\pm1.329\times10^{-4}$
$\mathrm{cov}(D_{AB},D_{AABb})$	$-1.840\times10^{-4}\pm1.329\times10^{-4}$	$-1.904\times10^{-4}\pm1.338\times10^{-4}$	$-2.368\times10^{-4}\pm1.313\times10^{-4}$
$\mathrm{cov}(D_{AB},D_{AaBb})$	$2.691\times10^{-4}\pm1.615\times10^{-4}$	$2.689\times10^{-4}\pm1.616\times10^{-4}$	$2.880\times10^{-4}\pm1.508\times10^{-4}$
$\mathrm{cov}(D_{AABB},D_{AaBB})$	$-6.525\times10^{-4}\pm2.086\times10^{-4}$	$-6.653\times10^{-4}\pm2.064\times10^{-4}$	$-7.266\times10^{-4}\pm1.877\times10^{-4}$
$\mathrm{cov}(D_{AABB},D_{AABb})$	$-6.550\times10^{-4}\pm2.088\times10^{-4}$	$-6.665\times10^{-4}\pm2.056\times10^{-4}$	$-7.280\times10^{-4}\pm1.874\times10^{-4}$
$\mathrm{cov}(D_{AABB},D_{AaBb})$	$5.107\times10^{-4}\pm2.043\times10^{-4}$	$5.256\times10^{-4}\pm2.025\times10^{-4}$	$6.088\times10^{-4}\pm1.870\times10^{-4}$
$\mathrm{cov}(D_{AaBB},D_{AABb})$	$5.500\times10^{-4}\pm2.083\times10^{-4}$	$5.636\times10^{-4}\pm2.059\times10^{-4}$	$6.428\times10^{-4}\pm1.855\times10^{-4}$
$\mathrm{cov}(D_{AaBB},D_{AaBb})$	$-6.816\times10^{-4}\pm2.344\times10^{-4}$	$-6.945\times10^{-4}\pm2.297\times10^{-4}$	$-7.631\times10^{-4}\pm2.035\times10^{-4}$
$\mathrm{cov}(D_{AABb},D_{AaBb})$	$-6.723\times10^{-4}\pm2.338\times10^{-4}$	$-6.892\times10^{-4}\pm2.316\times10^{-4}$	$-7.562\times10^{-4}\pm2.051\times10^{-4}$

*Three selection schemes are: $w_{AB}=1, w_{Ab}=0.98$, $w_{aB}=0.98$, and $w_{ab}=0.96$ for gametic selection only; $w_{AABB}=1$, $w_{AABb}=w_{AaBB}=0.98$, $w_{AAbb}=w_{AaBb}=w_{aaBB}=0.96$, $w_{Aabb}=w_{aaBb}=0.94$, and $w_{aabb}=0.92$ for zygotic selection only; and $w_{AB}=1, w_{Ab}=0.98$, $w_{aB}=0.98$, $w_{ab}=0.96$, $w_{AABB}=1$, $w_{AABb}=w_{AaBB}=0.98$, $w_{AAbb}=w_{AaBb}=w_{aaBB}=0.96$, $w_{Aabb}=w_{aaBb}=0.94$, and $w_{aabb}=0.92$ for both gametic and zygotic selection. Other parameter settings are the recombination rate = 5%, the immigration rate of pollen $m_P=0.08$ and seeds $m_S=0.04$, and the effective population size = 50. The genotypic frequencies in the continent and initial island populations are 0.1225 for *AABB*, 0.105 for *AABb*, 0.0225 for *AAbb*, 0.105 for *AaBB*, 0.245 for *AB/ab*, 0.045 for *ab/aB*, 0.105 for *Aabb*, 0.0225 for *aaBB*, 0.105 for *aaBb*, and 0.1225 for *aabb*. The steady-state results (mean ± S_d) are obtained from 10000 independent simulation runs.

doi:10.1371/journal.pone.0080538.t001

and B, respectively. Selection coefficients for genotypes are set as $s_{A_1A_1}=0$, $s_{A_1A_2}=-h_as_a$, and $s_{A_2A_2}=-s_a$ for locus A; and $s_{B_1B_1}=0$, $s_{B_1B_2}=-h_bs_b$, and $s_{B_2B_2}=-s_b$ for locus B.

From Eqs. (A1) ~ (A5) in Appendix S1, the deterministic changes in allelic frequency (Δp_A andΔp_B), gametic (ΔD_{AB}) and four independent zygotic LDs (ΔD_{AABB}, ΔD_{AABb}, ΔD_{AaBB}, and ΔD_{AaBb}), can be derived. Other functions of zygotic LDs can be calculated once the four independent zygotic LDs are available. After genetic drift, the means for the per-generation changes in allelic frequency, gametic and zygotic LDs, can be derived using the conventional approach [49] (Appendix S2). Note that one additional factor $(1-r)$timesD_{AB} in the formulae in Appendix S2 is because D_{AB} is termed from the preceding adults in a plant life cycle (one generation difference between adults and pollen and ovules; [41]).

Let $\Phi(p_A,p_B,D_{AB},D_{AABB},D_{AABb},D_{AaBB},D_{AaBb})$be the steady-state pdf at the two linked loci so that $\Phi dp_Adp_BdD_{AB}dD_{AABB}dD_{AABb}dD_{AaBB}dD_{AaBb}$ represents the expected number of two loci having the allele frequencies, gametic and zygotic LDs within the intervals (p_A,p_A+dp_A), (p_B,p_B+dp_B), …, and $(D_{AaBb}, D_{AaBb}+dD_{AaBb})$, respectively. Expectation of each individual variable can be calculated in theory from pdf Φ. For instance, an expectation of gametic LD can be obtained by$E(D_{AB})=\int\ldots\int\Phi D_{AB}dp_Adp_BdD_{AABB}\ldots dD_{AaBb}$. For a stationary distribution of a function of seven variables $g(p_A,p_B,D_{AB},D_{AABB},D_{AABb},D_{AaBB},D_{AaBb})$, the Kolmogorov backward equation can be derived in the following expression [12], [50]:

$$
\begin{aligned}
0=E\Big\{&M(\delta_{p_A})\frac{\partial g}{\partial p_A}+M(\delta_{p_B})\frac{\partial g}{\partial p_B}\\
&+M(\delta_{D_{AB}})\frac{\partial g}{\partial D_{AB}}+\ldots+M(\delta_{D_{AaBb}})\frac{\partial g}{\partial D_{AaBb}}\\
&+\mathrm{cov}(\delta_{p_A},\delta_{p_B})\frac{\partial^2 g}{\partial p_A\partial p_B}+\mathrm{cov}(\delta_{p_A},\delta_{D_{AB}})\frac{\partial^2 g}{\partial p_A\partial D_{AB}}\\
&+\ldots+\mathrm{cov}(\delta_{D_{AaBB}},\delta_{D_{AaBb}})\frac{\partial^2 g}{\partial D_{AaBB}\partial D_{AaBb}}\\
&+\frac{1}{2}V(\delta_{p_A})\frac{\partial^2 g}{\partial p_A^2}+\frac{1}{2}V(\delta_{p_B})\frac{\partial^2 g}{\partial p_B^2}+\ldots+\frac{1}{2}V(\delta_{D_{AaBb}})\frac{\partial^2 g}{\partial D_{AaBb}^2}\Big\}
\end{aligned}
\tag{3}
$$

Notation E in Eq. (3) means expectation with respect to pdf Φ, the same meaning as in the preceding section except that its calculation is based on numerical simulations.

In Eq. (3), there are seven items with the average change coefficients$M(\delta_{\cdot})$, seven items with the variance coefficients$V(\delta_{\cdot})$, and twenty-one items with the covariance coefficients. Appendix S3 gives the expressions for the variances of per-generation changes in allelic frequency, gametic and zygotic LDs, and all possible covariances among these per-generation changes.

