

# Mating system and the critical migration rate for swamping selection

XIN-SHENG HU\*

Department of Renewable Resources, 751 General Service Building, University of Alberta, Edmonton, Alberta T6G 2H1, Canada

(Received 14 October 2010; revised 12 January 2011; first published online 6 May 2011)

## Summary

Crow *et al.* (1990) and Barton (1992) have examined the critical migration rate for swamping selection in the nuclear system. Here, I use the same methodology to examine the critical migration rate in the cytonuclear system for hermaphrodite plants with a mixed mating system. Two selection schemes for a nuclear gene (heterozygote disadvantage and directional selection) and the directional selection scheme for organelle genes are considered. Results show that under random mating, the previous results are applicable to plant species by appropriate re-parameterization of the migration rate for nuclear and paternal organelle genes. A simple complementary relationship exists between seed and pollen flow in contributing to the critical migration rate. Under the mixed mating system, the critical migration rate of seeds and pollen for nuclear and paternal organelle genes can be changed due to the effects of selection and the cytonuclear linkage disequilibrium generated by migration and inbreeding. A negative but not complementary relationship exists between seed and pollen flow in contributing to the critical migration rate, varying with the mating system. Partial selfing can also adjust the critical seed flow for the maternal organelle gene, with a small critical migration rate for species of a high selfing rate. Both concordance and discordance among cytonuclear genes can occur under certain conditions during the process of swamping selection. This theory predicts the presence of various contributions of seed versus pollen flow to genetic swamping for plants with diverse mating systems.

## 1. Introduction

The joint effects of selection and migration have long been appreciated (Haldane, 1930; Wright, 1931), and their importance has received considerable examinations in evolutionary theory and conservation biology. In a simple case where a balance can take place between effects of selection (soft selection) and immigration, the allele frequency in the recipient population may vary from zero (extinction) to unity (fixation), depending on the relative strengths of selection and immigration (Wright, 1969, pp. 36–38). When the immigration rate is greater than a threshold value (the minimum migration rate required for fixing migrated alleles or genotypes), such a balance vanishes, and the immigrating allele can eventually swamp the locally dominant allele, irrespective of the adaptation or maladaptation of immigrating alleles to local habitats. This migration threshold is termed

as the critical migration rate by Crow *et al.* (1990) in addressing phase three of Wright's shifting balance theory (SBT; Wright, 1977, pp. 443–473), and its role in impeding population fitness peak shift is extensively studied in theory (Barton, 1992; Barton & Rouhani, 1993; Rouhani & Barton, 1993; Barton & Whittlock, 1997).

In flowering plants, this critical migration rate may include both haploid pollen and diploid seed flow. The existing theories are mainly developed for animal species except that Haldane (1930, p. 229) indicated the necessity of generalizing the critical ratio of selection to immigration to inbreeding plants. The contribution of seed and pollen dispersal to gene flow is related to both the mode of gene inheritance and the reproductive system of species (Ennos, 1994). For maternally inherited organelle genomes, gene flow is mediated by seed dispersal only. This can occur for chloroplast genomes in angiosperms and mitochondrial genomes in both most angiosperms and gymnosperms (Mogenson, 1996). For the nuclear and paternally inherited organelle genomes, gene flow can

\* Correspondence address: 1400 College Plaza, University of Alberta, Edmonton, AB, Canada T6G 2C8. Tel: 780-248-1739. Fax: -1900. e-mail: xin-sheng.hu@ualberta.ca

be mediated through both seed and pollen dispersal. Such differences in gene flow lead to unequal spreading rates in space, evidenced from differential introgression between cytonuclear genes in hybrid zones (Avisé, 1994, pp. 293–295). When combined with effects of natural selection, different critical migration rates are expected in the presence of unequal selection strengths among the three genomes, as implied from the discordance among cytonuclear clines (Hu & Li, 2002). In practical terms, the critical migration rate is of particular interest for insights into the contamination of natural forests resulting from the immigration of artificial plantations (genetically modified) with high frequencies of target genes (Ellstrand, 2003; Hu *et al.*, 2003). Contamination may come from genes on either nuclear or organelle genomes, or both, and the required migration rate for gene swamping in natural populations could be different through seed or pollen dispersal.

Ample evidence indicates that distinct population genetic structure, assessed using nuclear and/or organelle genome markers, exists in species of various mating systems (Ennos, 1994; Ennos *et al.*, 1999). A plant mating system directly affects pollen flow, and a plant population with a high outcrossing rate often receives a high immigration rate of alien pollen (Pakkad *et al.*, 2008). For a predominantly outcrossing species, pollen dispersal often plays a dominant role in bringing about gene flow among mature forest populations (Chen *et al.*, 2008), enhancing the migration rate of both nuclear and paternal organelle genes. For the predominant selfing species, pollen dispersal only contributes a small portion of gene flow, enhancing comparable migration rates among nuclear, paternal and maternal organelle genes. Thus, the mating system as an important agent is involved in altering the critical migration rate through adjusting effective pollen dispersal (Haldane, 1930).

In addition, the joint effects of migration and the mating system facilitate interaction between nuclear and organelle genomes. Previous theories show that seed and pollen dispersal can generate cytonuclear linkage disequilibrium (LD) in the mainland–island model (Asmussen & Schnabel, 1991; Schnabel & Asmussen, 1992) and in hybrid zones (Arnold, 1993). Such an outcome can also be implied from the presence of stable LD between selective nuclear loci in structured populations (Li & Nei, 1974; Slatkin, 1975). The cytonuclear LD as a biological barrier, generated by seed and pollen dispersal, can impede the spread of nuclear or organelle genes in space (Hu, 2008, 2010). Partial selfing reinforces the generation of cytonuclear LD in plant species (Asmussen *et al.*, 1987, 1989; Maroof *et al.*, 1992; Asmussen & Orive, 2000), similar to its function in generating LD between nuclear sites. Therefore, it is of both theoretical and practical significance to examine the similarity

and differences in the critical migration rate among cytonuclear genes.

The purpose of this study is to bring together different combinations of cytonuclear systems to show how mating pattern changes the critical migration rate of seeds and pollen. This necessarily generalizes the theory to the cytonuclear system from the previous studies on the solely nuclear system where LD is negligible under the assumption of weak selection and random mating (Crow *et al.*, 1990; Barton, 1992). The cytonuclear system cannot be approximated simply by setting the recombination rate as  $\frac{1}{2}$  in the nuclear system, as indicated from the study of comparing two-locus LD in finite populations (Fu & Arnold, 1992) or from the studies on spreading neutral alleles in structured populations (Hu, 2008, 2010). The mixed mating system can exert distinct effects on diploid nuclear genes from on haploid organelle genes due to different modes of inheritance in addition to generating cytonuclear LD. Also, the functional epistasis between cytonuclear genes cannot be simply approximated in theory by the epistasis between nuclear sites due to the asymmetric mode of inheritance between cytonuclear genes. Cytonuclear epistasis may take place at the different levels of metabolic pathways for the products from organelle and nuclear genomes (Elo *et al.*, 2003; Wolf, 2009), due to the long-term endosymbiotic co-evolution in a plant cell (Stoebe *et al.*, 1999; Rand *et al.*, 2004). Thus, it is important to consider the genetic swamping effects in shifting the fitness of the cytonuclear system.

In my analysis of the critical migration rate of seeds and pollen, three cytonuclear systems are separately addressed: the nuclear and maternally or paternally organelle genome system; the nuclear and maternal and paternal organelle genome system. In the following sections, I begin by describing the exact general model where the critical migration rates of seed and pollen can only be estimated through numerical simulations. I then separately examine each system under weak selection where the analytical approximations for critical migration rates are derived. These approximations are checked through the exact simulation results. Inferences on how the mating system shapes the critical migration rate of seeds and pollen are drawn from both analytical and simulation results under the cytonuclear system.

## 2. General assumptions

Consider one natural population of a hermaphrodite plant species, with constant immigration rates of pollen ( $m_p$ ) and seeds ( $m_s$ ) per generation from a source population. The source population may have either the same or different fitness from the recipient population. For simplicity, the reverse directional migration is not considered or is assumed to be not

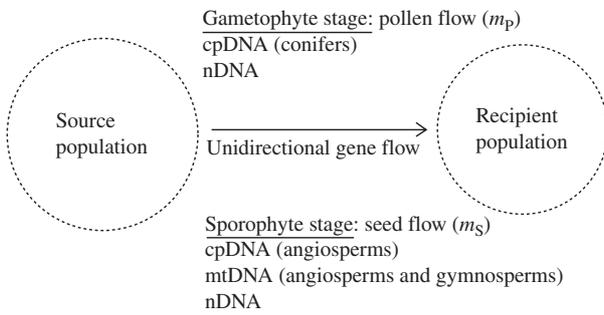


Fig. 1. A recipient population of a hermaphrodite plant species is subject to the influences of constant immigration of pollen and seeds from a source population. The source population is assumed to be stable in gametic and genotypic frequencies. At the gametophyte stage, chloroplast DNA (cpDNA) in conifers and nuclear DNA (nDNA) can be dispersed through pollen flow. At the sporophyte stage, all three genomes (mitochondrial DNA (mtDNA), cpDNA and nDNA), can be dispersed through seed flow. The present model considers the effects of unidirectional gene flow on swamping selection in the recipient population.

important (Fig. 1). Mutation rate is assumed to be very much smaller, and its effect is excluded. Each population size is assumed to be large so that genetic drift effect is negligible, similar to previous studies (Haldane, 1930; Crow *et al.*, 1990; Barton, 1992). The source population is assumed to be stable in gametic and genotypic frequencies. At the gametophyte stage, pollen and ovules are subject to natural selection (gametic selection) before combined to produce seeds. The same mating system in each population is assumed, with an arbitrary selfing rate. The plant life cycle follows a sequence of events: pollen and ovules generation, pollen flow, selection at the gametophyte stage, mixed mating, seed flow and selection at the sporophyte stage.

Consider three diallelic genes with contrasting modes of inheritance (bi-parental, paternal and maternal). Again, the three-gene cytonuclear system can naturally occur in some gymnosperms where chloroplast and mitochondrial genomes are paternally and maternally inherited, respectively (Mogenson, 1996). When the effects of one organelle genome are neglected, a two-gene cytonuclear system can occur between nuclear and chloroplast genomes in angiosperms or between nuclear and mitochondrial genomes in both angiosperms and gymnosperms. Selection strength can be either strong or weak, and cytonuclear epistasis at either the gametophyte or the sporophyte stage, or both, is allowed in theory.

### 3. Exact simulation model

Let  $A_1$  and  $A_2$  be the two alleles on the diploid nuclear genomes,  $B_1$  and  $B_2$  be the two alleles on the haploid organelle genome with paternal inheritance, and  $C_1$

and  $C_2$  be the two alleles on the haploid organelle genome with maternal inheritance. These genes are selectively non-neutral, and population fitness is decided by the joint cytonuclear genes. In the recipient population, let  $p_{A_iA_jB_kC_l}$ ,  $p_{A_iA_j}$ ,  $p_{B_k}$  and  $p_{C_l}$  be the frequencies of cytonuclear genotype  $A_iA_jB_kC_l$ , nuclear genotype  $A_iA_j$ , organelle genotypes  $B_k$  and  $C_l$  ( $i, j, k, l = 1, 2; i \leq j$ ), respectively. In order to extend the definitions of genotypic cytonuclear LD by Asmussen *et al.* (1987), the genotypic cytonuclear LD in the four-gene case,  $D_{A_iA_jB_kC_l}$ , can be expressed as

$$D_{A_iA_jB_kC_l} = p_{A_iA_jB_kC_l} - p_{A_iA_j}D_{B_kC_l} - p_{B_k}D_{A_iA_jC_l} - p_{C_l}D_{A_iA_jB_k}, \quad (1)$$

where  $D_{B_kC_l}$  is the LD between  $B_k$  and  $C_l$  ( $= p_{B_kC_l} - p_{B_k}p_{C_l}$ ),  $D_{A_iA_jC_l}$  is the LD between  $A_iA_j$  and  $C_l$  ( $= p_{A_iA_jC_l} - p_{A_iA_j}p_{C_l}$ ) and  $D_{A_iA_jB_k}$  is the LD between  $A_iA_j$  and  $B_k$  ( $= p_{A_iA_jB_k} - p_{A_iA_j}p_{B_k}$ ). The above expression can be readily derived using Bennett's method (Bennett, 1954). Some properties of cytonuclear LD are  $\sum_{i,j=1; i \leq j}^2 D_{A_iA_jB_kC_l} = \sum_{i,j=1; i \leq j}^2 D_{A_iA_jB_k} = \sum_{i,j=1; i \leq j}^2 D_{A_iA_jC_l} = \sum_{k=1}^2 D_{A_iA_jB_k} = \sum_{l=1}^2 D_{A_iA_jC_l} = \sum_{k=1}^2 D_{B_kC_l} = \sum_{l=1}^2 D_{B_kC_l} = 0$ . The following relationships between gametic and genotypic cytonuclear LDs hold:  $D_{A_iB_k} = D_{A_iA_iB} + D_{A_1A_2B_k}/2$ ,  $D_{A_iC_l} = D_{A_iA_iC_l} + D_{A_1A_2C_l}/2$  and  $D_{A_iB_kC_l} = D_{A_iA_iB_kC_l} + D_{A_1A_2B_kC_l}/2$ .

Similarly, let  $q_{A_iA_jB_kC_l}$ ,  $q_{A_iA_j}$ ,  $q_{B_k}$  and  $q_{C_l}$  be the frequencies of cytonuclear genotype  $A_iA_jB_kC_l$ , nuclear genotype  $A_iA_j$ , cytoplasmic genotypes  $B_k$  and  $C_l$  ( $i, j, k, l = 1, 2; i \leq j$ ) in the source population, respectively. Let  $\bar{D}_{B_kC_l}$ ,  $\bar{D}_{A_iA_jC_l}$  and  $\bar{D}_{A_iA_jB_k}$  be the LDs between  $B_k$  and  $C_l$ , between  $A_iA_j$  and  $C_l$  and between  $A_iA_j$  and  $B_k$ , respectively.  $\bar{D}_{A_iA_jB_kC_l}$  and some properties about cytonuclear LD in the source population can be derived in the same way as in the recipient population.

In the gametophyte stage, let  $w_{A_iB_k}$  and  $w_{A_iC_l}$  be the fitness for gametes  $A_iB_k$  in pollen and  $A_iC_l$  in ovules, respectively. The average fitness in pollen and ovules, denoted by  $w_P$  and  $w_O$ , respectively, can be calculated using the conventional method  $w_P = \sum_{i=1}^2 \sum_{k=1}^2 w_{A_iB_k} p_{A_iB_k}^*$  and  $w_O = \sum_{i=1}^2 \sum_{l=1}^2 w_{A_iC_l} p_{A_iC_l}^*$ , where  $p_{A_iB_k}^*$  and  $p_{A_iC_l}^*$  are the gametic frequencies after pollen dispersal. Let  $r_{A_iB_k} = w_{A_iB_k} / (w_{A_1B_k} + w_{A_2B_k})$ , the relative ratio of gametic fitness in pollen, and  $r_{A_iC_l} = w_{A_iC_l} / (w_{A_1C_l} + w_{A_2C_l})$ , the relative ratio of gametic fitness in ovules. In the sporophyte stage, let  $w_{A_iA_jB_kC_l}$  be the fitness for  $A_iA_jB_kC_l$  ( $i, j = 1, 2, i \leq j; k, l = 1, 2$ ). The average fitness in the sporophyte stage,  $w$ , can be calculated by  $w = \sum_{i=1}^2 \sum_{j=1, j \leq i}^2 \sum_{k=1}^2 \sum_{l=1}^2 w_{A_iA_jB_kC_l} p_{A_iA_jB_kC_l}^*$ , where  $p_{A_iA_jB_kC_l}^*$  is the genotypic frequency after seed dispersal.

