Theoretical Population Biology 77 (2010) 105–118

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

F_{ST} in the cytonuclear system

Xin-Sheng Hu

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

ARTICLE INFO

Article history:
Received 8 July 2009
Available online 11 December 2009

Keywords:
Cytonuclear disequilibrium
Population structure
Mating system
cpDNA
mtDNA

ABSTRACT

Selection on nuclear (or organelle) sites inevitably affects the spatial distribution of a neutral organelle (or nuclear) allele via transient cytonuclear disequilibrium. Here I examine this effect in terms of F_{st} for a neutral allele by bringing together cytonuclear genomes with contrasting modes of inheritance. The relationships between cytonuclear disequilibrium and increment in F_{st} are explored and confirmed through Monte Carlo simulations. Results show that the transient increment in F_{st} for a neutral allele is not only related to the vectors of seed and pollen dispersal but also to the mode of its inheritance. Such increments can be substantial under certain conditions. Seed dispersal is more effective than pollen dispersal in changing the transient increment. The cumulative effects from multiple selective nuclear sites can amplify the transient increment in F_{st} for a neutral paternal or maternal organelle allele. Selection on selective organelle sites facilitates the transient increment in F_{st} for a neutral nuclear allele. Partial selfing can significantly reinforce the transient increment in F_{st}. These theoretical insights highlight the roles of transient cytonuclear disequilibrium as a biological factor in evolving population differentiation and refine our practical interpretations of F_{st} with cytonuclear markers.

1. Introduction

Wright's F_{st}, a standardized variance of allele frequency among subpopulations, is an important measure for population differentiation (Wright, 1969). In an evolutionary model with the forces of migration and genetic drift, this measure has been extensively used for different purposes in practice, such as the indirect estimation of gene flow (Slatkin, 1985) and the neutrality test at individual loci (Lewontin and Krakauer, 1973). Other evolutionary forces (e.g., mutation and selection) can also modify F_{st} to different extents (Wright, 1969; Whitlock and McCauley, 1999). One important modification of F_{st} for a neutral allele comes from the effects of its linked selective site due to the linkage disequilibrium (LD) between them, broadly termed as genetic hitchhiking effects (Maynard Smith and Haigh, 1974; Charlesworth et al., 1993; Barton, 2000). Our current understanding of genetic hitchhiking effects on F_{st} is mainly attributable to the studies on the strictly nuclear genome system.

When the cytonuclear system, one of the most important symbiotic systems among living organisms (Rand et al., 2004; Dowling et al., 2007), is considered, the transformation from the nuclear to cytonuclear system cannot be simply realized in plant species, except for a large random mating population without genetic drift and migration effects. This is because the evolution for a cytonuclear system in a finite population with genetic drift and migration effects is related not only to the asymmetric vector of genome dispersal, but also to the mode of genome inheritance (Ennos, 1994; Mogenson, 1996). Furthermore, the presence of cytonuclear epistasis can reinforce the complexity of this transformation (Asmussen et al., 1987; Dowling et al., 2007; Wolf, 2009). In relation to the genetic hitchhiking effects, such differences between the two systems lead to the unsuitability of applying the theory for the nuclear system to the cytonuclear system. Incorrect prediction of F_{st} in the cytonuclear system can be resulted if we simply set the recombination rate as 1/2 in the genetic hitchhiking theory for the nuclear system. In reality, influences from cytonuclear genomes are often neglected in interpreting F_{st} with cytonuclear markers, partly due to the lack of appropriate theoretical support. In this study, I bring together different combinations of cytonuclear systems to show how genetic hitchhiking effects change F_{st} for a neutral allele in the cytonuclear system. The genetic basis for such a joint analysis is the presence of cytonuclear disequilibrium.

Ample evidence demonstrates the existence of cytonuclear disequilibrium, primarily in hybrid zones (Arnold, 1993; Monsen et al., 2007), but the type of cytonuclear disequilibrium between neutral and selective sites is rarely clarified for mechanism. As implied from previous studies (Hu and Li, 2002; Hu, 2008), migration can produce this type of cytonuclear disequilibrium in structured populations. The neutral allele frequency at a nuclear (or organelle) site can be altered by selective organelle (or nuclear) sites through cytonuclear disequilibrium. In the nuclear system, LD between selective and neutral nuclear sites generated by random genetic drift can transiently inflate the variance of neutral allele

E-mail address: xin-sheng.hu@ualberta.ca.

0040-5809/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.tpb.2009.12.002

frequencies among subpopulations (Barton, 2000). This is explicitly demonstrated in hermaphroditic plant species (Hu and He, 2005). Thus, it is of both theoretical and practical significance to study how cytonuclear disequilibrium affects F_{st} for a neutral nuclear (or organelle) site.

Like LD between selective and neutral sites in the nuclear system (Barton, 2000), cytonuclear disequilibrium between selective nuclear (or organelle) and neutral organelle (or nuclear) sites is transient. One reason is that recombination between physically unlinked cytonuclear genomes reduces cytonuclear disequilibrium with generations. Another reason is that mutation recurring at either nuclear or organelle sites changes the existing pattern of cytonuclear disequilibrium in structured populations and migrants. Further, genetic drift can facilitate the extinction of neutral alleles. Nevertheless, when the time required for mutation recurring at the focal nuclear or organelle site, ∼100 Myrs (Wolfe et al., 1987), is longer than the time required for the occurrence of a balance among effects of migration, selection, genetic drift, and recombination, stable cytonuclear disequilibrium can be maintained for a certain period of time. This is similar to a transient neutral cline in the nuclear system (Barton, 1979, 1986; Barton and Bengtsson, 1986), or in the cytonuclear system (Hu, 2008). Here, the steady-state cytonuclear disequilibrium, maintained by the joint effects of migration, genetic drift, and selection, is formulated for plants with a random mating system.