With the diffusion model, the expectations of zygotic LDs and the covariances between gametic and zygotic LDs or between different zygotic LDs can be calculated in theory. However, the algebraic deduction remains complicated when the joint effects of

Table 2. Effects of Dobzhansky-type epistatic selection on the steady-state gametic and zygotic LDs and other high-order covariances*

	Case I	Case II	Case III
p_A	0.5068± 0.1280	0.3090±0. 2127	0.2673±0.2343
p_B	0.5078±0.1292	0.7573±0.1937	0.8170±0.2009
D_{AB}	0.0788±0.0510	0.0329±0.0218	0.0165±0.0113
D_{AABB}	0.0533±0.0462	0.0219±0.0236	0.0100±0.0112
D_{AaBB}	−0.0235±0.0394	0.0257±0.0337	0.0310±0.0248
D_{AaABb}	−0.0231±0.0397	−0.0172±0.0205	−0.0079±0.0095
D_{AaBb}	0.0436±0.0474	−0.0120±0.0312	−0.0245±0.0215
$\mathrm{cov}(D_{AB},D_{AABB})$	$3.217\times10^{-4}\pm1.149\times10^{-4}$	$1.368\times10^{-4}\pm1.195\times10^{-4}$	$0.556\times10^{-4}\pm7.466\times10^{-4}$
$\mathrm{cov}(D_{AB},D_{AaBB})$	$-1.384\times10^{-4}\pm1.353\times10^{-4}$	$0.6650\times10^{-4}\pm1.158\times10^{-4}$	$8.250\times10^{-4}\pm6.508\times10^{-4}$
$\mathrm{cov}(D_{AB},D_{AABb})$	$-1.399\times10^{-4}\pm1.341\times10^{-4}$	$-0.9688\times10^{-4}\pm0.8927\times10^{-4}$	$-0.377\times10^{-4}\pm0.467\times10^{-4}$
$\mathrm{cov}(D_{AB},D_{AaBb})$	$2.692\times10^{-4}\pm1.642\times10^{-4}$	$0.1426\times10^{-4}\pm1.194\times10^{-4}$	$-0.499\times10^{-4}\pm0.505\times10^{-4}$
$\mathrm{cov}(D_{AABB},D_{AaBB})$	$-5.511\times10^{-4}\pm2.298\times10^{-4}$	$-0.928\times10^{-4}\pm1.275\times10^{-4}$	$0.1252\times10^{-4}\pm0.537\times10^{-4}$
$\mathrm{cov}(D_{AABB},D_{AABb})$	$-5.518\times10^{-4}\pm2.289\times10^{-4}$	$-1.987\times10^{-4}\pm1.753\times10^{-4}$	$-0.6276\times10^{-4}\pm0.823\times10^{-4}$
$\mathrm{cov}(D_{AABB},D_{AaBb})$	$3.942\times10^{-4}\pm2.105\times10^{-4}$	$0.8561\times10^{-4}\pm1.252\times10^{-4}$	$-0.092\times10^{-4}\pm0.486\times10^{-4}$
$\mathrm{cov}(D_{AaBB},D_{AABb})$	$4.386\times10^{-4}\pm2.212\times10^{-4}$	$0.8903\times10^{-4}\pm1.284\times10^{-4}$	$-0.091\times10^{-4}\pm0.489\times10^{-4}$
$\mathrm{cov}(D_{AaBB},D_{AaBb})$	$-5.628\times10^{-4}\pm2.536\times10^{-4}$	$-5.495\times10^{-4}\pm2.362\times10^{-4}$	$-2.307\times10^{-4}\pm1.544\times10^{-4}$
$\mathrm{cov}(D_{AABb},D_{AaBb})$	$-5.611\times10^{-4}\pm2.513\times10^{-4}$	$-0.7594\times10^{-4}\pm1.251\times10^{-4}$	$0.128\times10^{-4}\pm0.4867\times10^{-4}$

*Three selection schemes are: $w_{AABB}=w_{aabb}=1$, $w_{AaBB}=w_{aaBb}=0.99$, $w_{AABb}=w_{Aabb}=0.99$, $w_{aaBB}=0.98$, $w_{AAbb}=0.98$, and $w_{AaBb}=0.98$ for Case I; $w_{AABB}=w_{aabb}=1$, $w_{AaBB}=w_{aaBb}=0.99$, $w_{AABb}=w_{Aabb}=0.5$, $w_{aaBB}=0.98$, $w_{AAbb}=0.5$, and $w_{AaBb}=0.5$ for Case II; $w_{AABB}=w_{aabb}=1$, $w_{AaBB}=w_{aaBb}=0.99$, $w_{AABb}=w_{Aabb}=0.1$, $w_{aaBB}=0.98$, $w_{AAbb}=0.1$, and $w_{AaBb}=0.1$for Case III. Other parameter settings are the recombination rate = 5%, the immigration rate of pollen $m_P=0.08$ and seeds $m_S=0.04$, and the effective population size = 50. The genotypic frequencies in the continent and initial island populations are 0.1225 for *AABB*, 0.105 for *AABb*, 0.0225 for *AAbb*, 0.105 for *AaBB*, 0.245 for *AB/ab*, 0.045 for *ab/aB*, 0.105 for *Aabb*, 0.0225 for *aaBB*, 0.105 for *aaBb*, and 0.1225 for *aabb*. The steady-state results (mean ± S_d) are obtained from 10000 independent simulation runs.

doi:10.1371/journal.pone.0080538.t002

selection, migration, and genetic drift are considered. Here, I consider two specific cases. One case is that locus A is selective while locus B is neutral, with an emphasis on genetic hitchhiking effects [51], [52]. The other case is that both loci are neutral, with an emphasis on the effects of linkage distance.

Genetic Hitchhiking. How genetic hitchhiking effects evolve zygotic LDs is an important issue for studying the pattern of genotypic diversity along chromosomes. This may provide a genetic basis for forming zygotic LD blocks, analogous to the gametic LD blocks along chromosomes. Suppose that locus A is mainly subject to the balance between the effects of selection and immigration. The genetic drift effects are negligible for locus A. Locus B is subject to the balance among the effects of immigration, genetic drift, and recombination with locus A. This consideration is similar to the previous studies in examining associative overdominance or genetic hitchhiking effects on spreading neutral nuclear/organelle genes [44], [53], [54], [55]. All items withs_{bP}, s_{bO}, and s_b are eliminated for the average per-generation changes in allelic frequency, gametic and zygotic LDs in the formulae in Appendix S2. The variances for the per-generation changes in allelic frequency p_A and all covariances between p_A and gametic LD or between p_A and zygotic LDs are removed in the formulae in Appendix S3, but the remaining expressions hold except that the steady-state p_A is known. Similarly, the items containing $\delta_{p_A}, V(\delta_{p_A})$, and the covariances between δ_{p_A} and gametic LD or between δ_{p_A} and different zygotic LDs in Eq. (3) are removed.

The steady-state equation for allelic frequency at locus A can be obtained by setting $\Delta p_A=0$, the same as setting $M(\delta_{p_A})=0$, and $s_{bP}=s_{bO}=s_b=0$ in Appendix S2, i.e.

$$0=(m_S+m_P/2)(Q_A-p_A)+p_Ap_a((s_{aP}+s_{aO})/2+s_a(p_a-h_a(1-2p_A))). \quad (4)$$

The steady-state allelic frequency can be numerically calculated from the above cubic equation, given the condition of $0<p_A<1$ and other parameters. Like Ohta and Kimura [44], denote $\hat{p}_A$or $\hat{p}_a$ as the known frequencies calculated from Eq. (4). It can be seen that selection in the gametophyte and sporophyte stages is compounded in the case of $h_a=1/2$.

To calculate the expectations of the steady-state zygotic LDs and other types of covariances from Appendices S2 and S3, the following fourteen expectations are required: $E(p_B^{4-i})(i=0,1,2,3)$, $E(p_B^{3-j}D_{AB})(j=0,1,2,3)$, $E(p_B^{2-k}D_{AB}^2)(k=0,1,2)$, $E(p_B^{1-l}D_{AB}^3)$ $(l=0,1)$, and $E(D_{AB}^4)$. Expectations of a few low-order functions can be analytically derived. For instance, letting $g=p_B$ and $g=D_{AB}$ separately in Eq. (3), I can obtain:

$$E(D_{AB})=\frac{\tilde{m}(1-r)\bar{D}_{AB}}{r+(1-r)(\tilde{m}+1/2N-(1-\hat{p}_A-Q_A)\Delta)}, \quad (5)$$

$$E(p_B)=Q_B+\frac{(1-r)^2\Delta\bar{D}_{AB}}{r+(1-r)(\tilde{m}+1/2N-(1-\hat{p}_A-Q_A)\Delta)}, \quad (6)$$

where $\tilde{m}=m_S+m_P/2$, the joint migration rate, and

$\Delta=(s_{aP}+s_{aO})/2+s_a(\hat{p}_a-h_a(1-2\hat{p}_A))$, the selection component at locus A. Eq. (6) indicates the dependence of the allelic frequency at locus B on the allelic frequency at locus A.

Substitution of g in Eq. (3) by three functions, p_B^2, p_BD_{AB}, and D_{AB}^2, can yield three equations for calculating $E(p_B^2)$, $E(p_BD_{AB})$, and $E(D_{AB}^2)$:

$$\begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{11} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} E(p_B^2) \\ E(p_BD_{AB}) \\ E(D_{AB}^2) \end{pmatrix} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix}, \tag{7}$$

where $a_{11}=2\tilde{m}+1/2N$, $a_{12}=-2(1-r)\Delta, a_{21}=\tilde{m}(Q_A-\hat{p}_A)$, …, and $g_3=(2\tilde{m}Q_B(\hat{p}_A-Q_A)-2\tilde{m}(1-r)\bar{D}_{AB}-(1-2\hat{p}_A)(1-r)/2N)E(D_{AB})-\hat{p}_A\hat{p}_aE(p_B)/2N$.

Expectations of the remaining nine functions can be numerically calculated using *Mathematica* tool [56] by substituting g in Eq. (3) with p_B^3, p_B^4, $p_B^2D_{AB}$, $p_B^3D_{AB}$, $p_BD_{AB}^2$, $p_B^2D_{AB}^2$, D_{AB}^3, $p_BD_{AB}^3$, and D_{AB}^4, respectively. These calculations are not shown here.