Let  $\alpha$  ( $0 \leq \alpha \leq 1$ ) be the selfing rate. Following the life cycle mentioned above, the cytonuclear genotypic

frequency after selection in the sporophyte stage, denoted by  $p_{A_i A_j B_k C_l}^{**}$  ( $=w_{A_i A_j B_k C_l} p_{A_i A_j B_k C_l}^* / w$ ) for  $A_i A_j B_k C_l$  ( $i, k, l = 1, 2$ ), can be expressed as

$$\begin{aligned}
 p_{A_i A_j B_k C_l}^{**} &= (1 - m_S) \frac{w_{A_i A_j B_k C_l}}{w} \\
 &\times \left( (1 - \alpha) \frac{w_{A_i B_k} w_{A_i C_l}}{w_O w_P} ((1 - m_P) p_{A_i B_k} p_{A_i C_l} \right. \\
 &\quad \left. + m_P Q_{A_i B_k} p_{A_i C_l}) \right. \\
 &\quad \left. + \alpha (p_{A_i A_j B_k C_l} + r_{A_i B_k} r_{A_i C_l} p_{A_1 A_2 B_k C_l}) \right) \\
 &+ m_S \frac{w_{A_i A_j B_k C_l}}{w} Q_{A_i A_j B_k C_l}. \tag{2}
 \end{aligned}$$

Similarly, the general expression for  $p_{A_1 A_2 B_k C_l}^{**}$  ( $k, l = 1, 2$ ) is derived as

$$\begin{aligned}
 p_{A_1 A_2 B_k C_l}^{**} &= (1 - m_S) \frac{w_{A_1 A_2 B_k C_l}}{w} \\
 &\times \left( (1 - \alpha) \frac{w_{A_1 B_k} w_{A_2 C_l}}{w_O w_P} ((1 - m_P) p_{A_1 B_k} p_{A_2 C_l} \right. \\
 &\quad \left. + m_P Q_{A_1 B_k} p_{A_2 C_l}) \right. \\
 &\quad \left. + (1 - \alpha) \frac{w_{A_2 B_k} w_{A_1 C_l}}{w_O w_P} ((1 - m_P) p_{A_2 B_k} p_{A_1 C_l} \right. \\
 &\quad \left. + m_P Q_{A_2 B_k} p_{A_1 C_l}) \right. \\
 &\quad \left. + \alpha (r_{A_1 B_k} r_{A_2 C_l} + r_{A_2 B_k} r_{A_1 C_l}) p_{A_1 A_2 B_k C_l} \right) \\
 &+ m_S \frac{w_{A_1 A_2 B_k C_l}}{w} Q_{A_1 A_2 B_k C_l}. \tag{3}
 \end{aligned}$$

The above general model has several advantages. First, it allows an arbitrary level of selfing rate and can be used to assess the critical migration rate for any mating system. Second, no assumptions are imposed on the mode of joint selection for cytonuclear genotypes. Selection can be either weak or strong, or either the presence of cytonuclear epistasis or a purely linear additive–cytonuclear–viability model. Third, the general model considers natural selection at either the gametophyte or the sporophyte stage or both (Tanksley *et al.*, 1981; Clark, 1984). This may be of significance for some plant species where genes expressed in the sporophyte stage (~60% of structural genes) can also be expressed and potentially subject to selection in the gametophyte stage (Mulcahy *et al.*, 1996). Selection at the gametophyte stage can modify the critical migration rates for seeds and pollen. Finally, the model allows different cytonuclear systems, such as the system of nuclear and paternal (or maternal) organelle genomes or the system of sole organelle genomes, depending on the type of system in question.

Although the recursion expression for each allele frequency can be derived from the above general model, the migration rates of seeds and pollen are difficult to solve as the function of genotypic and

allelic frequencies and selection coefficients. Given a set of parameters, the above exact model can be applied to searching for the critical migration rates of seeds and pollen in the cytonuclear system. The simulation results are not separately presented but are used to check the analytical approximations under weak selection.

#### 4. Analytical model under weak selection

In order to derive the analytical expressions for the critical migration rates of seeds and pollen, a linear–additive–viability model with weak selection is considered. Let the gametic fitness be  $w_{A_i B_k} = 1 + s_{A_i} + s_{B_k}$  and  $w_{A_i C_l} = 1 + s_{A_i} + s_{C_l}$ , where  $s_{A_i}$ ,  $s_{B_k}$  and  $s_{C_l}$  are the selection coefficients for alleles  $A_i$ ,  $B_k$  and  $C_l$ , respectively. Similarly, let  $w_{A_i A_j B_k C_l} = 1 + s_{A_i A_j} + s_{B_k} + s_{C_l}$ , where  $s_{A_i A_j}$  is the selection coefficient for nuclear genotype  $A_i A_j$ . Under weak selection, all terms containing the second or higher orders of the selection coefficients are neglected. The immigration rates of seeds and pollen are assumed to be small as well so that the balance between the effects of migration and selection can be attained. The terms 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 and selection coefficient ( $s m_P$  or  $s m_S$ ) are neglected. For simplicity, the notation for alleles and subscripts is changed by  $A$  for  $A_1$  and  $a$  for  $A_2$ , ..., and  $c$  for  $C_2$ . The Appendix gives the recursion equations for allele frequencies and cytonuclear LD under a general mixed mating system.

Throughout this section, suppose that genotype  $AABC$  is fixed in the source population, and that genotype  $abc$  is initially fixed in the recipient population. Frequencies for  $A$ ,  $B$  and  $C$  are equal to unity in the source population, without cytonuclear LD, similar to the single locus case investigated by Barton (1992) for the nuclear system. In order to examine the swamping effects of migration, alleles  $A$ ,  $B$  and  $C$  that are adaptive to the source population are assumed to be maladaptive to the recipient population although other soft selection schemes can be examined in different scenarios. Two selection schemes for the migrating nuclear genes in the recipient population are considered: heterozygote disadvantage where the heterozygotes are less fit than the homozygotes ( $s_{AA} = 0$ ,  $s_{Aa} = -s_2$ , and  $s_{aa} = 0$ ), and directional selection where allele  $A$  is maladaptive ( $s_{AA} = -2s_2$ ,  $s_{Aa} = -s_2$  and  $s_{aa} = 0$ ). Selection in the gametophyte stage is set as  $s_A = -s_1$  and  $s_a = 0$  with directional selection, but is not considered in the scheme of heterozygote disadvantage. Selection coefficient for the organelle genomes is set as  $-s_B$  for  $B$  and 0 for  $b$ , and  $-s_C$  for  $C$  and 0 for  $c$  in both the gametophyte and sporophyte stages. The consequence for the migration swamping effects is that alleles  $A$ ,  $B$  and  $C$  are fixed in

the recipient population. In the following, the critical migration rates are analytically examined in three different cytonuclear systems. In each system, two specific mating systems ( $\alpha=0$  and 1) are firstly presented, and a mixed mating system is then examined.

(i) Nuclear and maternally inherited organelle genome system

(a) Random mating

The theory in this subsection is applicable to the nuclear and mitochondrial genome system in flowering plants or to the nuclear and chloroplast genome system in angiosperms with complete outcrossing ( $\alpha=0$ ). In the presence of heterozygote disadvantage, the change for the nuclear allele frequency,  $\Delta p_A = p_A^{**} - p_A$ , is derived from eqn (A1) in the Appendix:

$$\Delta p_A = (m_S + m_P/2)p_a - (p_a - p_A)s_2 p_A p_a - s_C D_{AC}. \quad (4)$$

The change for maternal organelle allele  $C$ ,  $\Delta p_C = p_C^{**} - p_C$ , can be derived from eqn (A4) in the Appendix:

$$\Delta p_C = m_S p_c - 2s_C p_C p_c - (p_a - p_A)s_2 D_{AC}. \quad (5)$$

Note that  $2s_C$  in the above equation comes from selection in both the gametophyte and sporophyte stages, and it becomes  $s_C$  in the absence of gametic selection. It can be seen that the frequencies of alleles  $A$  and  $C$  can consistently increase as long as their changes are greater than zero ( $\Delta p_A > 0$  and  $\Delta p_C > 0$ ), leading to the eventual fixation of alleles  $A$  and  $C$  in the recipient population.

Suppose that cytonuclear LD is mainly maintained by migration, the recombination rate ( $=1/2$ ) and the mating system. This consideration is similar to LD between nuclear sites in the hybrid zone studies where LD can be approximated by the balance between migration and recombination (Kruuk *et al.*, 1999; Barton & Shpak, 2000; Hu, 2005). Here, the mating system plays a role similar to recombination in the nuclear system in maintaining cytonuclear LD. A high selfing rate (or a high outcrossing rate) enhances (or erodes) cytonuclear LD. From eqn (A7) in the Appendix, the steady-state cytonuclear LD is approximated by  $D_{AC} \approx 2m_S p_a p_c$ , which is of the order similar to  $m_S$ .

Substituting  $D_{AC}$  into eqn (4) at the steady state and neglecting the term with the order of migration times selection yield

$$m_S + m_P/2 = (p_a - p_A)s_2 p_A. \quad (6)$$

The critical migration rate for joint seeds and pollen is  $(m_S + m_P/2)^* = s_2/8$ , essentially the same as Barton's result (Barton, 1992, p. 552) except that migration

here refers to the compounded seed and pollen flow. For a given selection coefficient  $s_2$ , the sum of  $m_S$  and  $m_P$  is a constant. A simple complementary relationship exists between seed and pollen flow in contributing to the total critical migration rate. Similarly, substituting  $D_{AC}$  into eqn (5) at the steady state and neglecting the term with the order of migration times selection yield

$$m_S = 2s_C p_C. \quad (7)$$

The critical migration rate for seeds is  $m_S^* = 2s_C$ , the maximum value of function  $2s_C p_C$ . It can be seen that the effects of cytonuclear LD can be neglected under random mating with weak selection.

In order to check the above analytical results, exact simulations are conducted according to the general model described in the preceding section. The critical migration rates are calculated using the binary search method detailed by Barton (1992, p. 557). Given the settings of selection coefficients, an analytical estimate of the migration rate is used as the starting value and then adjusted by the binary search procedure. When the difference between the upper (starting fixation of migrated alleles) and lower (sufficiently close to fixation of migrated alleles) values is smaller than their average divided by 20, the average of the upper and lower values is used as the estimate, ensuring the estimate within 5% of the true value. The dynamics for allele frequency is considered to reach steady state when its changes within 10 consecutive generations are sufficiently small ( $< 10^{-8}$ ; Crow *et al.*, 1990).

Figure 2a shows the comparisons between the exact simulation and analytical approximation results for the maternal organelle gene. Results indicate that slight biases exist when selection coefficients are large, but good agreement between them exists when selection coefficients are small. Figure 2b shows the comparisons between the exact simulation and analytical approximation results for the nuclear gene. Since seed and pollen migration rates are compounded, the critical migration rates of seeds are estimated, given a migration rate of pollen, or the critical migration rates of pollen are estimated, given a migration rate of seeds. In each case, the analytical approximation performs well for the compounded migration rates of seeds and pollen (Fig. 2b).

The relationships between  $m_S^*$  and  $(m_S + m_P/2)^*$  are complex in fixing the immigrated maternal and nuclear alleles.  $m_S^*$  for the maternal organelle gene may be smaller than  $(m_S + m_P/2)^*$  in the presence of weaker selection against allele  $C$  than against nuclear allele  $A$ . If the migration rate of seeds for both nuclear and maternal organelle genes is equal to  $m_S^*$ , the migration rate of pollen may be estimated as  $(1/4 - 4\gamma)s_2$ , where  $\gamma = s_C/s_2$ . When the selection against allele  $C$  is much stronger than selection against allele  $A$ ,  $m_S^*$  may be greater than  $(m_S + m_P/2)^*$ , and maternal organelle

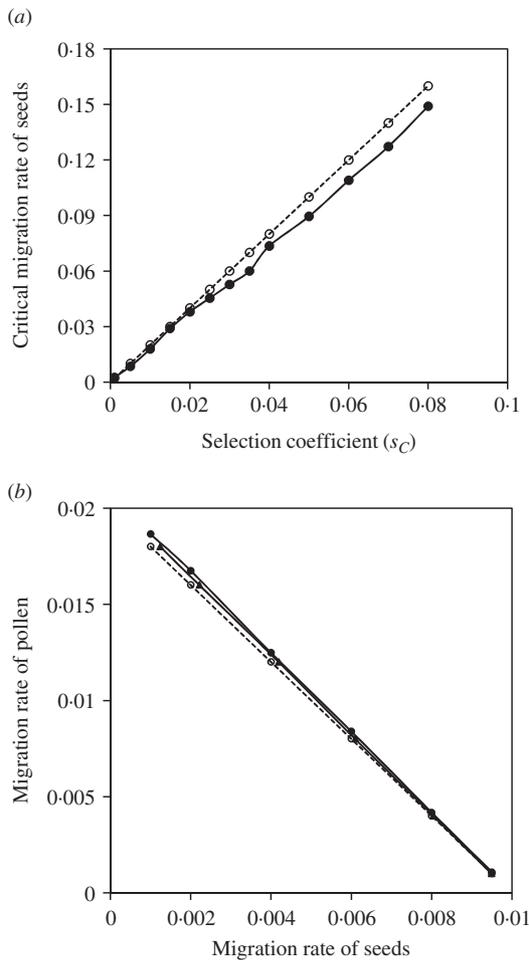


Fig. 2. (a) The critical migration rate of seeds as a function of selection strength for the maternal organelle gene. The dashed line with opened circles gives the analytic approximation,  $m_S^* = 2s_C$  from eqn (3). The solid line with closed circles is the exact simulation result. (b) A comparison of estimates of the critical migration rate from exact simulation and from analytical approximation. The dashed line with opened circles is the analytical approximation,  $(m_S + m_P/2)^* = s_2/8$  from eqn (4). The line with closed circles is the exact simulation result for the critical migration rate of pollen, given a set of migration rates of seeds. The line with closed triangles is the exact simulation result for the critical migration of seeds, given a set of migration rate of pollen. The selection coefficient in each case is set as  $s_2 = -0.08$ .

allele *C* is expected to be fixed earlier than nuclear allele *A*.

A concordance between nuclear and maternal allele frequencies is possible when certain conditions regarding selection strength, migration and the mating system are met. When a coincidence between cytonuclear allele frequencies occurs, let  $p_A = p_C = p$ . The recursion equation for the change of  $p$  can be approximated by combining eqns (A1) and (A4) in the Appendix. At the steady state, we obtained  $2m_S + m_P/2 = 2(1 + 2\gamma - 2p)ps_2$ . The critical value for compounded migration rate of seeds and pollen

$(m_S + m_P/4)^*$  is derived as  $(1 + 2\gamma)^2 s_2/16$ , which is difficult to numerically check using the exact simulation model described in the preceding section.

With directional selection, the effects of cytonuclear LD are negligible in altering allele frequency. The steady-state equation for the frequency of maternal organelle allele remains the same as eqn (7), yielding the same critical migration rate, i.e.  $m_S^* = 2s_C$ . The steady-state equation for the nuclear allele frequency becomes

$$m_S + m_P/2 = (1 + \rho)s_2 p_A, \tag{8}$$

where  $\rho = s_1/s_2$ , the ratio of selection coefficients between the gametophyte and sporophyte stages. The critical migration rate for compounded seeds and pollen flow is  $(m_S + m_P/2)^* = (1 + \rho)s_2$ , greater than the results in the scheme of heterozygote disadvantage. It reduces to the result obtained by Barton (1992, p. 552) in the absence of gametic selection ( $\rho = 0$ ). Here,  $m_S^*$  and the critical migration rate for compounded seed and pollen flow,  $(m_S + m_P/2)^*$ , are independently responsible for swamping maternal and nuclear alleles.  $m_S^*$  may be greater than  $(m_S + m_P/2)^*$  under certain conditions. Exact simulations confirm a very good performance for the analytical estimates of critical migration rate of pollen, given a migration rate of seeds, or the analytical estimates of critical migration rate of seeds, given a migration rate of pollen (data not given here).