The purpose of this study is to examine the effects of selective nuclear (or organelle) sites on F_{st} for a neutral organelle (or nuclear) allele. An earlier study has explored the roles of transient cytonuclear disequilibrium in impeding the spread of a neutral nuclear or organelle allele in an ecological hybrid zone (Hu, 2008). Here, I continue to explore the roles of transient cytonuclear disequilibrium on the spread of a neutral nuclear or organelle allele in space, in terms of F_{st} in the classical island model (Wright, 1969). In my analysis of F_{st} in the cytonuclear system, a mixed mating system is investigated. Selfing facilitates cytonuclear disequilibrium in plant species (Asmussen et al., 1987, 1989; Maroof et al., 1992; Asmussen and Orive, 2000; Overath and Asmussen, 2000), similar to its function in generating LD between nuclear sites. Partial selfing exists in plant species where nuclear and organelle genomes have contrasting modes of inheritance (Mitton, 1992; Mogenson, 1996). In the following sections, three systems are separately examined: effects of the selective nuclear sites on F_{st} for a neutral organelle allele with paternal or maternal inheritance; effects of selective organelle sites on F_{st} for a neutral nuclear allele. Monte Carlo (MC) simulations are used to evaluate transient increments in F_{st} for a neutral allele since few empirical studies are recorded. Inferences on the transient increment in F_{st} are drawn from both analytical and simulation results.

2. General assumptions

The models deal with a hermaphroditic plant species distributing in the form of classical island model of population structure (Wright, 1969), with constant immigration rates of seeds (m_S) and pollen (m_P) per generation in each subpopulation. Throughout the analysis, only diallelic sites are considered for either selective or neutral sites. This is appropriate for most single nucleotide polymorphism (SNP) sites because tri-/tetra-allelic SNP sites are infrequent in natural populations. The selective SNP sites may refer to those adaptive quantitative trait nucleotides (Mackay, 2004). Mutation rate is assumed to be much smaller than the migration rate and its effects are not included. Mutation effects could have a similar mathematical form to migration under the island/continent model (Datta et al., 1996). The genetic variation at the selective sites is subjected to a balance between the effects of selection and migration, similar to the previous considerations in addressing genetic hitchhiking effects or associative overdominance effects except that mutation is not included here (Ohta and Kimura, 1970; Nordborg et al., 1996; Hu and He, 2005). Weak selection is considered throughout this study at the sporophyte stage so that the items containing the second or higher order of selection coefficients are neglected. The cytonuclear epistatic effects on fitness are excluded (Asmussen et al., 1987; Hu, 2008). When multiple selective nuclear sites are included, a multiplicative-viability model is employed to calculate mean population fitness. The life cycle for each subpopulation follows a sequence of events: pollen flow, combination between pollen and ovules (with or without partial selfing), seed flow, diploid selection, genetic drift, and next adults. This procedure is the same as Hu and He (2005) except that the cytonuclear system is emphasized here.

For the effective population size of individual cytonuclear genomes, each subpopulation is assumed to have the same effective population size, N_e. The effective number of diploid nuclear genomes (biparentally inherited) is $2N_e$ in each subpopulation. The effective number of organelle genomes (paternally or maternally inherited) in each subpopulation is N_e by assuming haploid (homoplasmy) organelle genomes (Birky, 1995). This assumption can be relaxed using N_m and N_p for the effective numbers of maternally and paternally inherited genomes, respectively (Hu and Ennos, 1999).

3. Nuclear and paternally inherited organelle genome system

In this section, effects from selective nuclear sites on $F_{st(P)}$ for a neutral organelle allele with paternal inheritance are concentrated. I first look at the mixed mating system (Asmussen et al., 1987), which can naturally occur in many conifer species where chloroplast genomes are paternally inherited (Mitton, 1992; Mogenson, 1996). In this case, only MC simulations are used to evaluate the effects of partial selfing on $F_{st(P)}$ due to the difficulty of analytical deduction. I then look at the random mating system where $F_{st(P)}$ is formulated and evaluated through MC simulations. The results can be applied to plants with the random mating system, or used as an approximation to plants with a nearly complete outcrossing rate (Mitton, 1992).

3.1. Mixed mating system

For simplicity, consider a selective nuclear site and a neutral organelle site, with alleles A_i and a_i at the i nuclear site, and alleles B and b at the neutral organelle site, respectively. Let $p_{A_iA_iB(l)}$ ($l = 1, \ldots, L$) be the frequency of genotype A_iA_iB at the current generation (adults) in the lth subpopulation and $p_{A_i(l)}$ and $p_{B(l)}$ be the frequencies of alleles A_i and B in the lth subpopulation, respectively. Notation for the frequencies of the remaining genotypes and alleles can be denoted in the same way. Let $D_{A_iA_iB(l)}$ ($= p_{A_iA_iB(l)} - p_{A_iA_i(l)}p_{B(l)}$) and $D_{A_iB(l)}$ ($= p_{A_iB(l)} - p_{A_i(l)}p_{B(l)}$) be the genotypic and gametic disequilibria in the lth subpopulation, respectively. Notation for the remaining gametic and genotypic disequilibria can be set in the same way (Asmussen et al., 1987). Some statistical properties for estimating these disequilibria have been explored (Basten and Asmussen, 1997; Dean and Arnold, 1996; Sarchez et al., 2006) and are not repeated further.