With the availability of the above fourteen expectations, the expectations of some lower or the same order functions can be indirectly calculated. For instance, I can obtain $E(p_{AB}^*)=\hat{p}_AE(p_B)+(1-r)E(D_{AB})$ and $E(p_{AB}^*p_{ab}^*)=\hat{p}_A\hat{p}_aE(p_Bp_b)+(1-r)(\hat{p}_AE(p_BD_{AB})+\hat{p}_aE(p_bD_{AB}))+(1-r)^2E(D_{AB}^2)$. The expectation of any steady-state zygotic LD, $E(D_{....})$, can be calculated by substituting g in Eq. (3) with $D_{....}$ (one of the four independent zygotic LDs), resulting in $E(M(\delta_{D_{....}})=0$. For instance, $E(D_{AaBb})$can be calculated by$E(M(\delta_{D_{AaBb}}))=0$ from Appendix S2, i.e.

$$\begin{aligned} E(D_{AaBb}) &= m_P\big(Q_{AB}^*E(p_{ab}^*)+Q_{ab}E(p_{AB}^*) \\ &+Q_{Ab}E(p_{aB}^*)+Q_{aB}E(p_{Ab}^*) \\ &-2\hat{p}_A\hat{p}_a(Q_bE(p_B)+Q_BE(p_b)) \\ &+m_S(Q_{AaBb}-2\hat{p}_A\hat{p}_aQ_{Bb}) \\ &-2(m_S+m_P)(E(p_{AB}^*p_{ab}^*)+E(p_{Ab}^*p_{aB}^*)) \\ &-2(1-r)E(D_{AB})/N \\ &+(1-r)(2(1-2\hat{p}_A)+(s_{aP}+s_{aO})(1-6\hat{p}_A\hat{p}_a) \\ &+2s_a(\hat{p}_a^2(1-4\hat{p}_a) \\ &-h_a(1-4\hat{p}_A\hat{p}_a)(1-2\hat{p}_A))E((1-2p_B)D_{AB}) \\ &+2(1-r)^2(2-s_{aP}-s_{aO}+2s_a(\hat{p}_a(1-2\hat{p}_a) \\ &-h_a(1-4\hat{p}_A\hat{p}_a))E(D_{AB}^2) \end{aligned} \tag{8}$$

Eqs. (5) and (8) indicate that effects of seed and pollen flow are compounded in generating gametic LD, but can be separated in generating zygotic LDs.

The expectations of the steady-state variances of any zygotic LDs can be calculated using Fisher's delta method by omitting all items containing $m_{.}/N$, $s_{.}/N$, and higher orders. It is shown that these expectations can be calculated from the expectations of the variances of the per-generation change in zygotic LD in Appendix S3, i.e. $E(V(D_{....}))=E(V(\delta_{D_{....}}))$with a sufficient accuracy (Appendix S4; [12]). For instance, $E(V(D_{AaBb}))$ can be calculated from $E(V(\delta_{D_{AaBb}}))$, i.e.

$$\begin{aligned} E(V(D_{AaBb})) &= \frac{1}{N}(4\hat{p}_A\hat{p}_a(1-2\hat{p}_A\hat{p}_a)E((1-2p_Bp_b)p_Bp_b) \\ &+2(1-2\hat{p}_A)(1-4\hat{p}_A\hat{p}_a)(1-r)E((1-2p_B)(1-4p_Bp_b)D_{AB}) \\ &-16(1-r)^3(1-2\hat{p}_A)E((1-2p_B)D_{AB}^3)-16(1-r)^4E(D_{AB}^4)\big) \end{aligned} \tag{9}$$

The expectation of any steady-state covariance between gametic and different zygotic LDs can be calculated using Fisher's delta method by omitting all items with $m_{.}/N$, $s_{.}/N$, and higher orders. It is also shown that this expectation can be calculated from the expectation of the covariance in its per-generation change in Appendix S3, i.e. $E(\mathrm{cov}(D_{AB},D_{....}))=E(\mathrm{cov}(\delta_{D_{AB}},\delta_{D_{....}}))$ with a sufficient accuracy (Appendix S4). For instance, $E(\mathrm{cov}(D_{AB},D_{AaBb}))$ can be calculated from $E(\mathrm{cov}(\delta_{D_{AB}},\delta_{D_{AaBb}}))$in Appendix S3, i.e.

$$\begin{aligned} &E(\mathrm{cov}(D_{AB},D_{AaBb})) = \\ &\frac{1}{N}(\hat{p}_A\hat{p}_a(1-2\hat{p}_A)E(p_Bp_b(1-2p_B)) \\ &+(1-r)(1-2\hat{p}_A)^2E((1-2p_B)^2D_{AB}) \\ &+4(1-r)\hat{p}_A\hat{p}_aE(p_Bp_bD_{AB}) \\ &-(1-2\hat{p}_A)(1-r)^2E((1-2p_B)D_{AB}^2) \\ &-4(1-r)^3E(D_{AB}^3)\big) \end{aligned} \tag{10}$$

Similarly, expectations of other covariances in Appendix S3, such as the covariances between different zygotic LDs, can be calculated in the way similar to the above deductions. Expectations of high-order LDs, such as $E(D_{AaBb}^2)$, can be numerically calculated using multiple equations derived by substituting g in Eq. (3) with D_{AaBb}^2 and other high-order LDs. This needs more extensive algebraic analyses, and is not explored further.

Simulations confirm that the above analytical model performs well. For instance, consider the same parameter settings as in Figure 1 for the genotypic frequencies in the continent and island populations, $\alpha=0$, $m_P=0.08$ and $m_S=0.04$, $N=100$, $s_{a(O)}=s_{a(P)}=s_a/2=0.04$, and $h_a=0.5$. Gametic and zygotic LDs and other covariances can reach steady-state distributions ($\sim$50th generation; data not shown here), reflecting the equilibrium among the effects of migration, genetic drift, and genetic hitchhiking. All analytical predictions are distributed within the ranges of one-standard deviations of the simulation results (Table 3).

Figure 5 shows that genetic hitchhiking effects can produce different patterns among gametic and zygotic LDs and other covariances. The expected neutral allelic frequency, $E(p_B)$, increases as the frequency of favorite allele A increases with the selection coefficient. $E(D_{AB})$ gradually decreases while $E(D_{AABB})$ gradually increases with the selection coefficient (Figure 5a). $E(D_{AaBb})$ slightly increases while both $E(D_{AaBB})$ and $E(D_{AABb})$ decrease with the selection coefficient. The covariances between gametic and zygotic LDs for the genotypes with heterozygotes at one locus or two loci gradually decrease with the selection coefficient, except $E(\ cov(D_{AB},\ D_{AABB}))$ showing a different pattern (Figure 5b). The covariances between the zygotic LDs for the genotypes with one common genotype at a locus decrease with the selection coefficient, while the covariances between the zygotic LDs for the genotypes without a common genotype at one locus ($E(cov(D_{AABB},\ D_{AaBb}))$ and $E(cov(D_{AaBB},\ D_{AABb})))$ increase with the

Table 3. Comparison between the simulation results and analytical model predictions under genetic hitchhiking effects*

	MC simulations	Analytical model
p_A	0.7165±0.0716	0.7071
p_B	0.5916±0.0915	0.5509
D_{AB}	0.0777±0.0355	0.0536
D_{AABB}	0.0732±0.0405	0.0465
D_{AaBB}	−0.0533±0.0347	−0.0317
D_{AaABb}	−0.0367±0.0317	−0.0169
D_{AaBb}	0.0419±0.0375	0.0202
$\mathrm{cov}(D_{AB},D_{AABB})$	$2.011\times10^{-4}\pm4.323\times10^{-5}$	2.018×10^{-4}
$\mathrm{cov}(D_{AB},D_{AaBB})$	$-1.463\times10^{-4}\pm3.994\times10^{-5}$	-1.389×10^{-4}
$\mathrm{cov}(D_{AB},D_{AABb})$	$-1.059\times10^{-4}\pm 5.793\times10^{-5}$	-7.188×10^{-5}
$\mathrm{cov}(D_{AB},D_{AaBb})$	$1.541\times10^{-4}\pm6.185\times10^{-5}$	1.032×10^{-4}
$\mathrm{cov}(D_{AABB},D_{AaBB})$	$-3.918\times10^{-4}\pm7.381\times10^{-5}$	-4.125×10^{-4}
$\mathrm{cov}(D_{AABB},D_{AABb})$	$-3.496\times10^{-4}\pm8.167\times10^{-5}$	-3.514×10^{-4}
$\mathrm{cov}(D_{AABB},D_{AaBb})$	$3.140\times10^{-4}\pm7.641\times10^{-5}$	3.123×10^{-4}
$\mathrm{cov}(D_{AaBB},D_{AABb})$	$3.278\times10^{-4}\pm7.803\times10^{-5}$	3.217×10^{-4}
$\mathrm{cov}(D_{AaBB},D_{AaBb})$	$-3.310\times10^{-4}\pm9.101\times10^{-5}$	-3.366×10^{-4}
$\mathrm{cov}(D_{AABb},D_{AaBb})$	$-4.654\times10^{-4}\pm5.826\times10^{-5}$	-5.191×10^{-4}