When a coincidence between cytonuclear allele frequencies ( $p_A = p_C = p$ ) occurs, the steady-state equation for  $p$  can be approximated by combining eqns (A1) and (A4) in the Appendix, i.e.  $2m_S + m_P/2 = (1 + \rho + 2\gamma)ps_2$ . The critical migration rate for compounded seed and pollen flow becomes  $(m_S + m_P/4)^* = (1 + \rho + 2\gamma)s_2/2$ , unequal to the results in the scheme of heterozygote disadvantage.

(b) Complete selfing

Complete selfing ( $\alpha = 1$ ) removes the impacts of pollen flow. Cytonuclear LD is mainly generated by seed flow and selfing. From eqn (A7) in the Appendix,  $D_{AC}$  is approximated by  $p_a p_c$ . In the scheme of heterozygote disadvantage, the change for the frequency of allele *C* is

$$\Delta p_C = m_S p_c - s_C p_C p_c - s_2 D_{Ac} / 2. \tag{9}$$

In the above equation, the selection component comes from the selection in the sporophyte stage. From eqn (A11) in the Appendix, the genotypic cytonuclear LD can be approximated by  $D_{Ac} = 2(1 - \alpha)D_{AC}/(2 - \alpha) + \Lambda_1(s, m)$ , where  $\Lambda_1(s, m)$  is the function of the first-order migration rate and selection coefficients. The third term on the right-hand side of eqn (9) is negligible according to the assumption of weak

selection. Thus, the critical migration rate for seeds is  $m_S^* = s_C$ , a half value of the results under random mating due to the removal of selection effects in the gametophyte stage.

From eqn (A1) in the Appendix, the steady-state equation for the nuclear allele frequency becomes

$$0 = m_S p_a - s_2(p_a - p_A)p_{Aa}/4 - s_C D_{AC}. \tag{10}$$

According to eqn (A2) in the Appendix, the steady-state heterozygosity at the nuclear site can be approximated by  $p_{Aa} = 2(1 - \alpha)p_A p_a / (1 - \alpha/2) + \Lambda_2(s, m)$ , where  $\Lambda_2(s, m)$  is the function of the first-order migration rate and selection coefficients. Substituting  $p_{Aa}$  and  $D_{AC}$  in eqn (10) yields the same steady-state equation as the maternal allele frequency, with the critical migration rate for seeds,  $m_S^* = s_C$ , irrespective of the selection pressure at the nuclear site. This indicates the coincidence between maternal and nuclear allele frequencies, i.e.  $p_A = p_C = p$ , with approximately the same speed towards fixation.

Figure 3a shows the same trajectory for the change of migrated maternal (C) and nuclear (A) allele frequencies in the recipient population, calculated from the exact general model. The migrated alleles gradually replace the pre-existed alleles until fixation. Figure 3b shows a good agreement between the simulation results and the analytical approximation  $m_S^*$  under the complete selfing.

With directional selection at the nuclear site, the steady-state equation for the maternal organelle allele frequency is different from eqn (9),

$$0 = m_S p_c - s_C p_c p_c - 2s_2 D_{AC}. \tag{11}$$

By substituting  $D_{AC}$  into eqn (11), the critical migration rate for seeds is derived as  $m_S^* = (2 + \gamma)s_2$ . The steady-state equation for the nuclear allele frequency from eqn (A1) in the Appendix after substituting  $D_{AC}$  by  $p_a p_c$  yields the same critical migration rate. Simulation results confirm that  $m_S^* = (2 + \gamma)s_2$  is in very good agreement with the estimate from the exact model (data are not given here). Furthermore, a coincidence between maternal and nuclear allele frequencies occurs with this critical migration rate. Compared with the results for heterozygote disadvantage, a higher critical migration rate is needed for swamping selection with directional selection.

(c) Mixed mating system

From eqn (A7) in the Appendix, the steady-state cytonuclear LD can be approximated by

$$D_{AC} = \frac{2m_S}{1 - \alpha + (1 + \alpha)m_S} p_a p_c. \tag{12}$$

Equation (12) basically reflects the balance between the effects of migration and selfing that generate

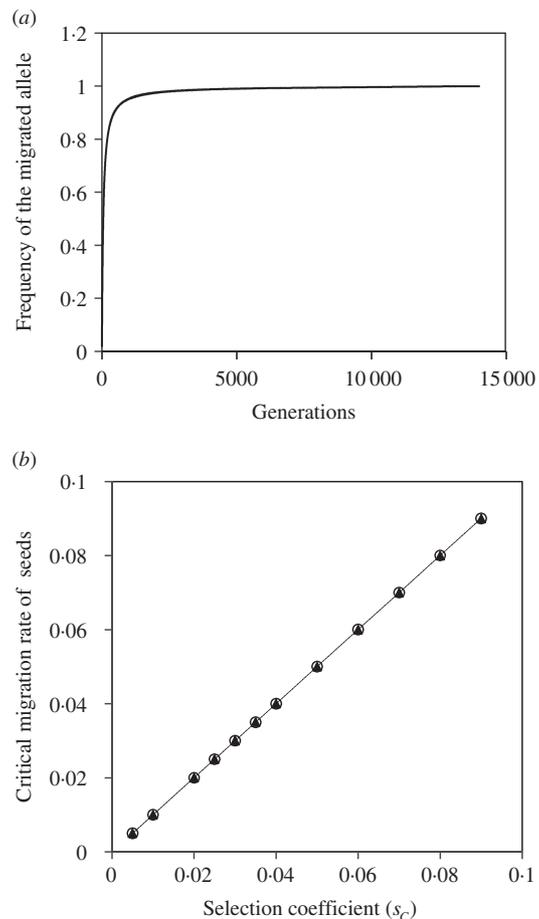


Fig. 3. (a) A coincidence for the change of migrated maternal organelle (C) or nuclear (A) allele frequencies was observed from the exact general model in the complete selfing system, with the parameter settings  $s_C = 0.02$  and  $m_S = 0.02$ . (b) Critical migration rate of seeds as a function of selection strength. The dashed line with opened circles gives the analytical approximation. The closed triangles are the exact simulation results.

cytonuclear LD and the effects of recombination (1/2) that erodes cytonuclear LD, analogous to LD in the nuclear system except for the inclusion of the mating system effects (Hu, 2005, p. 121). This is an intermediate case of the above two specific cases. In random mating ( $\alpha = 0$ ), eqn (12) can be approximated by  $D_{AC} \approx 2m_S p_a p_c$ . In the complete selfing case ( $\alpha = 1$ ), eqn (12) can be approximated by  $D_{AC} = p_a p_c$ . For many plant species with the mixed mating system, the condition of  $1 - \alpha \gg (1 + \alpha)m_S$  is appropriate. Numerical assessment indicates that plants of wide range of selfing rates meet this condition, given that migration rate is of the order similar to weak selection. Under this case, cytonuclear LD can be further simplified as  $D_{AC} \approx 2m_S p_a p_c / (1 - \alpha)$ . However, the condition of  $1 - \alpha \ll (1 + \alpha)m_S$  is applicable to the plant species with a predominant selfing system, and this leads to the approximation of  $D_{AC} = 2p_a p_c / (1 + \alpha)$ . For a more accurate calculation, these conditions are not separately addressed.

In the scheme of heterozygote disadvantage, substitution of  $D_{AC}$  into eqn (A4) in the Appendix yields the steady-state equation for the maternal organelle allele frequency:

$$\left(1 - \frac{4(1-\alpha)}{(2-\alpha)(1-\alpha+(1+\alpha)m_S)} s_2 p_a (p_a - p_A)\right) m_S = (2-\alpha) s_C p_C. \quad (13)$$

The maximum value of the right-hand side of eqn (13) is  $(2-\alpha)s_C$ . The allele frequency  $p_C$  consistently increases until the fixation of allele  $C$  as long as the left-hand side of eqn (13) is greater than  $(2-\alpha)s_C$ . However, the left-hand side of eqn (13) as a function of  $p_a$  indicates that an array of migration rate of seeds can be generated to meet the above condition. Note that searching for the critical migration rate  $m_S^*$  is different from searching for the maximum  $m_S$  as the function of  $p_A$  and  $p_C$ . Here, consider the critical migration rate of seeds at the point  $p_a=1/4$ , where  $-p_a(p_a-p_A)$  has a maximum value of  $1/8$ . Given the selfing rate and selection coefficients, the critical migration rate of seeds can be numerically solved (e.g., using the iterative method) from

$$\left(1 + \frac{(1-\alpha)s_2}{2(2-\alpha)(1-\alpha+(1+\alpha)m_S^*)}\right) m_S^* = (2-\alpha)s_C. \quad (14)$$

Figure 4a shows that the estimates of critical migration rate of seeds from the exact model with the binary search method are in good agreement with the results predicted from eqn (14) under various selfing rates.

From eqn (A1) in the Appendix and the use of the approximation for  $D_{AC}$ , the steady-state equation for the nuclear allele frequency can be written as

$$\left(1 - \frac{2s_C}{1-\alpha+(1+\alpha)m_S} p_c\right) m_S + \frac{1-\alpha}{2} m_P = (1-\alpha) \left(1 + \frac{\alpha^2}{2-\alpha}\right) s_2 (1-2p_A) p_A. \quad (15)$$

This compounded critical migration rate changes with the mating system. The maximum value for the right-hand side of eqn (15) (as the function of  $p_A$ ) is  $\frac{1}{8}(1-\alpha)(1+\frac{\alpha^2}{2-\alpha})s_2$  at  $p_A=1/4$ . The proportion of seed flow in the left-hand side of eqn (15) is the function of  $p_c$ . Thus, the critical migration rate for the compounded seed and pollen flow is expressed as

$$\left(\left(1 - \frac{2s_C}{1-\alpha+(1+\alpha)m_S}\right) m_S + \frac{1-\alpha}{2} m_P\right)^* = \frac{1}{8}(1-\alpha) \left(1 + \frac{\alpha^2}{2-\alpha}\right) s_2. \quad (16)$$

Given the migration rate of seeds (or pollen), the critical migration rate of pollen (or seeds) can be calculated from eqn (16). Given a mating system and

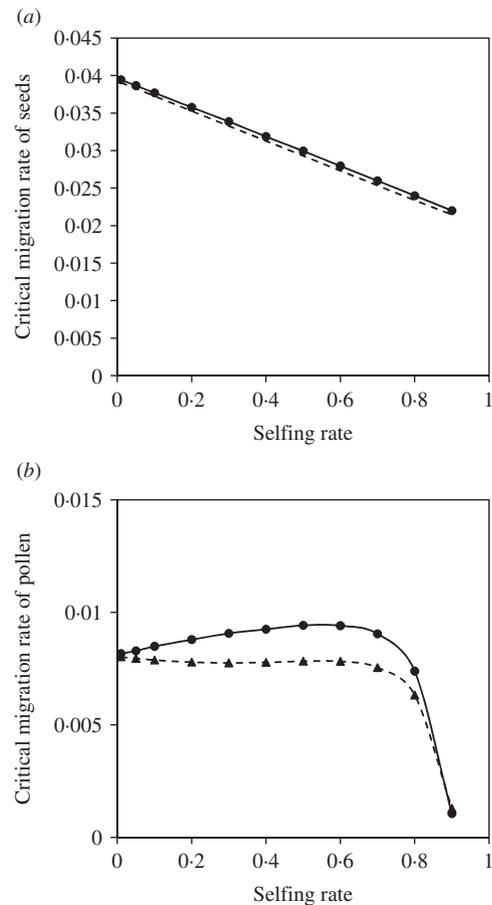


Fig. 4. (a) Critical migration rate of seeds as a function of selfing rate for the maternal organelle allele. The dashed line gives the analytical approximation from eqn (14). The line with closed circles is the exact simulation result. In each case, parameter settings are the migration rate of pollen  $m_P=0.01$ , the selection coefficient  $s_C=0.02$  and  $s_2=0.08$ . (b) Critical migration rate of pollen as a function of selfing rate for the nuclear allele. The dashed curve with closed triangles gives the analytical approximation from eqn (16). The curve with closed circles is the exact simulation result. In each case, parameter settings are the migration rate of seeds  $m_S=0.001$ , the selection coefficient  $s_C=0.01$  and  $s_2=0.04$ .

selection coefficients, the right-hand side of eqn (16) is constant. Numerical analysis shows that the migration rate of pollen required to satisfy eqn (16) reduces with an increase in migration rate of seeds, i.e. a negative but not complementary relationship between seed and pollen flow. Figure 4b shows the estimates of the critical migration rate of pollen under a fixed migration rate of seeds and various selfing rates. Estimates from the exact simulation results are slightly higher than those predicted from eqn (16) although the similar changing patterns with the selfing rate exist between them.

When eqns (14) and (16) are jointly considered, the maximum value of the right-hand side of eqn (14) may be greater than the maximum value of the right-hand

side of eqn (16) except for  $s_2 \gg 16s_C$ . This strict condition  $s_2 \gg 16s_C$  is applicable to the maternal alleles whose selection coefficients are much smaller than those of nuclear alleles. The more frequent situation is that the migration rate of seeds from eqn (14) that leads to the fixation of allele *C* can also lead to the fixation of nuclear allele *A*, without the need of pollen flow, which has been confirmed through exact simulations (data not given here).

When a coincidence between cytonuclear allele frequencies ( $p_A = p_C = p$ ) occurs, the steady-state equation for *p* can be obtained by combining eqns (A1) and (A4) in the Appendix:

$$(1 + f_1 s_2) m_S + \frac{1 - \alpha}{4} m_P = \frac{1}{2} p \left( (2 - \alpha) s_C - (1 - \alpha) \left( 1 + \frac{\alpha^2}{2 - \alpha} \right) (2p - 1) s_2 \right), \tag{17}$$

where  $f_1 = -\left( p\gamma - \frac{2(1-\alpha)}{2-\alpha} p(2p-1) \right) \frac{1}{1-\alpha+(1+\alpha)m_S}$ . Let  $\lambda_1 = \frac{(2-\alpha)^2 \gamma}{(1-\alpha)(2-\alpha+\alpha^2)}$ . It can be shown that the maximum value of the right-hand side of eqn (17) is  $\frac{1}{16}(1-\alpha) \left( 1 + \frac{\alpha^2}{2-\alpha} \right) (1 + \lambda_1)^2 s_2$  at the point  $p = (1 + \lambda_1)/4$ . Similarly, the minimum migration rate of seeds can be obtained by setting the maximum  $f_{1max}$  in the left-hand side of eqn (17), i.e.  $f_{1max} = \frac{1-\alpha}{4(2-\alpha)} \left( 1 - \frac{(2-\alpha)\gamma}{2(1-\alpha)} \right) \frac{1}{1-\alpha+(1+\alpha)m_S}$ . The critical migration rate of seeds or pollen can be numerically calculated from the following equation:

$$\left( (1 + f_{1max} s_2) m_S + \frac{1 - \alpha}{4} m_P \right)^* = \frac{1}{16} (1 - \alpha) \left( 1 + \frac{\alpha^2}{2 - \alpha} \right) \times (1 + \lambda_1)^2 s_2. \tag{18}$$

With the change in the mating system and selection coefficients, a negative but not complementary relation exists between seed and pollen flow in contributing to the compounded critical migration rate. Figure 5 shows the change of the critical migration rate of pollen, predicted by the above equation, with the selfing rate. These estimates are unequal to those under the discordance case. Generally, a high critical migration rate of pollen is needed for species with a high selfing rate, given a fixed migration rate of seeds and selection strengths.