Let $1 + s_{i(l)}$ be the genotypic fitness for A_iA_iB or A_iA_ib, $1 + h_{i(l)}s_{i(l)}$ for A_ia_iB or A_ia_ib, and $1 - s_{i(l)}$ for a_ia_iB or a_ia_ib in the lth subpopulation, where $s_{i(l)}$ is the selection coefficient and $h_{i(l)}$ is the degree of dominance. The selective nuclear site evolves independently from the neutral organelle site. The steady-state frequency of A_i in the lth subpopulation at the balance between the effects of migration and selection is derived as

$$0 = \left(m_S + \frac{1-\alpha}{2} m_P\right)(Q_A - p_{A_{i(l)}}) + s_{i(l)}(1-\alpha)p_{A_{i(l)}}p_{a_{i(l)}}\left(1 + (p_{a_{i(l)}} - p_{A_{i(l)}})h_{i(l)}\right) + s_{i(l)}\alpha p_{a_{i(l)}}\left(p_{a_{i(l)}} + \frac{1-(1-\alpha)h_{i(l)}}{2-\alpha}p_{A_{i(l)}}\right), \quad (1)$$

where α is the selfing rate and Q_A is the frequency of A_i in migrants. The steady-state $p_{A_{i(l)}}$ can be numerically calculated using Eq. (1) for each subpopulation.

Following the life cycle of hermaphrodite plants, the frequency of the neutral organelle allele before the occurrence of genetic drift, $p^{***}_{B(l)}$, is derived as

$$p^{***}_{B(l)} = (1-(1-\alpha)m_P - m_S)\,p_{B(l)} + ((1-\alpha)m_P + m_S)\,Q_B + (1-\alpha)s_{i(l)}\left(1 + (p_{a_i(l)} - p_{A_i(l)})h_{i(l)}\right)D_{A_iB(l)} + \alpha s_{i(l)}(D_{A_iA_iB(l)} - D_{a_ia_iB(l)} + h_{i(l)}D_{A_ia_iB(l)}/2), \quad (2)$$

where Q_B is the frequency of B in migrating seeds and pollen. In the presence of partial selfing ($\alpha \neq 0$), the genotypic frequencies in the preceding adults cannot be simply decomposed, leading to the difficulty for further analytical deduction for the cytonuclear disequilibria and the genotypic frequencies at the selective nuclear site. Biologically, Eq. (2) indicates that the change of the neutral allele frequency is dependent on the genetic variation at the selective nuclear site. Partial selfing gives an additional component to the change of the neutral allele frequency. When the selection coefficient equals zero, the cytonulear allele frequencies evolve independently.

MC simulation is employed to evaluate the effects of a mixed mating system. Simulations start from an initial adult reference population that begins subdivision into L subpopulations. All allele frequencies among subpopulations at the selective or neutral site are initially set to be the same as their initial frequencies in the reference population. The frequencies of cytonuclear gametes in migrating pollen and the genotypic frequencies in migrating seeds are assumed to be equal to those in the initial reference population. For simplicity, constant selection coefficients at the nuclear site are considered in all subpopulations, i.e. $s_{i(1)} = \cdots = s_{i(L)} = s$. There is no variation in selective allele frequency among subpopulations. The simulation procedure is based on the life cycle of plants mentioned in the preceding section, given the initial parameter settings. After sampling, genotypic and allelic frequencies are calculated, given the effective subpopulation size (N_e), and $F_{st(P)}$ is then calculated. These steps are repeated until the $F_{st(P)}$ reaches steady-state distribution from generation to generation. Independent simulation results are used to calculate mean and standard deviation of $F_{st(P)}$.

Simulation results show that partial selfing can increase $F_{st(P)}$ for the neutral organelle allele (Fig. 1a). Compared with the steady-state $F_{st(P)}(= (1 + 2N_e(m_S + m_P))^{-1} = 0.0769$, Ennos, 1994) in a pure neutral process, the steady increment is about 13.97%, 20.46%, and 71.68% at the 50th generation when α is 0.0, 5%, and 50%, respectively. Note that the increment at $\alpha = 0.0$ (random mating system) is due to the joint effects of seed flow ($m_S = 0.04$) and selection on the selective nuclear site ($s_{i(l)} = 0.04, l = 1, \ldots, L$). This can also be viewed from Eq. (2) by setting $\alpha = 0$. The standard deviation also approaches steady distribution although a high level of selfing rate increases the fluctuation of $F_{st(P)}$ and raises the standard deviations (Fig. 1b).

Fig. 2 shows that natural selection at the nuclear site facilitates the increment in $F_{st(P)}$ for the neutral organelle allele. Compared with the steady-state $F_{st(P)}$ under a pure neutral process, the steady increment can reach about 11.96% and 20.46% at the 50th generation when selection coefficient $s_{i(l)}(= s)$ is 0.01 and 0.05, respectively (Fig. 2a). Strong selection on the selective nuclear site can substantially increase $F_{st(P)}$ and bring about high fluctuation as well (Fig. 2b). The indirect effects from selection at the nuclear site are mediated through changing cytonuclear disequilibria (results not shown here).