*Parameter settings are the immigration rate of pollen $m_P=0.08$ and seeds $m_S=0.04$, the effective population size = 100, the selection coefficients in the gametophyte and sporophyte stages $s_{aO}=s_{aP}=s_a/2=0.04$, and the degree of dominance =0.5. The genotypic frequencies in the continent and the initial island populations are 0.1225 for *AABB*, 0.105 for *AABb*, 0.0225 for *AAbb*, 0.105 for *AaBB*, 0.245 for *AB/ab*, 0.045 for *ab/aB*, 0.105 for *Aabb*, 0.0225 for *aaBB*, 0.105 for *aaBb*, and 0.1225 for *aabb*. The steady-state simulation results (mean ± S_d) are obtained from 10000 independent runs.

doi:10.1371/journal.pone.0080538.t003

selection coefficient (Figure 5c). The covariances between the zygotic LDs for the genotypes with a common genotype at the selective locus, i.e. $E(cov(D_{AABB}, D_{AABb}))$ and $E(cov(D_{AaBB}, D_{AaBb}))$, are less sensitive to selection than the covariances between the zygotic LDs for the genotypes with a common genotype at the neutral locus, i.e. $E(cov(D_{AABB}, D_{AaBB}))$ and $E(cov(D_{AABb}, D_{AaBb}))$.

The above results indicate that the gametic LD can have a similarly changing pattern to some zygotic LDs with the selection pressure. This provides the genetic basis of using zygotic LDs to describe genetic hitchhiking effects at the diploid level. Furthermore, the covariances between gametic and zygotic LDs or between distinct zygotic LDs are informative to indicate genetic hitchhiking effects.

Neutral Process. How zygotic LDs evolve in a purely neutral process forms another important issue to study the pattern of genotypic diversity along chromosomes since most molecular population evolution is governed by the neutral process. This also provides theoretical perception of using zygotic LD to reveal the structure of genomic diversity. Suppose that both loci are subject to the balance among the effects of genetic drift, recombination, and immigration. All items with selection coefficients are removed in the formulae in Appendix S2, but all the formulae in Appendix S3 remain unaltered. To assess the steady-state zygotic LDs and other covariances (Appendices S2 and S3), I need to calculate the following fifty-four expectations: $E(p_A^{4-i}p_B^{4-j})(i, j=0, 1, 2, 3,4;$ except $i=j=4)$, $E(p_A^{3-i}p_B^{3-j}D_{AB})(i, j=0,1,2,3)$, $E(p_A^{2-i}p_B^{2-j}D_{AB}^2)(i, j=0,1,2)$, $E(p_A^{1-i}p_B^{1-j}D_{AB}^3)(i, j=0,1)$, and $E(D_{AB}^4)$. These expectations can be numerically calculated with *Mathematica* tool in different equations or different groups of equations.

Letting $g=p_A^i(i=1, 2, 3, 4)$ in Eq. (3), I can obtain

$$E(p_A^i)=\frac{4N\tilde{m}Q_A+i-1}{4N\tilde{m}+i-1}E(p_A^{i-1}). \tag{11}$$

The ith moment of allelic frequency is the same as that derived under a neutral process for individual loci since LD does not affect allelic frequency distribution. Let $F=(1+4N\tilde{m})^{-1}$(an inbreeding coefficient in the island population) [57], which is equal to $\left(E(p_A^2)-(E(p_A))^2\right)/Q_A(1-Q_A)$, analogous to the population differentiation coefficient F_{st} in the classical island model for plants [58]. Eq. (11) can be rewritten as $E(p_A^i)=\left(Q_A+\frac{(i-1)(1-Q_A)F}{1+(i-2)F}\right)E(p_A^{i-1})$, the same as Wright's expression except for plant species here ([11], p.450). Eq. (11) represents the steady-state moments of allelic frequency under the balance of migration-genetic drift, different from Robertson's results in a progressive inbreeding process [59]. Replacement of subscripts A with B in Eq. (11) yields $E(p_B^i)$ (i = 1, 2, 3, 4).

Substituting $g=p_Ap_B$ and $g=D_{AB}$ separately in Eq. (3) to yield two equations, I can obtain

$$E(D_{AB})=\frac{4N\tilde{m}(1-r)\bar{D}_{AB}}{4Nr+(1-r)(4N\tilde{m}+1)}, \tag{12}$$

$$E(p_Ap_B)=Q_AQ_B+\frac{(1-r)^2\bar{D}_{AB}}{4Nr+(1-r)(4N\tilde{m}+1)}. \tag{13}$$

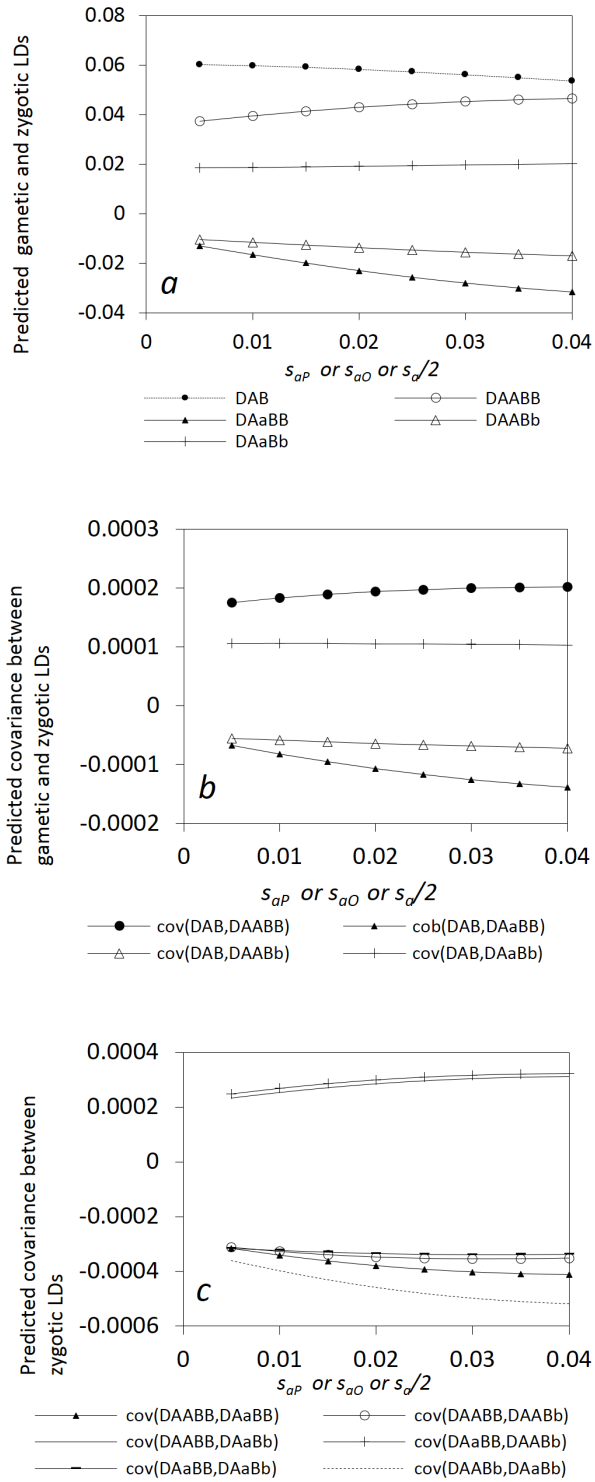

Figure 5. Genetic hitchhiking effects on the steady-state gametic and zygotic LDs, and other types of covariances. Gametic and zygotic LDs (a); covariances between gametic and zygotic LDs (b); and covariances between distinct zygotic LDs (c). Results are obtained from the analytical model in the section of Analytical Theory. Parameter settings are the immigration rate of pollen $m_P = 0.08$ and seeds $m_S = 0.04$, the effective population size = 100, the selection coefficients in the gametophyte and sporophyte stages $s_{aO} = s_{aP} = s_a/2$, and the degree of dominance = 0.5. The genotypic frequencies in the continent and initial island populations are 0.1225 for *AABB*, 0.105 for *AABb*, 0.0225 for *AAbb*, 0.105 for *AaBB*, 0.245 for *AB/ab*, 0.045 for *ab/aB*, 0.105 for *Aabb*, 0.0225 for *aaBB*, 0.105 for *aaBb*, and 0.1225 for *aabb*.
doi:10.1371/journal.pone.0080538.g005

Eq. (12) indicates that the expectation of gametic LD is equal to zero in the absence of LD in migrants, such as in a completely isolated population. Eq. (13) indicates that the expectation of joint allele frequencies at two loci is related to the gametic LD in migrants ($\bar{D}_{AB} \neq 0$) although expectations of individual allele frequencies are independent from each other (Eq. (11)).