With directional selection at the nuclear site, the steady-state equation for the change of maternal organelle allele frequency is derived as

$$\left( 1 - \frac{2(1-\alpha)(1+\rho)+4\alpha}{1-\alpha+(1+\alpha)m_S} s_2 p_a \right) m_S = (2-\alpha) s_C p_C. \tag{19}$$

The critical migration rate for seeds,  $m_S^*$ , can be numerically calculated from eqn (19) by letting  $p_a = 1$

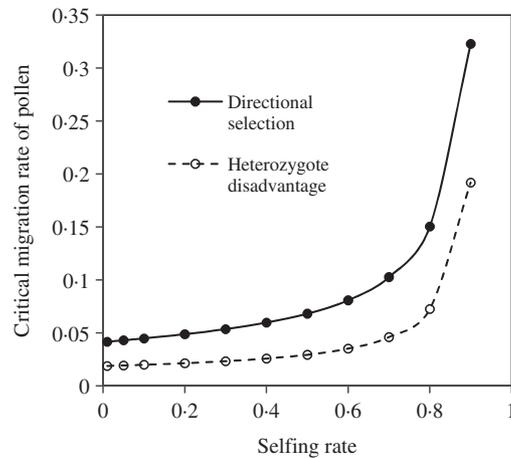


Fig. 5. Critical migration rate of pollen as a function of the selfing rate under the coincidence between the nuclear and maternal organelle allele frequencies. The curve with opened circles is the result estimated from eqn (16) in the scheme of heterozygote disadvantage, with the selection coefficients  $s_2 = 0.04$ ,  $s_C = 0.01$ , and the migration rate  $m_S = 0.001$ . The curve with closed circles is the result estimated from eqn (21) with directional selection, with the selection coefficients  $s_1 = s_2 = 0.01$ ,  $s_C = 0.01$ , and the migration rate  $m_S = 0.01$ .

and  $p_C = 1$ , i.e.

$$\left( 1 - \frac{2(1-\alpha)(1+\rho)+4\alpha}{1-\alpha+(1+\alpha)m_S^*} s_2 \right) m_S^* = (2-\alpha) s_C.$$

It can be shown that  $m_S^*$  is greater with directional selection than in the scheme of heterozygote disadvantage, eqn (14). Exact simulation results confirm that the analytical approximations from eqn (19) perform very well (data not shown here). The critical migration rate of seeds generally reduces with the selfing rate.

Similarly, the critical migration rate for the nuclear gene can be calculated from eqn (A1) in the Appendix,

$$\left( \left( 1 - \frac{2s_C}{1-\alpha+(1+\alpha)m_S} \right) m_S + \frac{1-\alpha}{2} m_P \right)^* = \frac{2(1-\alpha)}{2-\alpha} s_1 + \left( 1 + \alpha - \frac{\alpha^2(1-\alpha)}{2-\alpha} \right) s_2. \tag{20}$$

A negative but not complementary relation exists between seed and pollen flow in contributing to the compounded critical migration rate, given a mixed mating system and selection coefficients. Figure 6 demonstrates that good analytical approximations from eqn (20) can be obtained in comparison with the exact simulation results under various selfing rates. The critical migration rate of pollen increases with the selfing rate, given a fixed migration rate of seeds.

When a coincidence between nuclear and maternal organelle allele frequencies occurs, the critical

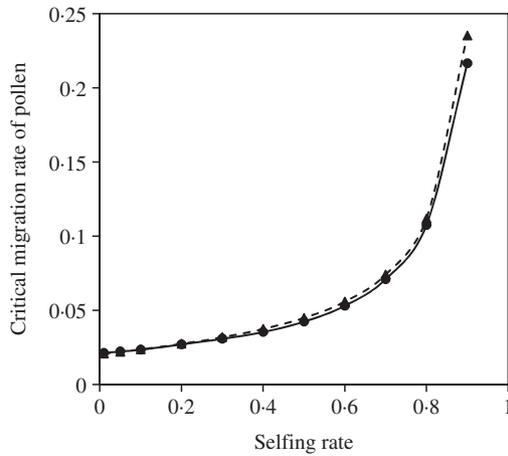


Fig. 6. Critical migration rate of pollen as a function of selfing rate for the nuclear allele. The dashed curve with closed triangles gives the analytical approximation from eqn (20). The curve with closed circles is the exact simulation result. In each case, parameter settings are the migration rate of seeds  $m_S=0.01$ , the selection coefficient  $s_C=0.01$  and  $s_1=s_2=0.01$ .

migration rate for the compounded seed and pollen flow can be derived by combining eqns (A1) and (A4) in the Appendix,

$$\left( (1-f_2s_2)m_S + \frac{1-\alpha}{4}m_P \right)^* = \frac{1-\alpha}{2-\alpha}s_1 + \frac{1}{2} \left( 1+\alpha - \frac{\alpha^2(1-\alpha)}{2-\alpha} \right) s_1 + \frac{1}{2}(2-\alpha)s_C. \tag{21}$$

where

$$f_2 = \frac{1+\alpha+\gamma+(1-\alpha)\rho}{1-\alpha+(1+\alpha)m_S}.$$

Similarly, given the migration rate of pollen (or seeds), critical migration rate of seeds (or pollen) can be estimated from the above equation. The critical migration rate of pollen may be greater with directional selection than that in the scheme of heterozygote disadvantage (e.g. Fig. 5). This critical migration rate of pollen (or seeds) is also unequal to that under the discordance between cytonuclear allele frequencies.

(ii) Nuclear and paternally inherited organelle genome system

The theory in this subsection is applicable to the nuclear and chloroplast genome system in conifer tree species. The difference from the preceding subsection is that both seed and pollen flow may contribute to swamping selection for the paternal organelle genes.

(a) Random mating

From eqn (A6) in the Appendix, the gametic cytonuclear LD between nuclear and paternal organelle

alleles can be approximated by  $D_{AB} \approx 2(m_S + m_P/2)p_a p_b$ , reflecting the balance between the effect of recombination (1/2) that reduces cytonuclear LD and the effect of migration that generates cytonuclear LD. In the scheme of heterozygote disadvantage, the equilibrium for the change of paternal allele frequency can be simplified from eqn (A3) in the Appendix. The effects of cytonuclear LD are negligible in altering allele frequency. The critical migration rate for the compounded seed and pollen flow is given by

$$(m_S + m_P)^* = 2s_B. \tag{22}$$

A complementary relationship exists between seed and pollen flow in contributing to the critical migration rate. Given the setting of seed flow (or pollen flow), the critical migration rate of pollen (or seeds) for the paternal allele can be estimated. Exact simulations confirm that the above analytical approximation performs well in predicting critical migration rate of seeds (or pollen), given a fixed migration rate of pollen (or seeds).

Equilibrium for the change of nuclear allele frequency can be simplified from eqn (A1) in the Appendix. The critical migration rate for the compounded seed and pollen flow is the same as that derived from eqn (6), i.e.  $(m_S + m_P/2)^* = s_2/8$ . The required migration rates of seeds and pollen may be unequal for the nuclear and paternal organelle alleles. Cytonuclear gene frequencies are basically independent due to the negligible effects of cytonuclear LD in random mating, and their alleles can be fixed at different speeds. In the condition of  $m_S = s_2/4 - 2s_B$  and  $m_P = 4s_B - s_2/4$ , which requires  $s_2/16 < s_B < s_2/8$ , both the nuclear and paternal alleles can be fixed even at different speeds.

When a coincidence between the nuclear and paternal organelle allele frequencies ( $p_A = p_B = p$ ) occurs, the critical migration rate for the compounded seed and pollen flow can be derived by combining eqns (A1) and (A3) in the Appendix:  $(m_S + 3m_P/4)^* = (1/2 + \beta)^2 s_2/4$ , where  $\beta = s_B/s_2$ , the ratio of selection coefficients between paternal organelle and nuclear alleles.

Similarly, with directional selection at the nuclear site, it can be shown that the critical migration rate for the paternal organelle allele remains the same as eqn (22). The critical migration rate for the nuclear gene is derived as  $(m_S + m_P/2)^* = (1 + \rho)s_2$ , the same as that derived from eqn (8). Cytonuclear genes are essentially independent but both can be fixed at unequal speeds when  $m_P = 2(2\beta - 1 - \rho)s_2$  and  $m_S = 2(1 + \rho - \beta)s_2$ . When a coincidence between cytonuclear allele frequencies ( $p_A = p_B = p$ ) occurs, the critical migration rate for compounded seed and pollen flow becomes  $(m_S + 3m_P/4)^* = (1 + \rho + 2\beta)s_2/2$ ,

obtained by combining eqns (A1) and (A3) in the Appendix.

(b) Complete selfing

Pollen flow does not contribute to swamping selection. At the steady state, cytonuclear LD can be approximated by  $D_{AB} \approx p_a p_b$  from eqn (A6) in the Appendix, which is much greater than that in random mating. From eqn (A10) in the Appendix, the genotypic cytonuclear LD can be approximated by  $D_{AaB} = 2(1-\alpha)D_{AB}/(2-\alpha) + \Lambda_2(s_\bullet, m_\bullet)$  where  $\Lambda_2(s_\bullet, m_\bullet)$  is the function of the first-order migration rate and selection coefficients. In the scheme of heterozygote disadvantage, it can be shown from eqn (A3) in the Appendix that the critical migration rate of seeds for the paternal organelle allele is  $m_S^* = \beta s_2$ .

From eqn (A1) in the Appendix and the approximation of cytonuclear LD, the steady-state equation for the nuclear allele frequency is  $0 = m_S p_a - s_B p_a p_C$ , which yields the critical migration rate for seeds,  $m_S^* = \beta s_2$ , irrespective of selection pressure at the nuclear site. Exact simulation confirms that this condition also leads to a coincidence between cytonuclear allele frequencies, i.e.  $p_A = p_C = p$ , with the same speed towards fixation (data not presented here).

With directional selection, the steady-state equation for the paternal organelle allele frequency can be readily obtained by replacing subscript C with B in eqn (11), yielding the critical seed migration rate  $m_S^* = (2 + \beta)s_2$ . This is greater than that in the scheme of heterozygote disadvantage. It can also be shown from eqn (A1) in the Appendix that  $m_S^* = (2 + \beta)s_2$  is the critical seed migration rate for the nuclear allele. Exact simulation results confirm that the analytical approximation can predict accurate results. A coincidence between paternal and nuclear allele frequencies occurs with this critical migration rate.

(c) Mixed mating system

From eqn (A6) in the Appendix, the steady-state cytonuclear LD can be approximated by

$$D_{AB} = \frac{2m_S + (1-\alpha)m_P}{1-\alpha + (1+\alpha)m_S + (1-\alpha)m_P} p_a p_b. \tag{23}$$

The above equation reflects the balance among the effects of migration, selfing and cytonuclear recombination. Different from eqn (12), both seed and pollen flow contribute to generating cytonuclear LD. In the random mating case ( $\alpha = 0$ ), eqn (23) can be approximated by  $D_{AB} \approx (2m_S + m_P)p_a p_b$ . In the complete

selfing case ( $\alpha = 1$ ), eqn (23) can be approximated by  $D_{AB} = p_a p_b$ .

In the scheme of heterozygote disadvantage at the nuclear site, the steady-state equation for the paternal organelle allele frequency can be derived from eqn (A3) in the Appendix:

$$(1 + 4f_3 s_2)m_S + (1 + 2f_3 s_2)(1 - \alpha)m_P = (2 - \alpha)s_B p_B, \tag{24}$$

where

$$f_3 = - \frac{(1 - \alpha)}{(2 - \alpha)(1 - \alpha + (1 + \alpha)m_S + (1 - \alpha)m_P)} p_a (p_a - p_A),$$

which is a function of the nuclear allele frequency. The maximum value of the right-hand side of eqn (24) is  $(2 - \alpha)s_B$ . When the left-hand side of eqn (24) is greater than  $(2 - \alpha)s_B$ , the paternal organelle allele frequency  $p_B$  increase towards the fixation. To minimize the migration rate that satisfies the above condition, the maximum  $f_1$  is chosen at  $p_a = 1/4$ . The critical migration rate for the compounded seed and pollen can be derived:

$$((1 + 4f_{3 \max} s_2)m_S + (1 + 2f_{3 \max} s_2)(1 - \alpha)m_P)^* = (2 - \alpha)s_B, \tag{25}$$

where

$$f_{3 \max} = - \frac{(1 - \alpha)}{8(2 - \alpha)(1 - \alpha + (1 + \alpha)m_S + (1 - \alpha)m_P)}.$$

A negative but not complementary relationship exists between seed and pollen flow in contributing to the critical migration rate of seed and pollen. Given the migration rate of seeds (or pollen), the critical migration rate of pollen (or seeds) can be numerically calculated through the iterative method. Figure 7 shows that the exact simulation results for the critical migration rate of pollen are in good agreement with the predictions from eqn (25) under various selfing rates, given a migration rate of seeds. The critical migration rate of pollen nonlinearly increases with the selfing rate when the migration rate of seeds is fixed.

Similarly, the steady-state equation for the nuclear allele frequency can be derived from eqn (A1) in the Appendix

$$\left(1 - \frac{2s_B}{1 - \alpha + (1 + \alpha)m_S + (1 - \alpha)m_P} p_b\right) \left(m_S + \frac{1 - \alpha}{2} m_P\right) = (1 - \alpha) \left(1 + \frac{\alpha^2}{2 - \alpha}\right) s_2 (1 - 2p_A) p_A, \tag{26}$$

which is analogous to eqn (15) except that pollen flow is involved in changing cytonuclear LD. Following the same consideration as in deriving eqn (16), the critical migration rate for the compounded

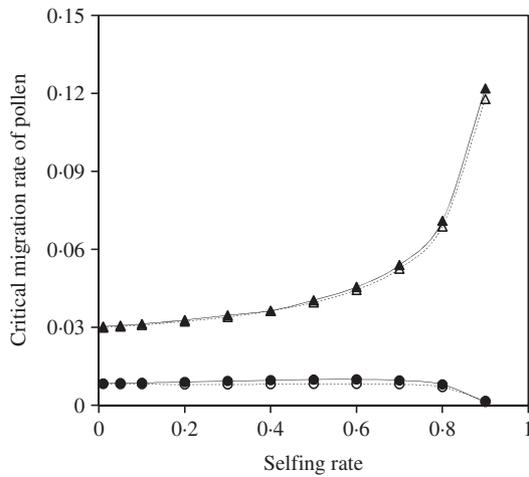


Fig. 7. Critical migration rate of pollen as a function of selfing rate in the scheme of heterozygote disadvantage. The curves with closed and opened triangles give the exact simulation results and the analytical approximation from eqn (25) for the paternal organelle allele, respectively. Parameter settings are the migration rate of seeds  $m_S=0.01$ , the selection coefficient  $s_B=0.02$  and  $s_2=0.04$ . The curves with closed and opened circles give the exact simulation results and the analytical approximation from eqn (27) for the nuclear allele, respectively. Parameter settings are the migration rate of seeds  $m_S=0.001$ , the selection coefficient  $s_B=0.01$  and  $s_2=0.04$ .