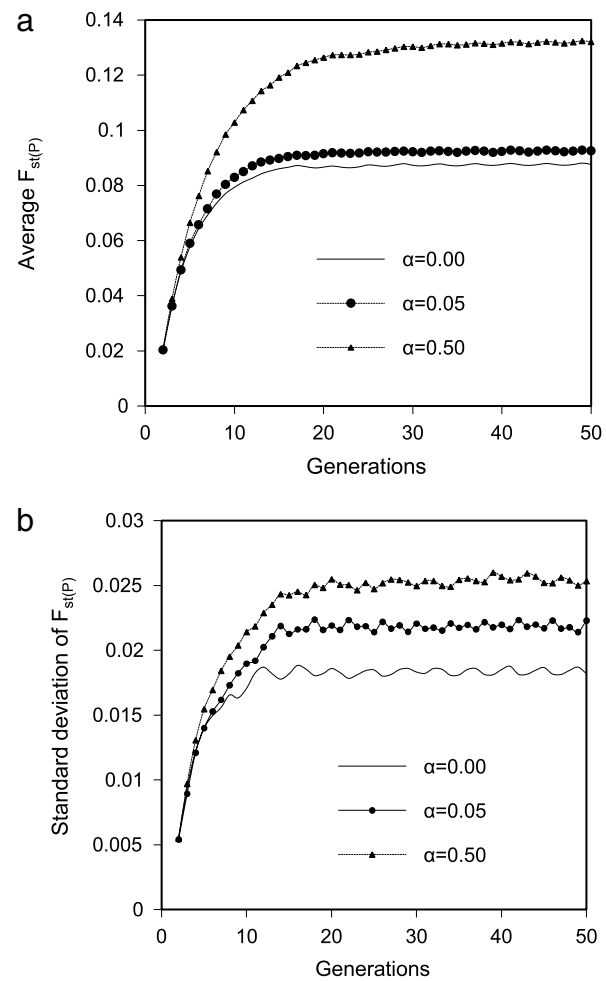

Fig. 1. Effects of partial selfing (α) on F_{st} for a neutral paternal organelle allele: (a) the change of average F_{st} with time; (b) the standard deviation of F_{st}. The results are obtained from 5000 independent simulations, with the migration rates of seeds $m_S = 0.04$ and pollen $m_P = 0.08$, the degree of dominance $h = 0.01$, the selection coefficient $s = 0.04$, the number of subpopulations $L = 30$, and the effective population size $N_e = 50$. The initial genotypic frequencies (the same as those in migrants) are set as $Q_{AAB} = 0.45$, $Q_{AaB} = 0.03$, $Q_{aaB} = 0.02$, $Q_{AAb} = 0.02$, $Q_{Aab} = 0.03$, and $Q_{aab} = 0.45$.

3.2. Random mating system

Consider $T(\geq 2)$ diallelic selective nuclear sites, with allele frequencies $p_{A_i(l)}$ and $p_{a_i(l)}$ ($p_{A_i(l)} + p_{a_i(l)} = 1$; $i = 1, 2, \ldots, T$; $l = 1, \ldots, L$) for A_i and a_i at the i site in the lth subpopulation, respectively. Similarly, let $1 + s_{i(l)}$ be the genotypic fitness for A_iA_i, $1+h_{i(l)}s_{i(l)}$ for A_ia_i and $1-s_{i(l)}$ for a_ia_i in the lth subpopulation. Mean population fitness can be calculated according to the assumptions of a multiplicative-viability model with weak selection. Following the life cycle of hermaphroditic plants, the changes in cytonuclear genotypic frequencies, in the neutral organelle allele frequency, and in the selective nuclear allele frequencies, can be formulated.

The T selective nuclear sites evolve independently from the neutral organelle site. Appendix A derives the results for the changes of allele frequencies and LDs in the case of T selective nuclear sites in one subpopulation. The steady-state allele frequencies can be reached under joint effects of migration that leads

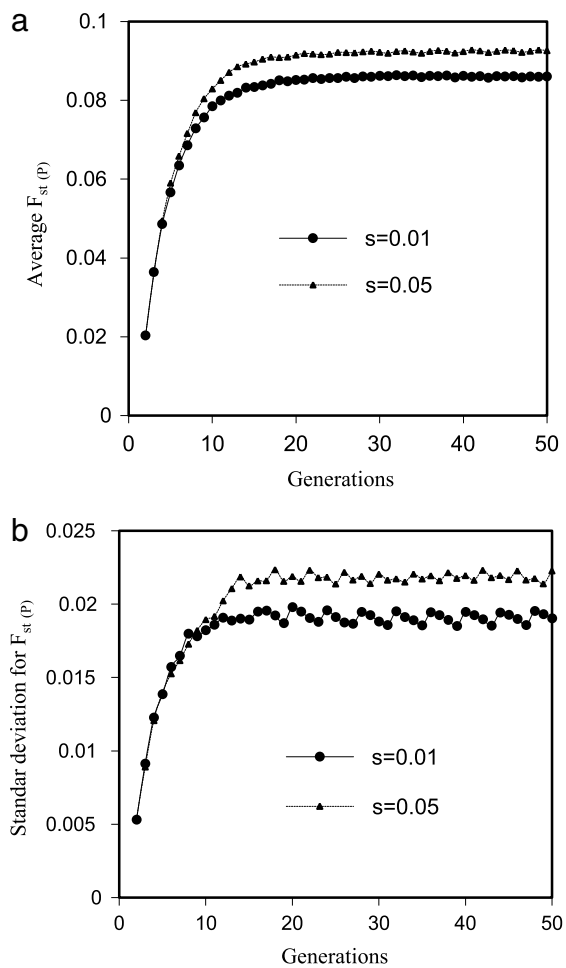

Fig. 2. Effects of selection on a nuclear site on F_{st} for a neutral paternal organelle allele: (a) the change of average F_{st} with time; (b) the standard deviation of F_{st}. The results are obtained from 5000 independent simulations, with the migration rates of seeds $m_S = 0.04$ and pollen $m_P = 0.08$, the degree of dominance $h = 0.01$, the selfing rate=5%, the selection coefficient $s = 0.04$, the number of subpopulations $L = 30$, and the effective population size $N_e = 50$. The initial genotypic frequencies (the same as those in migrants) are set as $Q_{AAB} = 0.45$, $Q_{AaB} = 0.03$, $Q_{aaB} = 0.02$, $Q_{AAb} = 0.02$, $Q_{Aab} = 0.03$, and $Q_{aab} = 0.45$.