Expectations of the remaining forty-four functions can be calculated in the following steps. Substitution of g in Eq. (3) by functions $p_A^{2-i}p_B^{2-j}(i,\ j=0,1;$ except $i=j=1)$, $p_A^{1-i}p_B^{1-j}D_{AB}$ $(i,\ j=0,1;$ except $i=j=1)$, and D_{AB}^2 can yield seven equations that can be used to numerically calculate their expectations: $E(p_A^{2-i}p_B^{2-j})(i,\ j=0,1;$ except $i=j=1)$, $E(p_A^{1-i}p_B^{1-j}D_{AB})(i,\ j=0,1;$ except $i=j=1)$, and $E(D_{AB}^2)$. Substitution of g in Eq. (3) by functions $p_A^3p_B^{3-i}(i=1,2)$, $p_A^{3-i}p_B^3(i=1,2)$, $p_A^2p_B^{2-i}D_{AB}(i=1,2)$, $p_A^{2-i}p_B^2D_{AB}(i=1,2)$, $p_AD_{AB}^2$, and $p_BD_{AB}^2$, yields ten equations to calculate their expectations. Substitution of g in Eq. (3) with two functions $p_A^4p_B$ and $p_A^3D_{AB}$ yields two equations to calculate $E(p_A^4p_B)$ and $E(p_A^3D_{AB})$. Substitution of g in Eq. (3) with two functions $p_Ap_B^4$, and $p_B^3D_{AB}$ yields two equations to calculate $E(p_Ap_B^3)$ and $E(p_B^3D_{AB})$. Substitution of g in Eq. (3) with functions $p_A^2p_B^4$, $p_A^4p_B^2$, $p_Ap_B^3D_{AB}$, $p_A^3p_BD_{AB}$, $p_A^2D_{AB}^2$, and $p_B^2D_{AB}^2$, yields six equations to calculate their expectations: $E(p_A^2p_B^4)$, …, and $E(p_B^2D_{AB}^2)$. Finally, substitution of g in Eq. (3) with the remaining seventeen functions $p_A^4p_B^4$, …, and D_{AB}^4, yields seventeen equations to calculate their expectations: $E(p_A^4p_B^4)$, …, and $E(D_{AB}^4)$.

The above order of g substitutions with different functions is sequentially arranged since calculations of the expected functions in the later equations need the expectations of the functions derived from the former equations. All these calculations can be done using *Mathematica* equation solution tool. Expectations of high-order LD functions, such as $E(D_{AaBb}^2)$, can also be calculated with additional equations by setting g in Eq. (3) with different functions. These are not explored further.

Once the expectations of the above fifty-four functions are available, the expectations of lower or the same order functions can be indirectly calculated. For instance, I can obtain $E(p_{AB}^*) = E(p_Ap_B) + (1-r)E(D_{AB})$ and $E(p_{AB}^*p_{ab}^*) = E(p_Ap_ap_Bp_b) + (1-r)(E(p_Ap_BD_{AB}) + E(p_ap_bD_{AB})) + (1-r)^2E(D_{AB}^2)$. The expectations of all possible zygotic LDs, the variances of zygotic LDs, and the covariances among different LDs at the steady state can be calculated according to Appendices S2 and S3.

For instance, the expectation of steady-state zygotic LDs for the genotype with double heterozygotes ($E(D_{AaBb})$ from $E(M(\delta_{DAaBb})) = 0$), its variance ($E(V(D_{AaBb})) = E(V(\delta_{D_{AaBb}}))$), and its covariance with gametic LD ($E(\mathrm{cov}(D_{AB}, D_{AaBb})) = E(\mathrm{cov}(\delta_{D_{AB}}, \delta_{D_{AaBb}}))$) are given by

$$\begin{aligned} E(D_{AaBb}) &= 2(1-r)E((1-2p_A)(1-2p_B)D_{AB}) + 4(1-r)^2E(D_{AB}^2) \\ &+ 2(m_S + m_P)\left(4E(p_Ap_ap_Bp_b) - E(p_{AB}^*p_{ab}^*) - E(p_{Ab}^*p_{aB}^*)\right) \\ &+ m_P\left(Q_{AB}^*E(p_{ab}^*) + Q_{ab}^*E(p_{AB}^*) + Q_{Ab}^*E(p_{aB}^*) + Q_{aB}^*E(p_{Ab}^*)\right. \\ &- 2(Q_AE(p_ap_Bp_b) + Q_aE(p_Ap_Bp_b) \\ &+ Q_BE(p_Ap_ap_b) + Q_bE(p_Ap_ap_B)) \\ &+ m_S(Q_{AaBb} - 2Q_{Aa}E(p_Bp_b) - 2Q_{Bb}E(p_Ap_a)) \end{aligned} \tag{14}$$

Table 4. Comparison between the simulation results and analytical model predictions under a neutral process*

	MC simulations	Analytical model
p_A	0.5079±0.0898	0.5
p_B	0.5102±0.0913	0.5
D_{AB}	0.0974±0.0399	0.0615
D_{AABB}	0.0623±0.0363	0.0360
D_{AaBB}	−0.0251±0.0312	−0.0100
D_{AaABb}	−0.0256±0.0312	−0.0100
D_{AaBb}	0.0462±0.0383	0.0200
$\mathrm{cov}(D_{AB},D_{AABB})$	$1.671\times10^{-4}\pm3.842\times10^{-5}$	1.689×10^{-4}
$\mathrm{cov}(D_{AB},D_{AaBB})$	$-0.909\times10^{-4}\pm4.71\times10^{-5}$	-0.634×10^{-4}
$\mathrm{cov}(D_{AB},D_{AABb})$	$-0.916\times10^{-4}\pm4.67\times10^{-5}$	-0.634×10^{-4}
$\mathrm{cov}(D_{AB},D_{AaBb})$	$1.774\times10^{-4}\pm5.876\times10^{-5}$	1.269×10^{-4}
$\mathrm{cov}(D_{AABB},D_{AaBB})$	$-3.164\times10^{-4}\pm8.517\times10^{-5}$	-2.921×10^{-4}
$\mathrm{cov}(D_{AABB},D_{AABb})$	$-3.167\times10^{-4}\pm8.527\times10^{-5}$	-2.921×10^{-4}
$\mathrm{cov}(D_{AABB},D_{AaBb})$	$2.318\times10^{-4}\pm8.527\times10^{-5}$	2.193×10^{-4}
$\mathrm{cov}(D_{AaBB},D_{AABb})$	$2.575\times10^{-4}\pm8.78\times10^{-5}$	2.267×10^{-4}
$\mathrm{cov}(D_{AaBB},D_{AaBb})$	$-2.977\times10^{-4}\pm1.002\times10^{-4}$	-3.237×10^{-4}
$\mathrm{cov}(D_{AABb},D_{AaBb})$	$-2.949\times10^{-4}\pm0.990\times10^{-4}$	-3.237×10^{-4}

*Parameter settings are the immigration rate of pollen $m_P=0.08$ and seeds $m_S=0.04$, the effective population size = 100, and the recombination rate = 0.05. The genotypic frequencies in the continent and the initial island populations are 0.1225 for *AABB*, 0.105 for *AABb*, 0.0225 for *AAbb*, 0.105 for *AaBB*, 0.245 for *AB/ab*, 0.045 for *ab/aB*, 0.105 for *Aabb*, 0.0225 for *aaBB*, 0.105 for *aaBb*, and 0.1225 for *aabb*. The steady-state simulation results (mean ± S_d) are obtained from 10000 independent runs.
doi:10.1371/journal.pone.0080538.t004

$$\begin{aligned}E(V(D_{AaBb})) &= \frac{1}{N}(4E(p_Ap_ap_Bp_b(1-2p_Ap_a)(1-2p_Bp_b)) \\ &+2(1-r)E((1-2p_A)(1-4p_Ap_a)(1-2p_B)(1-4p_Bp_b)D_{AB}) \\ &-16(1-r)^3E((1-2p_A)(1-2p_B)D_{AB}^3)-16(1-r)^4E(D_{AB}^4))\end{aligned} \quad (15)$$

$$\begin{aligned}E(Cov(D_{AB},D_{AaBb})) &= \frac{1}{N}(E(p_Ap_ap_Bp_b(1-2p_A)(1-2p_B)) \\ &+(1-r)E((1-2p_A)^2(1-2p_A)^2D_{AB}) \\ &+4(1-r)E(p_Ap_ap_Bp_bD_{AB}) \\ &-(1-r)^2E((1-2p_A)(1-2p_B)D_{AB}^2)-4(1-r)^3E(D_{AB}^3))\end{aligned} \quad (16)$$

Simulations confirm that the above analytical model performs well. The gametic and zygotic LDs and other covariances between two neutral loci can quickly reach steady-state distributions, reflecting the equilibrium among the effects of migration, recombination, and genetic drift. All analytical results are distributed within the range of one standard deviation of the simulation results (Table 4). Simulations also confirm that the expectations of D_{AABb} and D_{AaBB} and their covariances with gametic or other zygotic LDs are the same because both loci are neutral. This symmetry may help to reduce the number of expectations of distinct functions in theory.