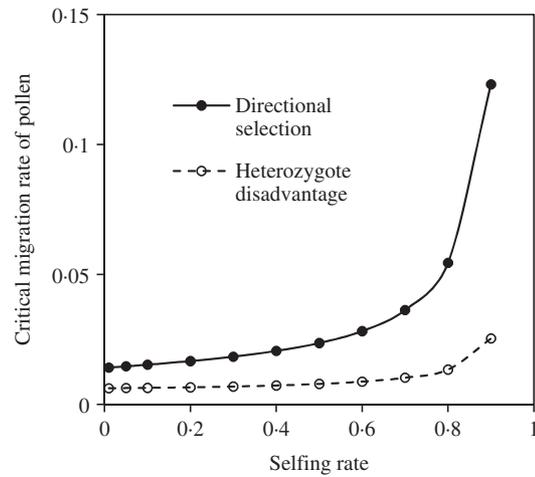


Fig. 8. Critical migration rate of pollen as a function of the selfing rate under the coincidence between the nuclear and paternal organelle allele frequencies. The curve with opened circles is the result estimated from eqn (29) in the scheme of heterozygote disadvantage, with the selection coefficients  $s_2=0.04$ ,  $s_B=0.01$ , and the migration rate  $m_S=0.001$ . The curve with closed circles is the result estimated from eqn (32) with directional selection, with the selection coefficients  $s_1=s_2=0.01$ ,  $s_B=0.01$  and the migration rate  $m_S=0.01$ .

seed and pollen flow is expressed as

$$\left( \left( 1 - \frac{2s_B}{1-\alpha+(1+\alpha)m_S+(1-\alpha)m_P} \right) \times \left( m_S + \frac{1-\alpha}{2} m_P \right) \right)^* = \frac{1}{8}(1-\alpha) \left( 1 + \frac{\alpha^2}{2-\alpha} \right) s_2. \tag{27}$$

Figure 7 shows that the exact simulation results are slightly higher than those predicted from eqn (27) although the similar changing patterns with the selfing rate exist between them.

Equations (25) and (27) are separately applied to the paternal and nuclear alleles. Nuclear and paternal organelle genes can be fixed at different speeds due to the needs of unequal critical migration rates. When a coincidence between cytonuclear allele frequencies ( $p_A=p_B=p$ ) occurs, the steady-state equation for  $p$  can be obtained by combining eqns (A1) and (A3) in the Appendix:

$$(1+f_4s_2)m_S+(3/4+f_4s_2/2)(1-\alpha)m_P = \frac{1}{2}p \left( (2-\alpha)s_B - (1-\alpha) \left( 1 + \frac{\alpha^2}{2-\alpha} \right) (2p-1)s_2 \right), \tag{28}$$

where

$$f_4 = -q \left( \frac{2(1-\alpha)}{2-\alpha} (q-p) + \beta \right) / (1-\alpha+(1+\alpha)m_S+(1-\alpha)m_P).$$

The function  $f_4$  has a maximum,  $f_{4\max} = \frac{1-\alpha}{4(2-\alpha)} \left( 1 - \frac{2-\alpha}{2(1-\alpha)} \beta \right)^2 \frac{1}{1-\alpha+(1+\alpha)m_S+(1-\alpha)m_P}$  at  $q = \frac{1}{4} \left( 1 - \frac{2-\alpha}{2(1-\alpha)} \beta \right)$ . Let  $\lambda_2 = (2-\alpha)^2\beta/(1-\alpha)(2-\alpha+\alpha^2)$ , the critical migration rate for the compounded seeds and pollen can be numerically calculated from the following equation:

$$\left( (1+f_{4\max}s_2)m_S + (3/2+f_{4\max}s_2)(1-\alpha)m_P/2 \right)^* = \frac{1}{16}(1-\alpha) \left( 1 + \frac{\alpha^2}{2-\alpha} \right) (1+\lambda_2)^2 s_2. \tag{29}$$

Figure 8 shows the change for the critical migration rate of pollen with the selfing rate, predicted by eqn (29).

With directional selection, the steady-state equation for the change of paternal organelle allele frequency is

$$(1-2f_5s_2p_a)m_S+(1-f_5s_2)(1-\alpha)m_P=(2-\alpha)s_Bp_B, \tag{30}$$

where

$$f_5 = \frac{(1-\alpha)(1+\rho)+2\alpha}{1-\alpha+(1+\alpha)m_S+(1-\alpha)m_P}.$$

The above equation is analogous to eqn (19) except for the inclusion of pollen flow. The critical migration rate for the compounded seed and pollen flow can be numerically calculated from eqn (30) by letting  $p_a=1$  and  $p_B=1$ . This analytical approximation performs

very well in comparison with the exact simulation results (e.g. Fig. 9).

Similarly, the critical migration rate for the nuclear gene can be calculated from eqn (A1) in the Appendix,

$$\left( \left( 1 - \frac{2s_B}{1-\alpha + (1+\alpha)m_S + (1-\alpha)m_P} \right) \times \left( m_S + \frac{1-\alpha}{2} m_P \right) \right)^* \quad (31)$$

$$= \frac{2(1-\alpha)}{2-\alpha} s_1 + \left( 1 + \alpha - \frac{\alpha^2(1-\alpha)}{2-\alpha} \right) s_2.$$

Given a migration rate of seeds (or pollen), the critical migration rate of pollen (or seeds) can be numerically calculated. Figure 9 demonstrates that estimates from eqn (31) are generally in good agreement with the exact simulation results.

Equations (30) and (31) separately address the critical migration rate for nuclear and paternal organelle alleles in the presence of cytonuclear LD. When a coincidence between nuclear and paternal organelle allele frequencies occurs, the critical migration rate for the compounded seed and pollen flow can be derived by combining eqns (A1) and (A3) in the Appendix:

$$\left( (1-f_5s_2)m_S + \left( \frac{3}{2} - f_5s_2 \right) \frac{1-\alpha}{2} m_P \right)^* = \frac{1-\alpha}{2-\alpha} s_1 + \frac{1}{2} \left( 1 + \alpha - \frac{\alpha^2(1-\alpha)}{2-\alpha} \right) s_2 + \frac{1}{2} (2-\alpha) s_B. \quad (32)$$

The above equation is analogous to eqn (21) but has a different  $f$  function. Figure 8 shows the pattern for the change of the critical migration rate of pollen, predicted from eqn (32), with the selfing rate. The required migration rate of pollen, given a fixed rate of seed flow, is higher with directional selection than in the scheme of heterozygote disadvantage.

(iii) Nuclear and maternally and paternally inherited organelle genome system

The three-genome system, each genome with different modes of inheritance, is applicable to most conifer tree species. Theoretical deductions indicate that under weak selection, LD between paternal and maternal organelle genomes essentially does not change their allele frequencies although each of them separately interacts with the nuclear genomes (Appendix). This simplifies the analysis of the three-genome system.

(a) Random mating

In random mating, three cytonuclear genomes are essentially independent since the effects of cytonuclear

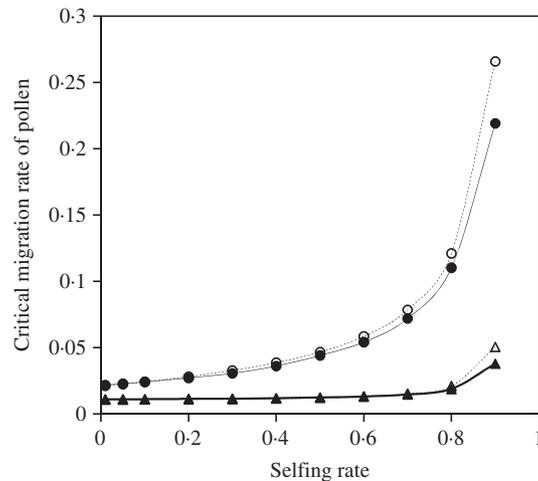


Fig. 9. Critical migration rate of pollen as a function of selfing rate with directional selection. The curves with closed and opened triangles give the exact simulation results and the analytical approximation from eqn (30) for the paternal organelle allele, respectively. Parameter settings are the migration rate of seeds  $m_S=0.01$ , the selection coefficient  $s_B=0.01$  and  $s_1=s_2=0.01$ . The curves with closed and opened circles give the exact simulation results and the analytical approximation from eqn (31) for the nuclear allele, respectively. Parameter settings are the migration rate of seeds  $m_S=0.01$ , the selection coefficient  $s_B=0.01$  and  $s_1=s_2=0.01$ .

LD, mainly maintained by migration and recombination, are negligible in changing cytonuclear allele frequencies. The swamping effects of migration for individual genomes have been addressed in the preceding sections. Here, only the coincidence among three genomes ( $p_A=p_B=p_C=p$ ) is presented.

In the scheme of heterozygote disadvantage, the critical migration rate for the compounded seed and pollen flow can be derived by combining eqns (A1), (A2) and (A3) in the Appendix:

$$(m_S + m_P/2)^* = (1 + 2\beta + 2\gamma)^2 s_2 / 24. \quad (33)$$

With directional selection, the critical migration rate for compounded seed and pollen flow can be derived by combining eqns (A1)–(A3) in the Appendix:

$$(m_S + m_P/2)^* = (1 + \rho + 2\beta + 2\gamma) s_2 / 3. \quad (34)$$

In each case, the critical migration rate of seeds (or pollen) is unequal to that in the two-genome case, given the same migration rate of pollen (or seeds).

(b) Complete selfing

In the complete selfing, the critical migration rates for paternal and maternal organelle alleles remain the same as in the preceding sections. The difference from the two-genome cytonuclear system is that both paternal and maternal organelle genomes contribute

to the swamping effects of migration for the nuclear allele.

In the scheme of heterozygote disadvantage, the critical migration rate of seeds for the nuclear allele is  $m_S^* = (\beta + \gamma)s_2$ , greater than the result at the two-genome system. When a coincidence among the three genomes occurs, the critical migration rate of seeds can be obtained by combining eqns (A1), (A3) and (A4) in the Appendix:  $m_S^* = (\beta + \gamma)s_2/3$ .

With directional selection, the critical migration rate of seeds for the nuclear allele is given by  $m_S^* = (2 + \beta + \gamma)s_2$ . When a coincidence among the three genomes occurs, the critical migration rate of seeds can be obtained by combining eqns (A1), (A3) and (A4) in the Appendix:  $m_S^* = (\beta + \gamma + 4)s_2/3$ . This critical value is greater with directional selection than in the scheme of heterozygote disadvantage.

(c) *Mixed mating system*

The approximations for the steady-state cytonuclear LD between the nuclear and maternal organelle alleles or between the nuclear and paternal organelle alleles remain the same as eqn (12) or (23), respectively. Expressions for the critical migration rate remain the same as eqn (13) or (19) for the seed flow for the maternal organelle gene, or the same as eqn (25) or (30) for the compounded seed and pollen flow for the paternal organelle gene.

In the scheme of heterozygote disadvantage, the critical migration rate for compounded seed and pollen flow for the nuclear gene can be derived from eqn (A1) in the Appendix:

$$\left( \left( 1 - \frac{2s_B}{1 - \alpha + (1 + \alpha)m_S + (1 - \alpha)m_P} - \frac{2s_C}{1 - \alpha + (1 + \alpha)m_S} \right) m_S + \left( 1 - \frac{2s_B}{1 - \alpha + (1 + \alpha)m_S + (1 - \alpha)m_P} \right) \frac{1 - \alpha}{2} m_P \right)^* = \frac{1}{8} (1 - \alpha) \left( 1 + \frac{\alpha^2}{2 - \alpha} \right) s_2. \tag{35}$$

Equation (35) is analogous to eqn (16) or (27) in the two-genome system except that effects of cytonuclear LD are linearly added to its left-hand side. Given the migration rate of seeds (or pollen), the critical migration rate of pollen (or seeds) can be numerically calculated using the iterative method. Figure 10 shows that the critical migration rate of pollen from the analytical approximation is slightly biased from the exact simulation results although a similar pattern as the function of selfing rate exists between them, similar to the results in the two-genome cases.

When a coincidence among the three-genome allele frequencies ( $p_A = p_B = p_C = p$ ) occurs, the steady-state

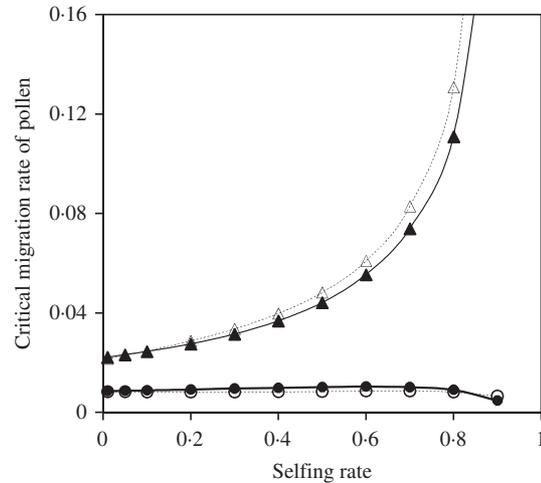


Fig. 10. Critical migration rate of pollen as a function of selfing rate for the nuclear allele. The curves with closed and opened circles give the exact simulation results and the analytical approximation from eqn (35) in the scheme of heterozygote disadvantage, respectively. Parameter settings are the migration rate of seeds  $m_S = 0.001$ , the selection coefficient  $s_B = s_C = 0.01$ ,  $s_1 = 0.0$  and  $s_2 = 0.04$ . The curves with closed and opened triangles give the exact simulation results and the analytical approximation from eqn (37) with directional selection, respectively. Parameter settings are the migration rate of seeds  $m_S = 0.01$ , the selection coefficient  $s_B = s_C = 0.01$  and  $s_1 = s_2 = 0.01$ .

equation for  $p$  can be calculated by simply combining eqns (A1), (A3) and (A4) in the Appendix. The critical migration rate for the compounded seed and pollen flow can be derived:

$$\left( \left( \frac{3}{2} + f_{4 \max} s_2 + f_{1 \max} s_2 \right) m_S + \left( \frac{3}{2} + f_{4 \max} s_2 \right) \frac{1 - \alpha}{2} m_P \right)^* = \frac{1}{16} (1 - \alpha) \left( 1 + \frac{\alpha^2}{2 - \alpha} \right) (1 + \lambda_1 + \lambda_2)^2 s_2. \tag{36}$$

Compared with eqns (18) and (29) in the two-genome systems, there is not a simple relationship among them. Figure 11 shows the pattern for the change of critical migration rate of pollen with the selfing rate, given a fixed migration rate of seeds. The results similar to Figs 5 and 8 can be observed.

With directional selection, the critical migration rate for compounded seed and pollen flow for the nuclear allele can be derived in the same method:

$$\left( \left( 1 - \frac{2s_B}{1 - \alpha + (1 + \alpha)m_S + (1 - \alpha)m_P} - \frac{2s_C}{1 - \alpha + (1 + \alpha)m_S} \right) m_S + \left( 1 - \frac{2s_B}{1 - \alpha + (1 + \alpha)m_S + (1 - \alpha)m_P} \right) \frac{1 - \alpha}{2} m_P \right)^* = \frac{2(1 - \alpha)}{2 - \alpha} s_1 + \left( 1 + \alpha - \frac{\alpha^2(1 - \alpha)}{2 - \alpha} \right) s_2. \tag{37}$$

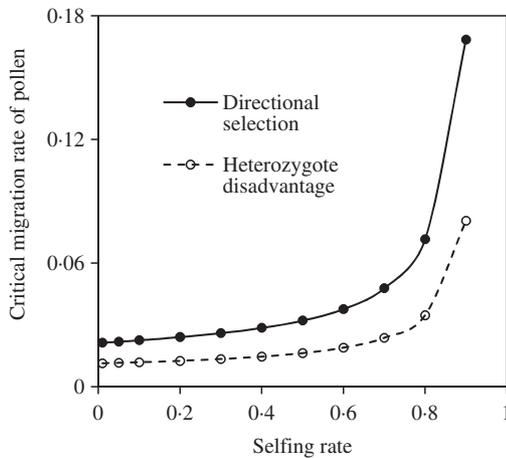


Fig. 11. Critical migration rate of pollen as a function of the selfing rate under the coincidence among nuclear, maternal and paternal organelle allele frequencies. The curve with opened circles is the result estimated from eqn (36) in the scheme of heterozygote disadvantage, with the selection coefficients  $s_2=0.04$ ,  $s_B=s_C=0.01$ , and the migration rate  $m_S=0.001$ . The curve with closed circles is the result estimated from eqn (38) with directional selection, with the selection coefficients  $s_1=s_2=0.01$ ,  $s_B=s_C=0.01$ , and the migration rate  $m_S=0.01$ .