allele frequencies to the frequencies in migrants and selection that increases adaptive allele frequencies. The steady-state LDs can be reached under the joint effects of recombination, migration, and selection. At the steady state, there are $T + T(T-1)/2$ joint equations that are difficult to solve analytically. Without loss of examining the cumulative effects from multiple selective nuclear sites on the neutral organelle site, a coincidence among T allele frequencies is considered. This consideration is similar to the analysis of coincidence of allele frequency clines in hybrid zones (Kruuk et al., 1999). Although the cumulative effects in this specific case may be different from a more general case (a discordance among allele frequencies), the qualitative results are expected to be the same. The steady-state allele frequencies and LDs can be calculated under the coincidence among the T allele frequencies (Appendix A). These results are applicable to any subpopulation.

Under the influences of T selective nuclear sites, the neutral allele frequency can be readily obtained from the results for the two nuclear sites shown in Eq. (A.3) in Appendix A. The frequency before the occurrence of genetic drift in the lth subpopulation is written as

$$p^{***}_{B(l)} = (1-\tilde{m}')p_{B(l)} + \tilde{m}'Q_B + \Lambda_{1(l)}, \tag{3}$$

where $\tilde{m}' = m_S + m_P$, $\Lambda_{1(l)} = \sum_{i=1}^{T} s_{i(l)}\left(1 + (p_{a_i(l)} - p_{A_i(l)})h_{i(l)}\right) D_{A_iB(l)}$ in which $D_{A_iB(l)}$ is the cytonuclear disequilibrium between A_i and B in the lth subpopulation. $\Lambda_{1(l)}$ represents the cumulative effects from the T selective nuclear sites in the lth subpopulation. Eq. (3) indicates that the neutral organelle allele evolves independently from the T selective nuclear sites when $D_{A_iB(l)} = 0$.

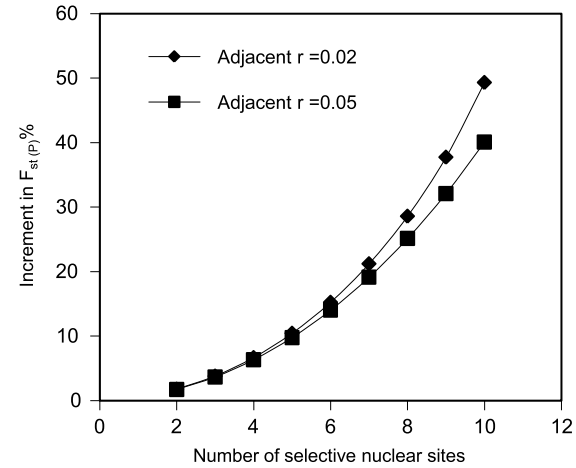

Fig. 3. Cumulative effects from multiple selective nuclear sites on the increment in F_{st} for a neutral paternal organelle allele. Results are calculated according to Eq. (4), with the migration rates of seeds $m_S = 0.04$ and pollen $m_P = 0.08$, the degree of dominance $h = 0.0$, the selection coefficient $s = 0.03$, the allele frequencies in migrants $Q_A = Q_B = 0.5$, the linkage disequilibrium between adjacent nuclear sites $\bar{D} = 0.15$, the cytonuclear disequilibrium $\bar{D}_{AB} = 0.15$, the number of subpopulations $L = 30$, and the effective population size $N_e = 50$.

The presence of transient cytonuclear disequilibrium can modify population differentiation to some extents. Appendix B details the analysis of population differentiation for the neutral organelle allele. At the steady state, population differentiation for the neutral organelle allele can be simply expressed as

$$F_{st(P)} = \frac{1}{1 + 2N_e\tilde{m}'}\,(1+\rho_1)\,, \tag{4}$$

where ρ_1 is the change induced by the cytonuclear disequilibria. Appendix B shows that ρ_1 is a complex function of the allele frequencies, migration rate, and the effective population size. Compared with the result under a pure neutral process (Ennos, 1994; Hu and Ennos, 1999), the relative increment in $F_{st(P)}$ is equal to $100\rho_1$%.

To assess quantitatively the cumulative effects of multiple selective nuclear sites, a specific case is considered, where equal selection coefficients exist among subpopulations and the coincidence of allele frequencies among the T sites takes place in each subpopulation. Under this specific situation, the expected effects of cytonuclear disequilibria on the frequency of the neutral organelle allele among subpopulations are equal among subpopulations as well. Appendix B details the methods for calculating ρ_1.

Fig. 3 shows how the increment of $F_{st(P)}$ at a neutral organelle site changes with the number of selective nuclear sites. Generally, the relative increment ($100\rho_1$%) increases with the number of selective nuclear sites. For example, the relative increment changes from 1.78% to 49.33% as the number of selective nuclear sites increases from 2 to 10 under certain parameter settings (Fig. 3). When more selective nuclear sites are closely linked on chromosomes, the cumulative effects are higher, especially in the case of many closely linked selective nuclear sites (Fig. 3). A large effective population size facilitates this increment due to its effects on cytonuclear disequilibrium (results not shown here).