Figure 6 shows that different patterns exist for the expectations of gametic and zygotic LDs and other covariances with the recombination rate. $E(D_{AB})$ decreases faster than the absolute expectations of zygotic LDs with the recombination rate in addition to their inequality in magnitude (Figure 6a). Figure 6b shows that the absolute expectations of the covaraines between gametic and zygotic LDs gradually decrease with the recombination rate for the genotypes with heterozygotes at one locus or two loci. $E(cov(D_{AB}, D_{AABB}))$ slightly decreases with the recombination rate, but does not approach zero in the presence of immigration that maintains gametic and zygotic LDs. Figure 6c shows that the covariances between different zygotic LDs are generally not as sensitive as some covariances between gametic and zygotic LDs to the change of linkage distance. $E(cov(D_{AABB}, D_{AaBb}))$ and $E(cov(D_{AaBB}, D_{AABb}))$ slightly decrease with the recombination rate, while the covariances between the zygotic LDs of the genotypes with one common genotype at a locus slightly increase with the recombination rate.

The above results indicate that a neutral process can generate a similar pattern between zygotic and gametic LDs along chromosomes, with strong LDs within short distances and weak LDs within long distances. The covariances between gametic and zygotic LDs or between distinct zygotic LDs are relatively insensitive to the linkage distance.

Discussion

In this study, I have developed the evolutionary theory of zygotic LDs in a local plant population, complementing the previous theories that mainly focus on the statistical issues [5], [6], [9]. The theory shows that evolutionary forces can generate different patterns among gametic and zygotic LDs, the covariances between gametic and zygotic LDs, and the covariances between different zygotic LDs. Zygotic LDs can be greater or smaller than, or comparable to gametic LD, depending on the major ecological and evolutionary processes involved in a local population. The covariances between gametic and zygotic LDs are more sensitive to the effects of mating system, linkage distance, and genetic drift, than to the effects of seed and pollen flow and selection. The covariances between different zygotic LDs are relatively robust to

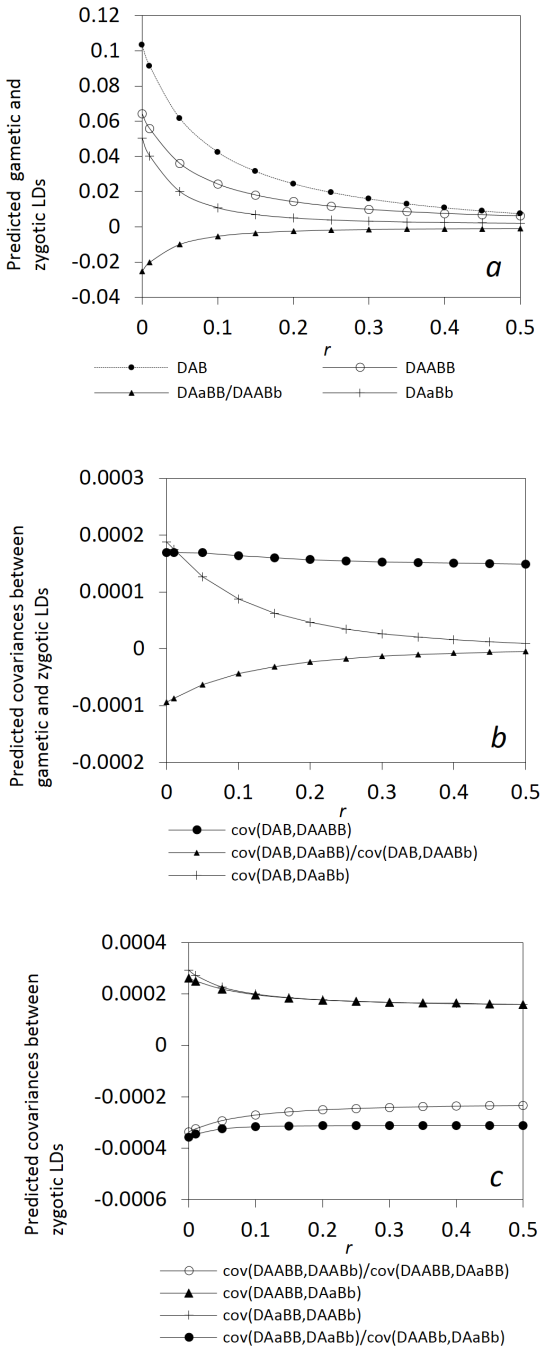

Figure 6. Effects of the linkage distance on the steady-state gametic and zygotic LDs and other types of covariances in a neutral process. Gametic and zygotic LDs (a); covariances between gametic and zygotic LDs (b); and covariances between distinct zygotic LDs (c). Results are obtained from the analytical model in the section of Analytical Theory. Parameter settings are the immigration rate of pollen $m_P = 0.08$ and seeds $m_S = 0.04$, the effective population size = 100, the selection coefficients in the gametophyte and sporophyte stages $s_{aO} = s_{aP} = s_a = 0$, and the degree of dominance = 0.0. The genotypic frequencies in the continent and initial island populations are 0.1225 for *AABB*, 0.105 for *AABb*, 0.0225 for *AAbb*, 0.105 for *AaBB*, 0.245 for *AB/ab*, 0.045 for *ab/aB*, 0.105 for *Aabb*, 0.0225 for *aaBB*, 0.105 for *aaBb*, and 0.1225 for *aabb*.
doi:10.1371/journal.pone.0080538.g006

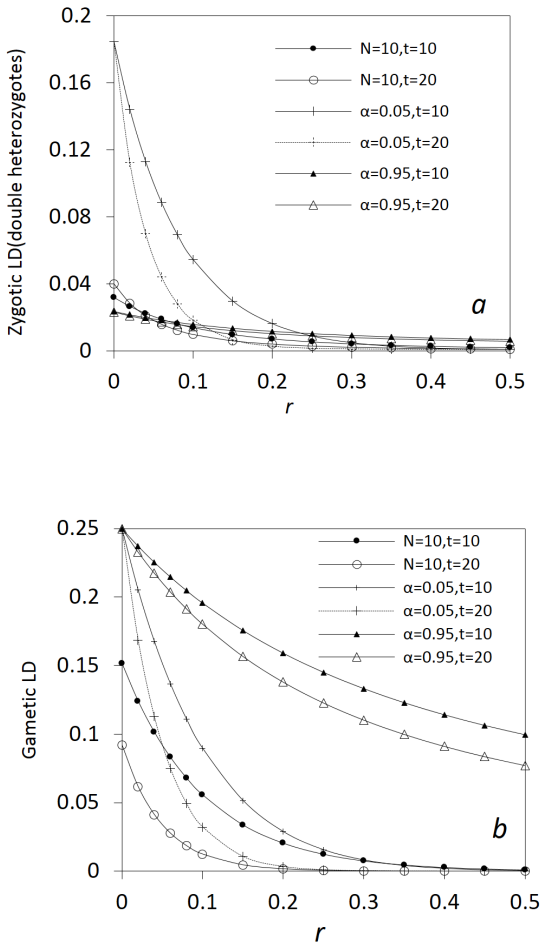

Figure 7. A comparison of transient gametic and zygotic LDs in a finite isolated population versus in an infinite population. Zygotic LDs in the finite population are calculated by synthesizing the theories of Robertson [59] and Ohta and Kimura [12]: $N = 10$, $t = 10$, and $N = 10$, $t = 20$ in (a). Gametic LD in the finite population is calculated from Hill and Robertson [8]: $N = 10$, $t = 10$, and $N = 10$, $t = 20$ in (b). Gametic and zygotic LDs in the infinite population are calculated from Weir and Cockerham [4]: $\alpha = 0.05$, $t = 10$; $\alpha = 0.05$, $t = 20$; $\alpha = 0.95$, $t = 10$; and $\alpha = 0.95$, $t = 20$ in (a) and (b). The initial settings for the finite population are $N = 10$ and the frequency of double heterozygotes (coupling) = 1 (gametic LD = 0.25 and the allelic frequency at each diallelic locus = 0.5). The initial setting for the infinite population is the frequency of double heterozygotes (coupling) = 1 (gametic LD = 0.25 and the allelic frequency at each diallelic locus = 0.5).
doi:10.1371/journal.pone.0080538.g007

the change in gene flow, selection, and genetic distance, but sensitive to the genetic drift effects and mating system. Consistent patterns exist for the covariances between zygotic LDs for the genotypes with a common genotype at one locus, or for the genotypes without any common genotype at each locus. These similarities and differences suggest the potential utility of zygotic LDs in revealing the ecological and evolutionary processes underlying the pattern of population genomic diversity at the diploid level.