Equation (37) is analogous to eqns (20) and (31) in the two-genome system except that effects of cytonuclear LD are linearly added. Exact simulation results indicate that the analytical approximation by eqn (37) performs very well (Fig. 10). The critical migration rate of pollen increases with the selfing rate under a fixed rate of seed flow.

Similarly, when a coincidence among the three-genome allele frequencies occurs, the critical migration rate for compounded seed and pollen flow is given by

$$\begin{aligned} & \left( \left( \frac{3}{2} - f_5 s_2 - f_2 s_2 \right) m_S + \left( \frac{3}{2} - f_5 s_2 \right) \frac{1-\alpha}{2} m_P \right)^* \\ &= \frac{(1-\alpha)}{2-\alpha} s_1 + \frac{1}{2} \left( 1 + \alpha - \frac{\alpha^2(1-\alpha)}{2-\alpha} \right) s_2 \\ &+ \frac{1}{2} (2-\alpha)(s_B + s_C). \end{aligned} \tag{38}$$

Compared with the two-genome system (eqns 21 and 32), effects of cytonuclear LD are linearly added into eqn (38). Figure 11 shows the effects of the mating system on the critical migration rate of pollen, given a fixed rate of seed flow. The critical migration rate of pollen may be much greater with directional selection than in the scheme of heterozygote disadvantage.

### 5. Discussion

This study demonstrates how the mating system changes the critical migration rate for swamping selection for plant genomes with contrasting modes of

inheritance (maternal, paternal and bi-parental inheritance), necessarily extending the previous studies to include the effects of mating systems and the joint cytonuclear system (Haldane, 1930; Wright, 1969; Crow *et al.*, 1990; Barton, 1992). The analytical and simulation results demonstrate that the mating system can alter the critical migration rates of seed and pollen through changing effective pollen flow, cytonuclear LD and gametic selection for both nuclear and paternal organelle genes. Partial selfing changes the critical migration rate of seeds for the maternal organelle genes through changing gametic selection and cytonuclear LD. Both concordance and discordance between cytonuclear genes can occur during the gene swamping process due to the needs of unequal critical migration rates of seeds and pollen. These theoretical results expand our understanding of the joint effects of migration and selection in evolving natural hermaphrodite plant populations.

It is important to understand the biological significance of incorporating the mating system into the joint effects of migration and selection, the very important type of interaction that is considered as one of mechanisms for gene spread, species' range expansion and incipient speciation (Wright, 1931, 1977; Fisher, 1937; Haldane, 1956; Barton & Charlesworth, 1984; Kirkpatrick & Barton, 1997). Complete random mating and selfing represent two extreme breeding systems, while the mixed mating system is considered as a transitional pattern (Goodwillie *et al.*, 2005; Charlesworth, 2006). Complete random mating allows a maximum probability of immigrating alien pollen and enhances differential fixation rates between cytonuclear genes. A low migration rate can swamp weak selection for the nuclear allele, but a high migration rate may be required for the maternal organelle gene. The effect of random mating has been addressed for the nuclear system (Crow *et al.*, 1990; Barton, 1992). Complete selfing does not allow immigration of alien pollen but enhances the concordant swamping process between cytonuclear genes. A typical flowering plant is hermaphroditic and self-compatible (Heilbut, 2000), with co-sexual functions for an individual. The evolutionary stability of this reproductive system remains controversial due to the use of two contrasting strategies with different proportions of gene transmission to progeny (Maynard Smith, 1982; Goodwillie *et al.*, 2005). The proportion of each strategy (selfing versus outcrossing) alters the relative pollen and seed flow and hence changes the swamping rate. The present theory confirms that the mixed mating system directly adjusts the contributions of alien pollen to swamping selection in a complicated non-linear way for the nuclear and paternal organelle genes.

In comparison with the nuclear system for animal species (Barton, 1992), the expression for critical

migration rate for a single nuclear gene is the same in the cytonuclear system under random mating with weak selection. The only step is to replace the critical migration rate for animal species with  $m_S + m_P/2$  for plant species. However, the inclusion of cytonuclear LD, attained by selfing and migration in the mixed mating system, falsifies such simple transformation. One more likely system is the involvement of multiple nuclear sites because nuclear genomes are much larger in size and have a more number of genes than organelle genomes (Barton, 1992; Birky, 1995). The effects of selection on multiple nuclear sites may substantially reduce effective gene flow, which makes swamping less likely. Also, the transformation from the nuclear to the cytonuclear system becomes complicated in the mixed mating system because additional LD between nuclear sites can alter the process of nuclear gene swamping. A more sophisticated model is needed to assess the critical migration rate and the relative contributions of cytonuclear LD versus LD between nuclear sites.

Similarly, in comparison with the sole organelle genome system (not separately addressed here), the critical migration rate for maternal or paternal organelle gene remains the same as in the cytonuclear system in random mating with weak selection. The inclusion of cytonuclear LD in the mixed mating system can modify this swamping process in a non-linear way. It is speculated that such a modification can be intensified when multiple nuclear sites are jointly considered, as implied from the studies on gene spreading in structured populations (Hu, 2008, 2010).

Biologically, both concordant and discordant processes of gene swamping between cytonuclear genes have significantly evolutionary meanings although the proportion of nuclear or organelle genes undergoing these distinct processes is unknown. The population fitness in the sporophyte stage here is determined by cytonuclear genes rather than by the individual nuclear or organelle genes. The same swamping rate, such as in the complete selfing case, implies that the whole three-genome system approaches maximum co-evolution. The population fitness can be shifted when different population fitness exists between the source and recipient populations, the phase III of Wright's SBT (Wright, 1977). The unequal swamping rates result in a different structure for cytonuclear genes, and leads to a lag in population adaptation between genomes.

This study assumes that genetic drift effects are negligible compared with the effects of seed and pollen flow and selection. This is plausible for a large population where a deterministic approach is approximated. For a small population, genetic drift could also cause population fitness shifting through the joint effects of drift and selection (Wright, 1977). An important extension in a future study is to include

genetic drift effects to derive the average critical migration rate of seed and pollen in the mixed mating system. Some insights into the drift and migration effects can be obtained from extensive explorations on Wright's SBT for both polygene quantitative traits and nuclear genes under random mating with classic island model (Barton & Rouhani, 1993). It is likely that the average critical migration rate of seed and pollen can be further modified by genetic drift effects.

The present study only addresses the case of one-way gene flow. The general exact model can be applied to the case of two-way gene flow as well. This can be done simply by setting the migrant cytonuclear genotypes not as constants and swapping the frequency notation  $p$  and  $Q$  to get the dynamics for the other population. The results in the Appendix need not be changed and are applicable to each population. What needs changing is the analytical deduction for the critical migration rate of seed and pollen, which becomes more complicated in algebra. Early studies under random mating for nuclear genes indicate that the critical migration rates for the forward and reverse directions may be unequal when the two population fitness is different (Crow *et al.*, 1990; Barton, 1992). It is speculated that similar results might occur, given the constant mixed mating system and soft selection schemes, which remains to be explored.

The analytical results only address two selection schemes for the nuclear genes and one selection scheme for organelle genes. The migrating alleles are assumed to be maladaptive to the recipient population and cause migration loads. Such migration load is the function of the mating system (Hu & Li, 2003). Only when the migration rate exceeds a threshold can the selection effects be swamped. However, the critical migration rate also exists when migrating genes are more adaptive in the scheme of heterozygote disadvantage, as demonstrated in previous studies (Crow *et al.*, 1990; Barton, 1992). This is because an equilibrium point between 0 and 1 for the nuclear gene is unstable. Under this situation, the application of the present theory simply needs to change selection coefficient settings. When other selection schemes are employed, such as the disruptive selection with asymmetry in fitness of the homozygotes (Barton & Rouhani, 1993), the analytical expression for the critical migration rate of seed and pollen needs modification under the mixed mating system.

Although the purpose of this study is to examine how the mating system changes the critical migration rate of seeds and pollen in the cytonuclear system, several implications can be deduced. One is to predict the potential impacts of the migration of genetically modified populations to natural populations through pollen (e.g. hybridization and introgression) and seed

flow, an important issue recently received considerable attention in conservation biology (Ellstrand *et al.*, 1999; Ellstrand, 2003). Hu *et al.* (2003) demonstrate that genetic variation and population fitness of natural populations can be altered, depending on the type of selection scheme (adaptive/maladaptive migrating genes). This result is derived for the nuclear genes under random mating. The present study furthers such impacts for both nuclear and organelle genes since the genetically modified genes may be located on organelle genomes as well (Verma & Daniell, 2007). The present results highlight the role of the mating system in swamping natural selection. Concerning the impacts through pollen flow, natural populations could be more seriously affected for species with a predominant outcrossing system (e.g. most conifer trees) than for species with the predominant selfing system. A small migration rate could sufficiently bring about gene swamping under weak selection, as implied from the present results. In addition, when the natural population is small, the probability of fixation of maladaptive genes from migrating pollen or seeds can be enhanced (Wright, 1931). The joint effects of genetic drift and migration might speed up or impede gene swamping, and this remains to be explored in a future study.

The second implication is to predict the spatial variation in critical migration rate of seeds and pollen in spreading either adaptive or maladaptive genes. Empirical studies on genetic swamping are recorded, primarily in hybrid zones (Antonovics, 1976; Raymond & Marquine, 1994; Childs *et al.*, 1996; Avise *et al.*, 1997). Ample evidence indicates the presence of population variation in the mating system since many biotic and abiotic factors can influence seed flow and pollination (Mitton, 1992; Goodwillie *et al.*, 2005; Charlesworth, 2006; Eckert *et al.*, 2009). The shift between insect and wind pollination can cause variation in the migration rate of pollen among local populations (Stebbins, 1970). The shift between animal and wind seed dispersal can cause variation in the migration rate of seeds among local populations (Montoya *et al.*, 2008). It is commonly held that heterogeneous selection often exists among natural populations in many plant species. Such heterogeneity may be attributable to the differences in either gametic selection, or zygotic selection, or both, which purges maladaptive or enhances adaptive genes. The present analytical results show that the critical migration rate of seeds and pollen is the function of the mating system and selection coefficients. Accordingly, the critical migration rates required for swamping natural selection are expected to vary in space for cytonuclear genes.

The third implication is to predict how the mating system adjusts the relative contributions of pollen and seed flow to the critical migration rate. Ennos (1994)

has demonstrated the relative rates of pollen to seed flow as a function of the mating system in terms of  $F_{st}$  for nuclear and organelle markers. Pollen flow often dominantly contributes to gene flow for the plant species with predominant outcrossing rate (Ennos, 1994; Ennos *et al.*, 1999; Dick *et al.*, 2008). The present study shows that a simple complementary relationship between pollen and seed flow exists in contributing to the total critical migration rate for nuclear and paternal organelle genes in random mating. However, a non-complementary relationship exists between seed and pollen flow in contributing to the total critical migration rate for a mixed mating system. Combining the present results with those obtained from  $F_{st}$ , it is speculated that pollen flow (or seed flow) might play a dominant role in order to swamp natural selection for the predominant outcrossing (or selfing) species.

Finally, the present theory predicts how the mating system influences gene swamping for the haploid maternal organelle genes. Although there are many empirical studies on mating systems and seed dispersal, the effects of the mating system on the spread of maternal organelle genes are rarely recorded. In a pure neutral process, a mixed mating system does affect the spread of maternal organelle genes, which can be inferred from eqn (A4) in the Appendix by setting selection coefficients to zero. The neutral maternal organelle gene spreads through seed dispersal, given that the gene attains a certain level of frequency. When the maternal organelle gene is selectively not neutral, the mixed mating system influences both gametic and zygotic selection, or the completely selfing influences the gametic selection. Here, the evolutionary significance is that genetic swamping from seed dispersal is enhanced for species with a higher selfing rate, given comparable natural selection strengths. This awaits empirical verification with collections of appropriate data.

#### Appendix: Cytonuclear LDs in a mixed mating system

Consider three diallelic cytonuclear sites, with alleles  $A$  and  $a$  at the nuclear site,  $B$  and  $b$  at the organelle genome site of paternal inheritance, and  $C$  and  $c$  at the organelle genome site of maternal inheritance. The same assumptions as the main text are considered about the model of population structure and the life cycle of hermaphrodite plants. With weak selection, the terms containing the second or higher orders of the selection coefficients are neglected. The terms containing the second or higher orders of the immigration rate ( $m_P^2$ ,  $m_S^2$  or  $m_S m_P$  or higher orders) and the products of the migration rates and selection coefficients ( $s m_P$  or  $s m_S$ ) are neglected as well. A linear-additive-viability model is assumed between cytonuclear genes. Let the

gametic fitness be  $w_{AB}=1+s_A+s_B$  and  $w_{AC}=1+s_A+s_C$ , where  $s_A$ ,  $s_B$  and  $s_C$  are the selection coefficients for alleles  $A$ ,  $B$  and  $C$ , respectively. Let  $w_{AABC}=1+s_{AA}+s_B+s_C$ , where  $s_{AA}$  is the selection coefficient for genotype  $AA$ . Fitness for other gametes and cytonuclear genotypes are set in the similar way.

From the general eqns (2) and (3) in the main text, the recursion equation for the nuclear allele frequency is derived as

$$\begin{aligned}
 p_A^{**} = & \left(1 - m_S - \frac{1 - \alpha}{2} m_P\right) + \left(m_S + \frac{1 - \alpha}{2} m_P\right) \\
 & Q_A + (s_A - s_a) \left( (1 - \alpha) p_A p_a + \frac{1}{4} \alpha p_{Aa} \right) \\
 & + (p_A (s_{AA} - s_{Aa}) + p_a (s_{Aa} - s_{aa})) p_A p_a \\
 & + \alpha (p_A (s_{Aa} - s_{Aa}) + p_a (s_{AA} - s_{Aa})) \left( p_A p_a - \frac{1}{4} p_{Aa} \right) \\
 & + (s_B - s_b) D_{AB} + (s_C - s_c) D_{AC}. \tag{A1}
 \end{aligned}$$

Cytonuclear LD ( $D_{AB}$  and  $D_{AC}$ ) gives additional sources to change the nuclear allele frequency in addition to migration and selection. The expression for the change of nuclear heterozygosity is given by

$$\begin{aligned}
 p_{Aa}^{**} = & (1 - m_S) \left( 2(1 - \alpha) p_A p_a + \frac{1}{2} \alpha p_{Aa} \right) \\
 & + (1 - \alpha) m_P (p_a - p_A) (Q_A - p_A) + m_S Q_{Aa} \\
 & + 2(s_A - s_a) (1 - \alpha) p_A p_a (p_a - p_A) \\
 & - \left( (s_{AA} - s_{Aa}) \left( (1 - \alpha) p_A^2 + \alpha \left( p_A - \frac{1}{4} p_{Aa} \right) \right) \right. \\
 & \left. + (s_{aa} - s_{Aa}) \left( (1 - \alpha) p_a^2 + \alpha \left( p_a - \frac{1}{4} p_{Aa} \right) \right) \right) \\
 & \times \left( 2(1 - \alpha) p_A p_a + \frac{1}{2} \alpha p_{Aa} \right) \\
 & + (s_B - s_b) \left( 2(1 - \alpha) (p_a - p_A) D_{AB} + \frac{1}{2} \alpha D_{AaB} \right) \\
 & + (s_C - s_c) \left( 2(1 - \alpha) (p_a - p_A) D_{AC} + \frac{1}{2} \alpha D_{AaC} \right). \tag{A2}
 \end{aligned}$$

In the absence of cytonuclear LD, the above two expressions reduce to the results of Hu & Li (2003; eqns 5 and 6) by resetting selection coefficients.