To confirm the above-predicted pattern, MC simulations are conducted. The number of cytonuclear genotypes, 2×3^T (excluding the difference in linkage phases) for one diallelic organelle and T diallelic nuclear sites, increases quickly as the number nuclear sites increases, which makes the simulation difficult to implement.

Thus, only the simulation results for the two and three nuclear sites are presented.

Fig. 4(a and b) shows that the average $F_{st(P)}$ at the steady state increases as the number of selective nuclear sites changes from two to three, with the average relative increment being 2.84% ($F_{st(P)} = 0.0791 \pm 0.013$; $F_{st(P)} = 0.0769$ for the pure neutral case) and 7.67% ($F_{st(P)} = 0.0828 \pm 0.02$), respectively. Note that the parameter settings in Fig. 4 are comparable for allele frequencies and cytonuclear disequilibria between the two- and three-site cases except for one more selective site in the three-site case. The theoretical predicts for the relative increment of $F_{st(P)}$ under the comparable parameter settings are 1.78% and 3.81% for the two- and three-site cases (Fig. 3), respectively. The simulation results are greater than the theoretical predicts because of the limited number of subpopulations. The theoretical results (Appendix B) are calculated according to the density distribution of neutral allele frequency and cytonuclear disequilibrium where a large number of subpopulations are assumed.

MC simulations confirm the predicted pattern about the effects of effective population size on the relative increment of $F_{st(P)}$. For example, let $m_S = 0.04$, $m_P = 0.08$, $h = 0.0$, $s = 0.03$, the adjacent recombination rate = 0.02, $L = 30$, $Q_{A_1} = Q_{A_2} = 0.6$, $Q_B = 0.5$, $\bar{D}_{A_1A_2B} = \bar{D}_{a_1a_2b} = 0.15$ and all other cytonuclear disequilibria = −0.05. Simulation results show that the relative increment in $F_{st(P)}$ at steady state is about 24.0% ($F_{st(P)} = 0.0496 \pm 0.007$; $F_{st(P)} = 0.04$ for the pure neutral case) with the effective population size being 100 in the case of two selective nuclear sites. The average relative increment is greater in the case of $N_e = 100$ than in the case of $N_e = 50$ in Fig. 4. Note that all other parameter settings are comparable between the cases of $N_e = 50$ and 100. Again, the simulation result ρ_1 is greater than the predict from Eq. (4) ($\rho_1\% = 2.96\%$) under the comparable parameter settings owing to the limited number of subpopulations. However, the theoretical predicted results in each case are within the range of one standard deviation of the simulation results.

4. Nuclear and maternally inherited organelle genome system

This system is applicable to the combination of nuclear and mitochondrial genomes in both angiosperms and gymnosperms where mitochondrial genomes are maternally inherited, or to the combination of nuclear and chloroplast genomes in angiosperms where chloroplast genomes are maternally inherited (Mogenson, 1996). Only seed dispersal can spread the maternally inherited organelle genomes. Effects of pollen dispersal on $F_{st(M)}$ for a neutral organelle allele are realized through changing the cytonuclear disequilibrium mainly generated by pollen and seed dispersal. Some properties of this type of cytonuclear disequilibrium have been explored in different models (Datta et al., 1996; Datta and Arnold, 1998). In the following, I first look at the case of the mixed mating system occurring in some angiosperms and gymnosperms (Mitton, 1992; Jesson and Barrett, 2002) and then the case of random mating system.

4.1. Mixed mating system

Consider one diallelic selective nuclear site, with alleles A_i and a_i, and one diallelic neutral organelle site, with alleles E and e. Let $p_{A_iA_iE(l)}$ be the frequency of genotype A_iA_iE at the current generation (adults) in the lth subpopulation ($l = 1, \ldots, L$), and $p_{E(l)}$ be the frequency of the E allele in the lth subpopulation. The frequencies of remaining genotypes and alleles can be denoted in the same way. Let $D_{A_iA_iE(l)} (= p_{A_iA_iE(l)} - p_{A_iA_i(l)}p_{E(l)})$ and $D_{A_iE(l)} (= p_{A_iE(l)} - p_{A_i(l)}p_{E(l)})$ be the genotypic and gametic disequilibria in the lth subpopulation, respectively. All remaining genotypic and gametic disequilibria for the cytonuclear system can be defined in a similar way to Asmussen et al. (1987).