It is important to understand that in a pure drift process, LD is transient in a completely isolated population of random mating. Expectations of both gametic and zygotic LDs are zero although the expectations of their squared values are nonzero [8], [44],

[60]. Zygotic LDs are smaller than gametic LDs in magnitude, but decay more slowly than gametic LDs with time [12], [46]. This is because the genotypic association may primarily arise from the effects other than the linkage distance although the linkage distance can affect the frequencies of genotypic recombinants. Weir and Cockerham[4] and Cockerham and Weir [3] have decomposed gametic and zygotic LDs in terms of descent measures for an infinite population with a mixed mating system. Both gametic and zygotic LDs decay faster in an infinite than in a finite population within a short linkage distance when the genetic drift effects are in the same order as the selfing rate ($\alpha\tilde{1}/2N_e$; e.g., Figure 7). A predominantly selfing population reduces the rates of decay of both gametic and zygotic LDs. When additional driving forces are involved, the above "null" expectation and the rates of decay in gametic and zygotic LDs could be changed [1],[2].

Note that the theory only addresses the constant immigration of seeds and pollen. In reality, a frequent situation is the stochastic migration of seeds and pollen due to the influences of biotic and abiotic factors [61], [62]. This occurs particularly when the source populations or the pollen and seed pools are unstable. Under this situation, the gametic and genotypic frequencies fluctuate in migrating seeds and pollen, and so do the gametic and zygotic LDs in migrants. Zygotic and gametic LDs can exhibit more fluctuations under the joint effects of genetic drift and stochastic migration. This can weaken the relationships between zygotic and gametic LDs. Nevertheless, the explored qualitative relationships between zygotic LDs and migration remain valid. How the stochastic migration of seeds and pollen affects the relative gametic and zygotic LDs remains unclear, and this forms a topic for further study.

Also, note that a plant mating system in a natural population may exhibit a dynamic property [25]. Mating system can be naturally changed through different ways [63], such as the change of pollen pool and the shift from wind to animal pollinations [64], [65]. Since zygotic LDs and other covariances are sensitive to the change of mating system (Figure 1), an unstable mating system enhances the fluctuation of these covariances. Nevertheless, the non-linear relationships between the selfing rate and zygotic LDs remain valid.

Apart from the above assumptions, the theory suggests several useful implications [1], [2]. First, the newly explored relationships between gametic and zygotic LDs under different evolutionary forces, not the purely statistical relationships [3], [5], [6], suggest their different or similar practical applications. Different patterns between gametic and zygotic LDs indicate that they can be applied for different purposes. Zygotic LDs provide additional information for inferring population history. Previous studies emphasize the use of gametic LD for this purpose [61], [66]. The present theory shows that zygotic LDs exhibit more diverse patterns in response to different driving forces, which can reinforce our inference on the major ecological and evolutionary processes.

The occurrence of a weak gametic LD combined with strong zygotic LDs suggests epistatic interactions at the diploid level (e.g., postzygotic isolation due to the Dobzhansky-Muller incompatibility [27], [28], [47]). The occurrence of a strong gametic LD combined with weak zygotic LDs suggests the involvement of non epistatic processes at the diploid level, including migration, linear-additive selection, and genetic drift processes. When both strong gametic and zygotic LDs arise from the tight linkage, they can be applied for the same purposes. For example, in analyzing normalized gametic and zygotic LDs in a human population, the same SNP markers exist when both gametic and zygotic LDs are very strong (say, the squares of normalized gametic and zygotic LDs >0.9; [67]). The relatively stable patterns in zygotic and gametic LDs and in covariances between gametic and zygotic LDs across multiple populations suggest the impacts of seed and pollen flow or weak selection. Patterns from multiple samples of a given population or from multiple different natural populations can strengthen such inferences.

Second, the theory provides a genetic basis of using zygotic LDs for QTL mapping that has been recently addressed [23]. A similar pattern between zygotic and gametic LDs with the linkage distance implies the common utility for QTL mapping. Zygotic LD-based QTL mapping can be conducted in nonrandom mating populations [23]. One caution is that spurious and unstable non-random associations can occur in natural populations under the influences of the driving forces other than the recombination process. This can influence the accuracy and precision of QTL mapping. QTL mapping based on the linkage maps from a single family, such as a half-sib family from a single tree or a full-sib family from a single cross, is not affected. However, the population-based linkage maps could be affected although this approach is commonly suggested to search for LDs within a short linkage distance at a finer scale [68], [69]. Thus, the patterns of zygotic LDs can be used to preliminarily screen markers for QTL mapping through a high criterion [46], or to effectively remove spurious LDs through a deliberate experiment [70]. This may improve QTL mapping with the population-based linkage maps.

Third, the theory aids in predicting the effects of seed and pollen flow on zygotic LDs in a local population. Previous studies use gametic LD to estimate gene flow in a specific case, such as in hybrid zones [71]. The present theory shows that gametic LD is more sensitive than zygotic LDs to either seed flow or pollen flow. Seed flow has greater effects than pollen flow on gametic LD. In natural populations of flowering plants, pollen flow is often more extensive than seed flow among mature populations, especially for the predominantly outcrossing species [72]. The cumulative effects on gametic LD from pollen flow could be substantial. The robust pattern of zygotic LDs to the impacts of seed or pollen flow enables their utility for inferring if gametic LD is generated by the forces other than migration. One extreme case is the admixture of two or more plant or animal populations, such as cross breeding, which results in the same consequence as that produced by a large proportion of immigrating seeds. This produces extensive gametic LDs rather than zygotic LDs [61], [66]. Only those tightly linked loci can maintain strong zygotic and gametic LDs [46]. Thus, the multilocus patterns of joint gametic and zygotic LDs can be used to judge if immigration is an important process to shape gametic LDs in local populations.

Fourth, the theory aids in assessing the selection mode (additive or epistatic) in the gametophyte and sporophyte stages in generating gametic and zygotic LDs. "Bulmer effects" mainly emphasize the impacts of selection on gametic LD [26], but gametic LD does not provide the information on the genotypic interaction at the diploid level. Extensive reports are recorded in the literature about the use of gametic LD for detecting selection signature along chromosomes [73]. So far, zygotic LDs have not been applied to detecting the genetic basis of adaption at the diploid level. In the linear additive-viability model, selection from the two stages is compounded. Gametic LD is greater than zygotic LDs in magnitude because selection affects gametic LD at each stage but affects zygotic LDs only in the sporophyte stage, similar to the effects of haploid pollen and diploid seed flow. However, in the presence of epistatic selection at the diploid level, some genotypes have zygotic LDs larger than gametic LD while other genotypes have zygotic LDs smaller than gametic LD. Such divergent patterns can aid in our inference on epistatic selection. One typical situation is a natural hybrid zone (a tension zone)[29]

where epistatic selection can cause zygotic LDs greater than gametic LD [41], which provides the information complementary to two non-allele interaction at the haploid level [27],[28], [30]. The joint patterns of gametic and zygotic LDs can be used to infer the selection mode (additive or epistatic) at the diploid level.

In addition, the genotypic interaction on fitness may arise from the dominance by dominance effects for D_{AaBb}, or the additive by dominance effects for D_{AABb} or D_{AaBB}, or the additive by additive effects for D_{AABB} at two loci. One further study is to assess the genetic mechanisms of these epistases in distinct zygotic LDs at the sporophyte stage.

Finally, it is of interest to discuss the utility of the covariances between distinct zygotic LDs since few studies have examined such high-order LDs [9], [10]. The present theory suggests one robust property of these high-order LDs, i.e. the presence of a consistent pattern for the genotypes with one common genotype at one locus or for the genotypes without any common genotype at each locus. This property can be used to effectively determine the impacts from migration, recombination, and additive weak selection, and to assess the effects of effective population size and/or a mating system. Given a stable effective population size and a stable mating system, a significant bias from the robust property implies epistatic selection (Table 2) or very diverse selection systems among genotypes. This requires further empirical verification with appropriate data collections.

Supporting Information

Appendix S1 Derivations of genotypic frequencies
(DOC)

Appendix S2 Average per-generation changes in allelic frequency, gametic and zygotic LDs
(DOC)

Appendix S3 Variances for the per-generation changes in allelic frequency, gametic and zygotic LDs
(DOC)

Appendix S4 Expectations of variances of zygotic LDs and the covariances between gametic and zygotic LDs
(DOC)

Acknowledgments

I sincerely appreciate three anonymous reviewers for very helpful comments that substantially improved the presentation of this article.

Author Contributions

Conceived and designed the experiments: XSH. Performed the experiments: XSH. Analyzed the data: XSH. Contributed reagents/materials/analysis tools: XSH. Wrote the paper: XSH.