The change for the frequency of the organelle gene with paternal inheritance is derived as

$$\begin{aligned}
 p_B^{**} = & p_B + (m_S + (1 - \alpha) m_P) (Q_B - p_B) \\
 & + (s_B - s_b) (2 - \alpha) p_B p_b + (s_A - s_a) (1 - \alpha) D_{AB} \\
 & - (1 - \alpha) (-p_A (s_{AA} - s_{Aa}) + p_a (s_{aa} - s_{Aa})) D_{AB} \\
 & - \alpha ((s_{aa} - s_{AA}) D_{AB} + \frac{1}{4} (s_{AA} - 2s_{Aa} + s_{aa}) D_{AaB}). \tag{A3}
 \end{aligned}$$

Similarly, the change for the frequency of the organelle gene with maternal inheritance is given by

$$\begin{aligned}
 p_C^{**} = & p_C + m_S (Q_C - p_C) + (s_C - s_c) (2 - \alpha) \\
 & \times p_C p_c + (s_A - s_a) (1 - \alpha) D_{AC} \\
 & - (1 - \alpha) (-p_A (s_{AA} - s_{Aa}) + p_a (s_{aa} - s_{Aa})) D_{AC} \\
 & - \alpha ((s_{aa} - s_{AA}) D_{AC} \\
 & + \frac{1}{4} (s_{AA} - 2s_{Aa} + s_{aa}) D_{AaC}). \tag{A4}
 \end{aligned}$$

A mixed mating system can modify the organelle allele frequency through cytonuclear LD. In the case of complete selfing ( $\alpha=1$ ), the selection at the nuclear site does not change the organelle gene frequency in the gametophyte stage, but does in the sporophyte stage. In the case of complete outcrossing ( $\alpha=0$ ), the genotypic cytonuclear LD does not affect organelle gene frequency, but the gametic cytonuclear LD does.

Let  $\Delta_1 = (1 - \alpha) (p_{AB} p_{aC} + p_{aB} p_{AC}) / 2 + \alpha p_{AaBC} / 4$  and  $\Delta_2 = (1 - \alpha) p_{AB} p_{AC} + \alpha (p_{AABC} + p_{AaBC} / 4)$ . The change for the frequency of three-allele cytonuclear combination is given by

$$\begin{aligned}
 p_{ABC}^{**} = & (1 - \alpha) \left( p_A p_B p_C + \frac{1}{2} p_C D_{AB} + \frac{1}{2} p_B D_{AC} \right) + \alpha p_{ABC} \\
 & + m_P (1 - \alpha) \left( \frac{1}{2} Q_{AB} p_C + \frac{1}{2} Q_{AC} p_B - p_A p_B p_C \right. \\
 & \left. - \frac{1}{2} p_B D_{AC} - \frac{1}{2} p_C D_{AB} \right) \\
 & + m_S \left( Q_{ABC} - \alpha p_{ABC} - (1 - \alpha) \right. \\
 & \left. \times \left( p_A p_B p_C + \frac{1}{2} p_B D_{AC} + \frac{1}{2} p_C D_{AB} \right) \right) \\
 & + (s_A - s_a) \left( (1 - \alpha) \left( p_A p_a p_B p_C + D_{AB} D_{AC} \right. \right. \\
 & \left. \left. + \frac{1}{2} p_B D_{AC} + \frac{1}{2} p_C D_{AB} \right) + \frac{1}{4} \alpha p_{AaCD} \right) \\
 & + (s_B - s_b) p_b \left( (1 - \alpha) (2p_A p_B p_C + p_C D_{AB} + p_B D_{AC}) \right. \\
 & \left. + \alpha p_{ABC} \right) + (s_C - s_c) p_c \left( (1 - \alpha) \right. \\
 & \left. \times (2p_A p_B p_C + p_C D_{AB} + p_B D_{AC}) + \alpha p_{ABC} \right) \\
 & + (s_{AA} - s_{Aa}) \left( \Delta_2 - \left( (1 - \alpha) p_A^2 + \alpha \left( p_A - \frac{1}{4} p_{Aa} \right) \right) \right. \\
 & \left. \times (\Delta_1 + \Delta_2) \right) - (s_{aa} - s_{Aa}) \\
 & \times \left( (1 - \alpha) p_a^2 + \alpha \left( p_a - \frac{1}{4} p_{Aa} \right) \right) (\Delta_1 + \Delta_2). \tag{A5}
 \end{aligned}$$

The above equation is used for calculating three-allele cytonuclear LD as the function of migration, selection and mating system.

The change for the gametic LD between the nuclear and paternal organelle sites is given by

$$\begin{aligned}
 D_{AB}^* = & \frac{1+\alpha}{2} D_{AB} + \frac{1-\alpha}{2} m_P((Q_A - p_A) \\
 & \times (Q_B - p_B) + \bar{D}_{AB} - D_{AB}) \\
 & + m_S \left( (Q_A - p_A)(Q_B - p_B) + \bar{D}_{AB} - \frac{1+\alpha}{2} D_{AB} \right) \\
 & + (s_A - s_a) \left( \frac{1-\alpha}{2} (p_a - p_A) D_{AB} + \frac{1}{4} \alpha D_{AaB} \right) \\
 & + (s_B - s_b)(p_b - p_B) D_{AB} \\
 & + (s_C - s_c) \alpha (p_{ABC} - p_A p_B p_C - p_B D_{AC} - p_C D_{AB}) \\
 & + (s_{AA} - s_{Aa}) \left( p_A \left( p_a - \frac{1}{2} p_A \right) D_{AB} \right. \\
 & \left. + \alpha \left( \left( p_a^2 - \frac{1}{2} \left( p_A - \frac{1}{4} p_{Aa} \right) \right) \right. \right. \\
 & \left. \left. \times D_{AB} - \frac{1}{4} p_a D_{AaB} \right) \right. \\
 & \left. - \alpha^2 \frac{1}{2} \left( p_A p_a - \frac{1}{4} p_{Aa} \right) D_{AB} \right) \\
 & + (s_{aa} - s_{Aa}) \left( p_a \left( p_A - \frac{1}{2} p_A \right) D_{AB} \right. \\
 & \left. + \alpha \left( \left( p_A^2 - \frac{1}{2} \left( p_a - \frac{1}{4} p_{Aa} \right) \right) \right. \right. \\
 & \left. \left. \times D_{AB} - \frac{1}{4} p_A D_{AaB} \right) \right. \\
 & \left. - \alpha^2 \frac{1}{2} \left( p_A p_a - \frac{1}{4} p_{Aa} \right) D_{AB} \right). \tag{A6}
 \end{aligned}$$

In random mating ( $\alpha=0$ ) and the absence of gametic selection, the above equation reduces to the result equivalent to eqn (3) of Hu & Li (2002).

The change for the gametic LD between the nuclear and maternal organelle sites is given by

$$\begin{aligned}
 D_{AC}^* = & \frac{1+\alpha}{2} D_{AC} + m_S \left( (Q_A - p_A)(Q_C - p_C) \right. \\
 & \left. + \bar{D}_{AC} - \frac{1+\alpha}{2} D_{AC} \right) \\
 & + (s_A - s_a) \left( \frac{1-\alpha}{2} (p_a - p_A) D_{AC} + \frac{1}{4} \alpha D_{AaC} \right) \\
 & + (s_C - s_c)(p_c - p_C) D_{AC} \\
 & + (s_B - s_b) \alpha (p_{ABC} - p_A p_B p_C - p_B D_{AC} - p_C D_{AB}) \\
 & + (s_{AA} - s_{Aa}) \left( p_A \left( p_a - \frac{1}{2} p_A \right) D_{AC} \right. \\
 & \left. + \alpha \left( \left( p_a^2 - \frac{1}{2} \left( p_A - \frac{1}{4} p_{Aa} \right) \right) \right. \right. \\
 & \left. \left. \times D_{AC} - \frac{1}{4} p_a D_{AaC} \right) \right)
 \end{aligned}$$

$$\begin{aligned}
 & - \alpha^2 \frac{1}{2} \left( p_A p_a - \frac{1}{4} p_{Aa} \right) D_{AC} \\
 & + (s_{aa} - s_{Aa}) \left( p_a \left( p_A - \frac{1}{2} p_A \right) D_{AC} \right. \\
 & \left. + \alpha \left( \left( p_A^2 - \frac{1}{2} \left( p_a - \frac{1}{4} p_{Aa} \right) \right) \right. \right. \\
 & \left. \left. \times D_{AC} - \frac{1}{4} p_A D_{AaC} \right) \right. \\
 & \left. - \alpha^2 \frac{1}{2} \left( p_A p_a - \frac{1}{4} p_{Aa} \right) D_{AC} \right). \tag{A7}
 \end{aligned}$$

Similarly, in random mating ( $\alpha=0$ ), the above equation reduces to the result equivalent to eqn (12a) of Hu & Li (2002). Pollen flow does not contribute to the change of the gametic LD between the nuclear and maternal organelle sites.

The recursion equation for the change of LD between the paternal and maternal organelle sites is derived as

$$\begin{aligned}
 D_{BC}^{**} = & \alpha D_{BC} + m_S((Q_B - p_B)(Q_C - p_C) + \bar{D}_{BC} - \alpha D_{BC}) \\
 & + (s_B - s_b) \alpha p_b D_{BC} + (s_C - s_c) \alpha p_c D_{BC} \\
 & + (s_{AA} + s_{aa} - 2s_{Aa}) \left( (1-\alpha) D_{AB} D_{AC} + \alpha(1-\alpha) \right. \\
 & \left. \times \left( p_A p_a - \frac{1}{4} p_{Aa} \right) D_{BC} - \frac{1}{4} \alpha D_{AaC} D \right) \\
 & + (s_{AA} - s_{aa}) \alpha D_{ABC}. \tag{A8}
 \end{aligned}$$

In random mating ( $\alpha=0$ ), LD between the paternal and maternal organelle sites transiently exists at the current generation and is of the order similar to the migration rate of seeds or selection coefficients, essentially the same as previous studies (Hu & Li, 2002; Hu, 2008). Partial selfing can produce stable LD between the paternal and maternal organelle genomes.

From eqns (A5)–(A8), the recursion equation for the three-allele cytonuclear LD,  $D_{ABC}^*$  ( $=p_{ABC}^{**} - p_A^* p_B^* p_C^* - p_A^* D_{BC}^* - p_B^* D_{AC}^* - p_C^* D_{AB}^*$ ), is derived as

$$\begin{aligned}
 D_{ABC}^* = & \alpha D_{ABC} - \frac{1}{2} m_P \alpha (1-\alpha) ((Q_B - p_B) D_{AC} \\
 & + (Q_A - p_A) D_{BC}) + m_S (Q_A Q_B Q_C - p_A p_B p_C - p_A \\
 & \times (Q_B - p_B)(Q_C - p_C) - p_B (Q_A - p_A) \\
 & \times (Q_C - p_C) - p_C (Q_A - p_A)(Q_C - p_C) \\
 & + (Q_A - p_A)(\bar{D}_{CD} - p_C p_D - \alpha D_{CD}) + (Q_B - p_B) \\
 & \times \left( \bar{D}_{AC} - p_A p_C - \frac{1+\alpha}{2} D_{AC} \right) + (Q_C - p_C) \\
 & \times \left( \bar{D}_{AB} - p_A p_B - \frac{1+\alpha}{2} D_{AC} \right) \\
 & + \bar{D}_{ABC} - \alpha D_{ABC} + (s_A - s_a) \alpha
 \end{aligned}$$

$$\begin{aligned}
 & \times \left( -(1-\alpha) \left( D_{AB}D_{AC} + \left( p_{APa} - \frac{1}{4}p_{Aa} \right) D_{BC} \right) \right. \\
 & + \left. \frac{1}{4}D_{AaBC} \right) + (s_B - s_b)\alpha \left( -\frac{1-\alpha}{2}p_Bp_bD_{AC} \right. \\
 & - (p_{APB} + D_{AB})D_{BC} + (p_b - p_B)D_{ABC} \left. \right) \\
 & + (s_C - s_c)\alpha \left( -\frac{1-\alpha}{2}p_Cp_cD_{AB} \right. \\
 & - (p_{APC} + D_{AC})D_{BC} + (p_c - p_C)D_{ABC} \left. \right) \\
 & + (s_{AA} - s_{Aa}) \left( (p_a - p_A)D_{AB}D_{AC} \right. \\
 & + \alpha \left( p_a \left( p_{APa} - \frac{1}{4}p_{Aa} \right) D_{BC} - (1 + p_a)D_{AB}D_{AC} \right. \\
 & + (p_a - p_A + p_{APa})D_{ACD} + \frac{1}{8}D_{AB}D_{AaC} \\
 & + \left. \left. \frac{1}{8}D_{AC}D_{AaB} - \frac{1}{4}p_bD_{AaBC} \right) \right) \\
 & + \alpha^2 \left( - \left( p_{APa} - \frac{1}{4}p_{Aa} \right) (p_aD_{BC} + D_{ABC}) \right. \\
 & - p_aD_{AB}D_{AC} + \frac{1}{8}D_{AB}D_{AaC} + \left. \frac{1}{8}D_{AC}D_{AaB} \right) \left. \right) \\
 & + (s_{Aa} - s_{aa}) \left( -(p_a - p_A)D_{AB}D_{AC} \right. \\
 & + \alpha \left( p_A \left( p_{APa} - \frac{1}{4}p_{Aa} \right) D_{BC} - (1 + p_A)D_{AB}D_{AC} \right. \\
 & + (p_a - p_A - p_{APa})D_{ACD} - \frac{1}{8}D_{AB}D_{AaC} \\
 & - \left. \left. \frac{1}{8}D_{AC}D_{AaB} - \frac{1}{4}p_A D_{AaBC} \right) \right) \\
 & + \alpha^2 \left( - \left( p_{APa} - \frac{1}{4}p_{Aa} \right) (p_A D_{BC} - D_{ABC}) \right. \\
 & - p_A D_{AB}D_{AC} - \left. \frac{1}{8}D_{AB}D_{AaC} - \frac{1}{8}D_{AC}D_{AaB} \right) \left. \right). \tag{A9}
 \end{aligned}$$

Like the LD between the paternal and maternal organelle sites, the three-allele cytonuclear LD,  $D_{ABC}$ , transiently exists in the case of random system due to the effects of seed migration and selection in the sporophyte stage. Note that the above expression is different from the  $D$  of Hu & Li (2002, eqn 15) where  $D$  is a composite cytonuclear LD (here, equal to  $p_{ABC} - p_{APB}p_C$ ), and the effects of the LD between the paternal and maternal organelle alleles ( $D_{BC}$ ) are neglected due to their magnitudes of the second order of selection coefficients or migration rates under the conditions of random mating and weak selection.