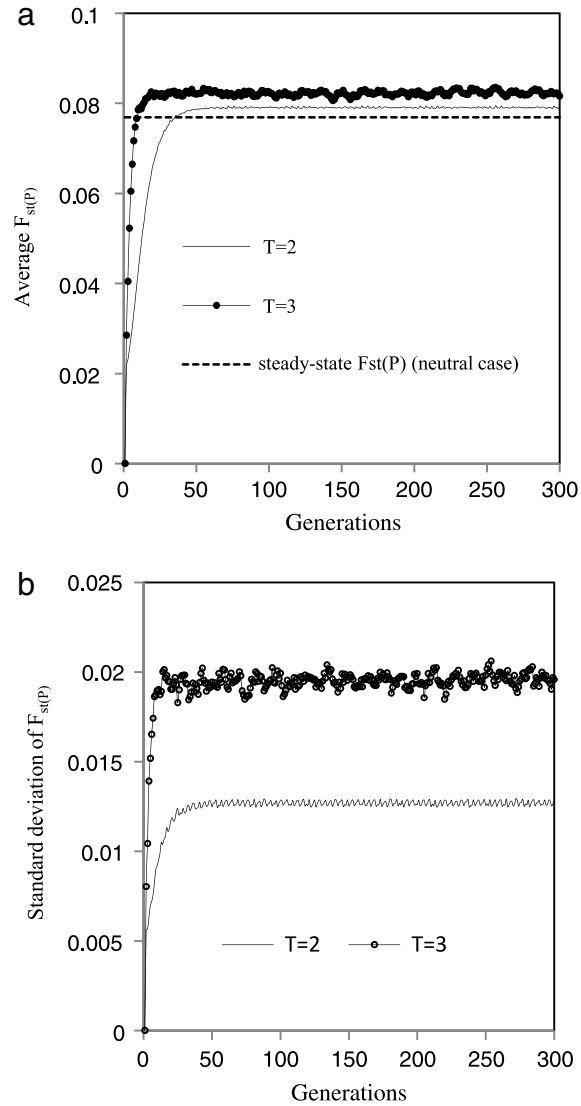

Fig. 4. Cumulative effects from two and three selective nuclear sites on F_{st} for a neutral paternal organelle allele: (a) the change of average F_{st} with time; (b) the standard deviation of F_{st}. The results are obtained from 2000 independent simulations, with the migration rates of seeds $m_S = 0.04$ and pollen $m_P = 0.08$, the degree of dominance $h = 0.0$, the selection coefficient $s = 0.03$, the adjacent recombination rate=0.02, the number of subpopulations $L = 30$, the effective population size $N_e = 50$. In the two-site case, the selective allele frequencies in the initial genotypic frequencies (or in migrants) $Q_{A_1} = Q_{A_2} = 0.6$, the neutral allele frequency $Q_B = 0.5$, the genotypic cytonuclear disequilibra $\bar{D}_{A_1A_2B} = \bar{D}_{a_1a_2b} = 0.15$ and all other cytonuclear disequilibria =−0.05. In the three-site case, $Q_{A_1} = Q_{A_2} = Q_{A_3} = 0.6$, $Q_B = 0.5$, and $\bar{D}_{A_1A_2A_3B} = \bar{D}_{a_1a_2a_3b} = 0.175$ and all other cytonuclear disequilibra=−0.025.

Let genotypic fitness be $1 + s_{i(l)}$ for A_iA_iE or A_iA_ie, $1 + h_{i(l)}s_{i(l)}$ for A_ia_iE or A_ia_ie, and $1 - s_{i(l)}$ for a_ia_iE or a_ia_ie in the lth subpopulation. Again, the selective nuclear site evolves independently from the neutral organelle site. The steady-state allele frequency at the selective site can be described by Eq. (1) and calculated through the iterative approach.

Following the life cycle of plants, the frequency of the neutral organelle allele before the occurrence of genetic drift, $p^{***}_{E(l)}$, is derived as

$$\begin{aligned} p^{***}_{E(l)} = {} & (1 - m_S)\, p_{E(l)} + m_S Q_E \\ & + (1 - \alpha) s_{i(l)} \left(1 + (p_{a_i(l)} - p_{A_i(l)}) h_{i(l)}\right) D_{A_iE(l)} \\ & + \alpha s_{i(l)} (D_{A_iA_iE(l)} - D_{a_ia_iE(l)} + h_{i(l)} D_{A_ia_iE(l)}/2), \end{aligned} \tag{5}$$

where Q_E is the frequency of E in migrating seeds. Again, the presence of partial selfing ($\alpha \neq 0$) introduces an additional part, i.e. genotypic cytonuclear disequilibria, in changing neutral allele frequency. In the case of random mating ($\alpha = 0$), the effects from the selective nuclear site are mediated through gametic cytonuclear disequilibrium. Eq. (5) explicitly describes the evolutionary dependence of the neutral organelle site on the selective nuclear site. The difference between Eqs. (1) and (5) is that the migration rate of pollen (m_P) is absent in Eq. (5) because pollen dispersal does not spread the neutral organelle allele with maternal inheritance.

Using the same method as in the preceding section, MC simulations are employed to examine the effects of selfing on $F_{st(M)}$ for the neutral organelle allele. Constant selection coefficients at the nuclear site are considered in all subpopulations, i.e. $s_{i(1)} = \cdots = s_{i(L)} = s$. Fig. 5 demonstrates how partial selfing affects $F_{st(M)}$ for the neutral maternal organelle allele. The steady-state $F_{st(M)}$ increases when the selfing rate increases (Fig. 5a). Compared with the steady-state $F_{st(M)}$ under a pure neutral process ($= (1 + 2N_e m_S)^{-1} = 0.2$; Ennos, 1994), the steady-state increment is about 4.08%, 5.75%, and 15.15% at the 100th generation when α is 0.0, 5%, and 50%, respectively. The steady-state standard deviations increase when the selfing rate increases (Fig. 5b). These results indicate that a certain increment in $F_{st(M)}$ can be produced under the influences of a single selective nuclear site and high selfing rates.