References

1. Weir BS, Allard RW, Kahler AL (1972) Analysis of complex allozyme polymorphisms in a barley population. Genetics 72: 505–523.
2. Weir BS, Allard RW, Kahler AL (1974) Further analysis of complex allozyme polymorphisms in a barley population. Genetics 78: 911–919.
3. Cockerham CC, Weir BS (1973) Descent measures for two loci with some applications. Theoretical Population Biology 4: 300–330.
4. Weir BS, Cockerham CC (1973) Mixed self and random mating at two loci. Genetical Research 21: 247–262.
5. Yang RC (2002) Analysis of multilocus zygotic associations. Genetics 161: 435–445.
6. Yang RC (2003) Gametic and zygotic associations. Genetics 165: 447–450.
7. Robbins RB (1918) Some applications of mathematics to breeding problems iii. Genetics 3: 375–389.
8. Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theoretical and Applied Genetics 38: 226–231.
9. Weir BS (1996) Genetic Data Analysis II. Sinauer Associates Sunderland, MA.
10. Slatkin M (2008) Linkage disequilibrium-understanding the evolutionary past and mapping the medical future. Natature Review Genetics 9: 477–485.
11. Wright S (1969) Evolution and the Genetics of Populations. Vol. 2: The Theory of Gene Frequencies. Chicago, IL: The University of Chicago Press.
12. Ohta T, Kimura M (1969) Linkage disequilibrium due to random genetic drift. Genetical Research 13: 47–55.
13. Sved JA, Felman MW (1973) Correlation and probability methods for one and two loci. Theoretical Population Biology 4: 129–132.
14. Hernandez-Sanchez J, Haley CS, Woolliams JA (2004) On the prediction of simultaneous inbreeding coefficients at multiple loci. Genetical Research 83:113–120.
15. Hill WG, Weir BS (2007) Prediction of multilocus inbreeding coefficients and relation to linkage disequilibrium in random mating populations. Theoretical Population Biology 72: 179–185.
16. Haldane JBS (1949) The association of characters as a result of inbreeding and linkage. Annals of Eugenics 15: 15–23.
17. Bennett J, Binet F (1956) Association between Mendelian factors with mixed selfing and random mating. Heredity 10: 51–56.
18. Avery PJ, Hill WG (1979) Variance in quantitative traits due to linked dominant genes and variance in heterozygosity in small populations. Genetics 91: 817–844.
19. Ohta T (1980) Linkage disequilibrium between amino acid sites in immunoglobulin genes and other multigene families. Genetical Research 36: 181–197.
20. Sabatti C, Risch N (2002) Homozygosity and linkage disequilibrium. Genetics 160: 1707–1719.
21. Rosenberg NA, Blum MGB (2007) Sampling properties of homozygosity-based statistics for linkage disequilibrium. Mathematical Biosciences 208:33–47.
22. Jiang Q, Wang Z, Moore SS, Yang RC (2012) Genome-wide analysis of zygotic linkage disequilibrium and its components in crossbred cattle. BMC Genetics 13: 65.
23. Wu S, Yang J, Wu R (2010) Mapping quantitative trait loci in a non-equilibrium population. Statistical Applications in Genetics and Molecular Biology *9*, Article 32.
24. Weir BS (1979) Inferences about linkage disequilibrium. Biometrics 35: 235–254.
25. Mitton JB (1992) The dynamic mating system of conifers. New Forests 6: 197–216.
26. Bulmer MG (1971) The effect of selection on genetic variability. American Naturalist 105: 201–211.
27. Dobzhansky Th (1937) Genetics and the origin of species. Columbia University Press, New York.
28. Muller HJ (1942) Isolating mechanisms, evolution, and temperature. Biol. Symp. 6: 71–125.
29. Barton NH, Gale KS (1993) Genetic analysis of hybrid zones. In: Harrison RG (ed) Hybrid Zones and the Evolutionary Process, Oxford University Press: New York, pp 13–45.
30. Teeter KC, Payseur BA, Harris LW, Bakewell MA, Thibodeau LM, et al.(2008) Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Research 18: 67–76.
31. Snow AA, Spira TP (1996) Pollen-tube competition and male fitness in *Hibiscus moscheutos*. Evolution 50: 1866–1870.
32. Mitchell RJ (1997) Effects of pollen quantity on progeny vigor: evidence from the desert mustard *Lesquerella fendleri*. Evolution 51: 1679–1684.
33. Niesenbaum RA (1999) The effects of pollen load size and donor diversity on pollen performance, selective abortion, and progeny vigor in *Mirabilis jalapa* (Nyctaginaceae). American Journal of Botany 86, 261–268.
34. Sorensen FC (1982) The roles of polyembryony and embryo viability in the genetic system of conifers. Evolution 36:725–733.
35. Charlesworth D (1989) Why do plants produce so many more ovules than seeds? Nature 338: 21–22.
36. Haldane JBS (1932) The time of action of genes, and its bearing on some evolutionary problems. American Naturalist 66: 5–24.
37. Tanksley S, Zamir D, Rick CM (1981) Evidence for extensive overlap of sporophytic and gametophytic gene expression in *Lycopersicon esculentum*. Science 213: 453–455.
38. Clark AG (1984) Natural selection with nuclear and cytoplasmic transmission. I. A deterministic model. Genetics 107: 679–701.
39. Charlesworth D, Charlesworth B (1992) The effects of selection in the gametophyte stage on mutational load. Evolution 46: 703–720.
40. Mulcahy DL, SariGorla M, Mulcahy GB (1996) Pollen selection-past, present and future. Sexual Plant Reproduction 9: 353–356.
41. Hu XS (2005) Tension versus ecological zones in a two-locus system. Theoretical Population Biology 68: 119–131.
42. Fisher RA (1954) Statistical methods for research workers, Ed. 12, Oliver and Boyd, London.
43. Lewontin RC (1964) The interaction of selection and linkage I. general considerations: heterotic models. Genetics 49: 49–67.
44. Ohta T, Kimura M (1970) Development of associative overdominance through linkage disequilibrium in finite populations. Genetical Research 16: 165–177.

45. Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1991) Numerical Recipes in C: The Art of Scientific Computing. Cambridge: Cambridge University Press.
46. Hu XS (2011) Mating system and the critical migration rate for swamping selection. Genetics Research 93: 233–254.
47. Coyne JA, Orr HA (2004) Speciation. Sinauer Associates, Inc., Sunderland, MA.
48. Gavrilets S (1997) Hybrid zones with Dobzhansky-type epistatic selection. Evolution 51: 1027–1035.
49. Kimura M (1964) Diffusion models in population genetics. Journal of Applied Probability 1: 177–232.
50. Gardiner CW (1985) Handbook of stochastic methods for physics, chemistry and the natural sciences. Springer-Verlag, Berlin.
51. Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favorable gene. Genetical Research 23: 23–35.
52. Barton NH (2000) Genetic hitchhiking. Philosophical Transactions of the Royal Society 355: 1553–1562.
53. Nordborg M, Charlesworth B, Charlesworth D (1996) The effect of recombination on background selection. Genetical Research 67: 159–174.
54. Hu XS, He FL (2005) Background selection and population differentiation. Journal of Theoretical Biology 235: 207–219.
55. Hu XS (2008) Barriers to the spread of neutral alleles in the cytonuclear system. Evolution 62: 2260–2278.
56. Wolfram Research, Inc. (2008) Mathematica. Version 7.0, Champaign, IL.
57. Hudson RR (1998) Island models and the coalescent process. Molecular Ecology 7: 413–418.
58. Hu XS, Ennos RA (1999) Impacts of seed and pollen flow on population differentiation for plant genomes with three contrasting modes of inheritance. Genetics 152: 441–450.
59. Robertson A (1952) The effect of inbreeding on the variation due to recessive genes. Genetics 37:189–207.
60. Sved JA (1968) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theoretical Population Biology 2: 125–141.
61. Levin DA (1988) Consequences of stochastic elements in plant migration. American Naturalist 132: 643–651.
62. Adams WT, Hipkins VD, Burczyk J, Randall WK (1997) Pollen contamination trends in a maturing Douglas-fir seed orchard. Canadian Journal of Forest Research 27: 131–134.
63. Eckert CG, Kalisz S, Geber MA, Sargent R, Elle E, et al. (2010) Plant mating systems in a changing world. Trends in Ecology and Evolution 25: 35–43.
64. Stebbins GL (1970) Adaptive radiation of reproductive characteristics in angiosperms. (I): pollination mechanisms. Annual Review of Ecology and Systematics 1: 307–326.
65. Montoya D, Zavala MA, Rodrıguez MA, Purves DW (2008) Animal versus wind dispersal and the robustness of tree species to deforestation. Science 320: 1502–1504.
66. Zöllner S (2008) Population history and linkage disequilibrium. In: Encyclopedia of life sciences (ELS). John Wiley and Sons, Ltd., Chester.
67. Hu XS, Yeh FC, Wang Z (2011) Structural genomics: correlation blocks, population structure, and genome architecture. Current Genomics 12: 55–70.
68. Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nature Review Genetics 10: 381–391.
69. Hu XS, Goodwillie C, Ritland K (2004) Joining linkage maps using a joint likelihood function. Theoretical and Applied Genetics 109: 996–1004.
70. Liu T, Todhunter RJ, Lu Q., Schoettinger L, Li H, et al. (2006) Modeling extent and distribution of zygotic disequilibrium: implications for a multigenerational canine pedigree. Genetics 174: 439–453.
71. Mallet J, Barton N, Lamas MG, Santisteban CJ, Muedas MM, et al. (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in *Heliconius* hybrid zones. Genetics 124: 921–936.
72. Ennos RA (1994) Estimating the relative rates of pollen and see migration among plant populations. Heredity 72: 250–259.
73. Oleksyk TK, Smith MW, O'Brien SJ (2010) Genome-wide scans for footprints of natural selection. Philosophical Transactions of the Royal Society B 365: 185–205.