The change for genotypic cytonuclear LD is derived as

$$\begin{aligned}
 D_{AaB}^* &= (1-\alpha)(p_a - p_A)D_{AB} + \frac{1}{2}\alpha D_{AaB} \\
 & + m_P \left( (1-\alpha)(Q_B - p_B) \left( p_A Q_a + p_a Q_A - 2 \right. \right.
 \end{aligned}$$

$$\begin{aligned}
 & \left. \left. \times (1-\alpha)p_{APa} - \frac{1}{2}\alpha p_{Aa} \right) + (1-\alpha)(p_a - p_A) \right. \\
 & \times \left( \bar{D}_{AB} - D_{AB} \right) + m_S \left( Q_{AaB} - p_B Q_{Aa} - \frac{1}{2}\alpha D_{AaB} \right. \\
 & - (1-\alpha)(p_a - p_A)D_{AB} - \left. \left( 2(1-\alpha)p_{APa} + \frac{1}{2}\alpha p_{Aa} \right) \right. \\
 & \times \left( Q_B - p_B \right) + (s_A - s_a)(1-\alpha)D_{AB} \\
 & \times \left( 1 - 4p_{APa} - 2(1-\alpha)p_{APa} - \frac{1}{2}\alpha p_{Aa} \right) \\
 & + (s_B - s_b)(2(1-\alpha)(\alpha p_{APa} p_B p_b + (p_b - p_B) \\
 & \times (p_a - p_A)D_{AB}) + \frac{1}{2}\alpha \left( -(1-\alpha)p_{Aa} p_B p_b \right. \\
 & + (p_b - p_B)D_{AaB} \left. \right) + (s_C - s_c) \\
 & \times \left( -4(1-\alpha)D_{AB}D_{AC} + \frac{1}{2}\alpha (p_{Aa}D_{BC} + D_{AaBC}) \right) \\
 & + (s_{AA} - s_{Aa}) \left( -(1-\alpha)^2 p_A^2 (4p_a - 1)D_{AB} \right. \\
 & - \alpha(1-\alpha) \left( \left( p_A - \frac{1}{4}p_{Aa} \right) (p_a - p_A)D_{AB} \right. \\
 & - \frac{1}{2}p_{APa}D_{AaB} \left. \right) - \frac{1}{2}\alpha^2 \left( p_A - \frac{1}{2}p_{Aa} \right) D_{AaB} \left. \right) \\
 & + (s_{aa} - s_{Aa}) \left( -(1-\alpha)^2 p_a^2 (1 - 4p_a)D_{AB} \right. \\
 & - \alpha(1-\alpha) \left( \left( p_a - \frac{1}{4}p_{Aa} \right) (p_a - p_A)D_{AB} \right. \\
 & - \frac{1}{2}p_{APa}D_{AaB} \left. \right) - \frac{1}{2}\alpha^2 \left( p_a - \frac{1}{2}p_{Aa} \right) D_{AaB} \left. \right) \\
 & + (s_{AA} - s_{aa})\alpha D_{AB} \left( -2(1-\alpha)p_{APa} - \frac{1}{2}\alpha p_{Aa} \right) \tag{A10}
 \end{aligned}$$

and

$$\begin{aligned}
 D_{AaC}^* &= (1-\alpha)(p_a - p_A)D_{AC} + \frac{1}{2}\alpha D_{AaC} \\
 & + 2m_P(1-\alpha)(Q_a - p_a)D_{AC} \\
 & + m_S \left( Q_{AaC} - p_C Q_{Aa} - \frac{1}{2}\alpha D_{AaC} - (1-\alpha) \right. \\
 & \times (p_a - p_A)D_{AC} - \left. \left( 2(1-\alpha)p_{APa} + \frac{1}{2}\alpha p_{Aa} \right) \right. \\
 & \times (Q_C - p_C) \left. \right) + (s_A - s_a)(1-\alpha)D_{AC} \\
 & \times \left( 1 - 4p_{APa} - 2(1-\alpha)p_{APa} - \frac{1}{2}\alpha p_{Aa} \right) \\
 & + (s_C - s_c)(2(1-\alpha)(\alpha p_{APa} p_C p_c + (p_c - p_C) \\
 & \times (p_a - p_A)D_{AC}) + \frac{1}{2}\alpha \left( -(1-\alpha)p_{Aa} p_C p_c \right. \\
 & + (p_b - p_B)D_{AaC} \left. \right) + (s_B - s_b) \\
 & \times \left( -4(1-\alpha)D_{AB}D_{AC} + \frac{1}{2}\alpha (p_{Aa}D_{BC} + D_{AaBC}) \right) \\
 & + (s_{AA} - s_{Aa}) \left( -(1-\alpha)^2 p_A^2 (4p_a - 1) \right.
 \end{aligned}$$

$$\begin{aligned}
& \times D_{AC} - \alpha(1-\alpha) \left( \left( p_A - \frac{1}{4} p_{Aa} \right) \right. \\
& \times \left. (p_a - p_A) D_{AC} - \frac{1}{2} p_{APa} D_{AaC} \right) \\
& - \frac{1}{2} \alpha^2 \left( p_A - \frac{1}{2} p_{Aa} \right) D_{AaC} \\
& + (s_{aa} - s_{Aa}) \left( -(1-\alpha)^2 p_a^2 (1-4p_A) D_{AC} - \alpha(1-\alpha) \right. \\
& \times \left. \left( \left( p_a - \frac{1}{4} p_{Aa} \right) (p_a - p_A) D_{AC} - \frac{1}{2} p_{APa} D_{AaC} \right) \right. \\
& \left. - \frac{1}{2} \alpha^2 \left( p_a - \frac{1}{2} p_{Aa} \right) D_{AaC} \right) \\
& + (s_{AA} - s_{Aa}) \alpha D_{AC} \left( -2(1-\alpha) p_{APa} - \frac{1}{2} \alpha p_{Aa} \right).
\end{aligned}
\tag{A11}$$

Selection in both the gametophyte and the sporophyte stages can affect the dynamics of genotypic cytonuclear LD. Partial selfing ( $\alpha \neq 0$ ) results in the involvement of higher-order disequilibrium in changing genotypic cytonuclear LD.

I sincerely appreciate Nick H. Barton and two anonymous reviewers for valuable comments and corrections that substantially improved the earlier version of this article.

## References

- Antonovics, J. (1976). The nature of limits to natural selection. *Annals of the Missouri Botanical Garden* **63**, 224–247.
- Arnold, J. (1993). Cytonuclear disequilibrium in hybrid zones. *Annual Review of Ecology and Systematics* **24**, 521–554.
- Asmussen, M. A., Arnold, J. & Avise, J. C. (1987). Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. *Genetics* **115**, 755–768.
- Asmussen, M. A., Arnold, J. & Avise, J. C. (1989). The effects of assortative mating and migration on cytonuclear associations in hybrid zones. *Genetics* **122**, 923–934.
- Asmussen, M. A. & Orive, M. E. (2000). The effects of pollen and seed migration on nuclear-dicytoplasmic systems. I. Nonrandom associations and equilibrium structure with both maternal and paternal cytoplasmic inheritance. *Genetics* **155**, 813–831.
- Asmussen, M. A. & Schnabel, A. (1991). Comparable effects of pollen and seed migration on the cytonuclear structure of plant populations. I. Maternal cytoplasmic inheritance. *Genetics* **128**, 639–654.
- Avise, J. C. 1994. *Molecular Markers, Natural History and Evolution*. New York: Chapman & Hall.
- Avise, J. C., Pierce, P. C., Vanden Avyle, M. J., Smith, M. H., Nelson, W. S. & Asmussen, M. A. (1997). Cytonuclear introgressive swamping and species turnover of bass after an introduction. *Journal of Heredity* **88**, 14–20.
- Barton, N. H. (1992). On the spread of new gene combination in the third phase of Wright's shifting-balance. *Evolution* **46**, 551–557.
- Barton, N. H. & Charlesworth, B. (1984). Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics* **15**, 133–164.
- Barton, N. H. & Rouhani, S. (1993). Adaptation and the shifting balance. *Genetical Research* **61**, 57–74.
- Barton, N. H. & Shpak, M. (2000). The effect of epistasis on the structure of hybrid zones. *Genetical Research* **75**, 179–198.
- Barton, N. H. & Whitlock, M. C. (1997). The evolution of metapopulations. In *Metapopulation Biology: Ecology, Genetics, and Evolution* (eds. I. A. Hanski & M. E. Gilpin), pp. 183–210. San Diego, CA: Academic Press.
- Bennett, J. H. (1954). On the theory of random mating. *Annals of Eugenics* **18**, 311–317.
- Birky, C. W. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences of the USA* **92**, 11331–11338.
- Charlesworth, D. (2006). Evolution of plant breeding systems. *Current Biology* **16**, R726–R735.
- Chen, X. Y., Fan, X. X. & Hu, X. S. (2008). Roles of seed and pollen dispersal in natural regeneration of *Castanopsis fargesii* (Fragaceae): implications for forest management. *Forest Ecology and Management* **256**, 1143–1150.
- Childs, M. R., Echelle, A. A. & Dowling, T. E. (1996). Development of the hybrid swarm between pecos pupfish (Cyprinodontidae: *Cyprinodon pecosensis*) and sheepshead minnow (*Cyprinodon variegatus*): a perspective from allozymes and mtDNA. *Evolution* **50**, 2014–2022.
- Clark, A. G. (1984). Natural selection with nuclear and cytoplasmic transmission. I. A deterministic model. *Genetics* **107**, 679–701.
- Crow, J. F., Engels, W. R. & Denniston, C. (1990). Phase three of Wright's shifting-balance theory. *Evolution* **44**, 233–247.
- Dick, C. W., Hardy, O. J., Jones, F. A. & Petit, R. J. (2008). Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. *Tropical Plant Biology* **1**, 20–33.
- Eckert, C. G., Kalisz, S., Geber, M. A., Sargent, R., Elle, E., Cheptou, P.-O., Goodwillie, C., Johnston, M. O., Kelly, J. K., Moeller, D. A., Porcher, E., Ree, R. H., Vallejo-Marin, M. & Winn, A. A. (2009). Plant mating systems in a changing world. *Trends in Ecology and Evolution* **25**, 35–43.
- Ellstrand, N. C. (2003). Current knowledge of gene flow in plants: implications for transgene flow. *Philosophical Transactions of the Royal Society London B* **358**, 1163–1170.
- Ellstrand, N. C., Prentice, H. C. & Hancock, J. F. (1999). Gene flow and introduction from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* **30**, 539–563.
- Elo, A., Lyznik, A., Gonzalez, D. O., Kachman, S. D. & Mackenzie, S. A. (2003). Nuclear genes that encode mitochondrial proteins for DNA and RNA metabolism are clustered in the *Arabidopsis* genome. *Plant Cell* **15**, 1619–1631.
- Ennos, R. A. (1994). Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72**, 250–259.
- Ennos, R. A., Sinclair, W. T., Hu, X.-S. & Langdon, A. (1999). Using organelle markers to elucidate the history, ecology and evolution of plant populations. In *Molecular Systematics and Plant Evolution* (eds. P. M. Hollingsworth, R. M. Bateman & R. J. Gornall), pp. 1–19. London: Taylor and Francis.

- Fisher, R. A. (1937). The wave of advance of advantageous genes. *Annals of Eugenics* **7**, 255–369.
- Fu, Y. X. & Arnold, J. (1992). Dynamics of cytonuclear disequilibrium in finite populations and comparison with a two-locus nuclear system. *Theoretical Population Biology* **41**, 1–25.
- Goodwillie, C., Kalisz, S. & Eckert, C. G. (2005). The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. *Annual Review of Ecology and Evolution* **36**, 47–79.
- Haldane, J. B. S. (1930). A mathematical theory of natural and artificial selection. Part VI. Isolation. *Proceedings of Cambridge Philosophical Society* **26**, 220–230.
- Haldane, J. B. S. (1956). The relation between density regulation and natural selection. *Philosophical Transactions of The Royal Society London B* **145**, 306–308.
- Heilbuth, J. C. (2000). Lower species richness in dioecious clades. *American Naturalist* **156**, 221–241.
- Hu, X. S. (2005). Tension versus ecological zones in a two-locus system. *Theoretical Population Biology* **68**, 119–131.
- Hu, X. S. (2008). Barriers to the spread of neutral alleles in the cytonuclear system. *Evolution* **62**, 2260–2278.
- Hu, X. S. (2010). Fst in the cytonuclear system. *Theoretical Population Biology* **77**, 105–118.
- Hu, X. S. & Li, B. (2002). Seed and pollen flow and cline discordance among genes with different modes of inheritance. *Heredity* **88**, 212–217.
- Hu, X. S. & Li, B. (2003). On migration load of seeds and pollen grains in a local population. *Heredity* **90**, 162–168.
- Hu, X. S., Zeng, W. & Li, B. L. (2003). Impacts of one-way gene flow on genetic variance components in a natural population. *Silvae Genetica* **52**, 18–24.
- Kirkpatrick, M. & Barton, N. H. (1997). Evolution of a species' range. *American Naturalist* **150**, 1–23.
- Kruuk, L. E. B., Baird, S. J. E., Gale, K. S. & Barton, N. H. (1999). A comparison of multilocus clines maintained by environmental adaptation or by selection against hybrids. *Genetics* **153**, 1959–1971.
- Li, W. H. & Nei, M. (1974). Stable linkage disequilibrium without epistasis in subdivided populations. *Theoretical Population Biology* **6**, 173–183.
- Maroof, M. A. S., Zhang, Q., Neale, D. B. & Allard, R. W. (1992). Associations between nuclear loci and chloroplast genotypes in wild barley. *Genetics* **131**, 225–231.
- Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University.
- Mitton, J. B. (1992). The dynamic mating system of conifers. *New Forests* **6**, 197–216.
- Mogenson, H. L. (1996). The hows and whys of cytoplasmic inheritance in seeds plants. *American Journal of Botany* **83**, 383–404.
- Montoya, D., Zavala, M. A., Rodríguez, M. A. & Purves, D. W. (2008). Animal versus wind dispersal and the robustness of tree species to deforestation. *Science* **320**, 1502–1504.
- Mulcahy, D. L., SariGorla, M. & Mulcahy, G. B. (1996). Pollen selection-past, present and future. *Sexual Plant Reproduction* **9**, 353–356.
- Pakkad, G., Ueno, S. & Yoshimaru, H. (2008). Gene flow pattern and mating system in a small population of *Quercus semiserrata* Roxb. (Fagaceae). *Forest Ecology and Management* **255**, 3819–3826.
- Rand, D. M., Haney, R. A. & Fry, A. J. (2004). Cytonuclear coevolution: the genomics of cooperation. *Trends in Ecology and Evolution* **19**, 645–653.
- Raymond, M. & Marquine, M. (1994). Evolution of insecticide resistance in *Clus pipiens* populations – the Corsican paradox. *Journal of Evolutionary Biology* **7**, 315–337.
- Rouhani, S. & Barton, N. H. (1993). Group selection and the shifting balance. *Genetics Research* **61**, 127–135.
- Schnabel, A. & Asmussen, M. A. (1992). Comparable effects of pollen and seed migration on the cytonuclear structure of plant populations. II. Paternal cytoplasmic inheritance. *Genetics* **132**, 253–267.
- Slatkin, M. (1975). Gene flow and selection in a two-locus system. *Genetics* **81**, 787–802.
- Stebbins, G. L. (1970). Adaptive radiation of reproductive characteristics in angiosperms. (I): pollination mechanisms. *Annual Review of Ecology and Systematics* **1**, 307–326.
- Stoebe, B., Hansmann S., Goremykin V., Kowallik, K. V. & Martin, W. (1999). Proteins encoded in sequenced chloroplast genomes: an overview of gene content, phylogenetics information and endosymbiotic gene transfer to the nucleus. In *Molecular Systematics and Plant Evolution* (eds. P. M. Hollingsworth, R. M. Bateman & R. J. Gornall), pp. 327–352. London: Taylor & Francis.
- Tanksley, S., Zamir, D. & Rick, C. M. (1981). Evidence for extensive overlap of sporophytic and gametophytic gene expression in *Lycopersicon esculentum*. *Science* **213**, 453–455.
- Verma, D. & Daniell, H. (2007). Chloroplast vector systems for biotechnology applications. *Plant Physiology* **145**, 1129–1143.
- Wolf, J. B. (2009). Cytonuclear interactions can favor the evolution of genomic imprinting. *Evolution* **63**, 1364–1371.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics* **16**, 97–159.
- Wright, S. (1969). *Evolution and the Genetics of Populations. Vol. 2: The Theory of Gene Frequencies*. Chicago, IL: The University of Chicago Press.
- Wright, S. (1977). *Evolution and the Genetics of Populations. Vol. 3: Experimental Results and Evolutionary Deductions*. Chicago, IL: The University of Chicago Press.