MC simulations show the presence of different effects of pollen and seed dispersal on $F_{st(M)}$. For example, let $\alpha = 5\%$, $h = 0.01$, $s = 0.04$, $L = 30$, $N_e = 50$, and the same initial genotypic frequencies as the settings in Fig. 5. Results show that in comparison with the steady-state $F_{st(M)}$ with $m_S = 0.04$ and $m_P = 0.08$ ($F_{st(M)} = 0.2115 \pm 0.0388$), the average increment can reach about 137.90% with $m_S = 0.01$ and $m_P = 0.08$ ($F_{st(M)} = 0.5031 \pm 0.0567$), and 6.02% with $m_S = 0.04$ and $m_P = 0.02$ ($F_{st(M)} = 0.2242 \pm 0.0381$), the same for the reduction of the migration rate ($0.04 - 0.01 = 0.03$ for the change of the seeds dispersal, $(0.08 - 0.02)/2 = 0.03$ for the pollen dispersal). This indicates that the effect of pollen dispersal is not as high as that of seed dispersal, given the same changing rate in migration. However, when pollen dispersal is substantially greater than seed dispersal (Ennos, 1994; Ennos et al., 1999), the indirect effects of pollen dispersal can be large as well.

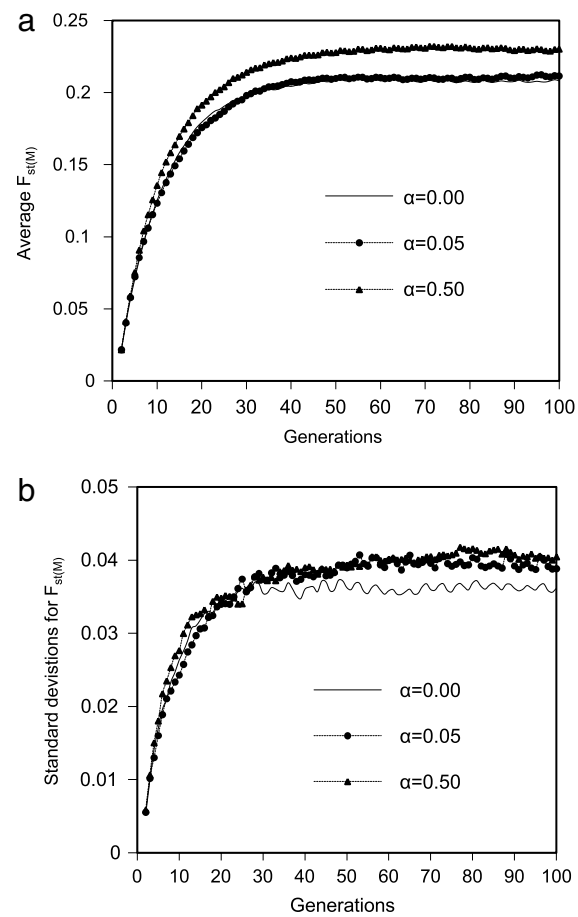

Fig. 5. Effects of selfing (α) on F_{st} for a neutral maternal organelle allele: (a) the change of average F_{st} with time; (b) the standard deviation of F_{st}. The results are obtained from 5000 independent simulations, with the migration rates of seeds $m_S = 0.04$ and pollen $m_P = 0.08$, the degree of dominance $h = 0.01$, the selection coefficient $s = 0.04$, the number of subpopulations $L = 30$, and the effective population size $N_e = 50$. The initial genotypic frequencies (the same as those in migrants) are set as $Q_{AAE} = 0.45$, $Q_{AaE} = 0.03$, $Q_{aaE} = 0.02$, $Q_{AAe} = 0.02$, $Q_{Aae} = 0.03$, and $Q_{aae} = 0.45$.

4.2. Random mating system

Assume $T(\geq 2)$ diallelic selective nuclear sites, with allele frequencies $p_{A_i(l)}$ and $p_{a_i(l)}$ ($p_{A_i(l)} + p_{a_i(l)} = 1$; $i = 1, 2, \ldots, T$; $l = 1, \ldots, L$) for A_i and a_i at the i site in the lth subpopulation, respectively. Similarly, let $1 + s_{i(l)}$ be the fitness of genotype A_iA_i, $1 + h_{i(l)}s_{i(l)}$ for A_ia_i and $1 - s_{i(l)}$ for a_ia_i in the lth subpopulation. The results for the changes of all T selective nuclear sites remain the same as those in the preceding cytonuclear systems.

Based on the results in Appendix C, the allele frequency at the neutral organelle site under the influences of T selective nuclear sites before the occurrence of genetic drift can be expressed as

$$p^{***}_{E(l)} = (1 - m_S)p_{E(l)} + m_S Q_B + \Lambda_{2(l)}, \tag{6}$$

where $\Lambda_{2(l)} = \sum_{i=1}^{T} s_{i(l)} \left(1 + (p_{a_i(l)} - p_{A_i(l)})h_{i(l)}\right) D_{A_iE(l)}$ in which $D_{A_iE(l)}$ is the cytonuclear disequilibrium between A_i and E in the lth subpopulation. The change for $D_{A_iE(l)}$ before the occurrence of genetic drift is given in Appendix C. Eq. (6) is analogous to Eq. (3), describing the cumulative effects from the T selective nuclear sites.

After genetic drift, population differentiation for the neutral organelle allele is derived in Appendix C. At the steady state, $F_{st(M)}$ is given by

$$F_{st(M)} = \frac{1}{1 + 2N_e m_S}(1 + \rho_2), \tag{7}$$

where ρ_2 is the change due to cytonuclear disequilibrium. Compared with the result under a pure neutral process (Hu and Ennos, 1999), the relative increment in $F_{st(M)}$ is equal to $100\rho_2\%$.

To examine the cumulative effects from the T selective nuclear sites, a specific case is analyzed where equal selection coefficients exist among subpopulations and the coincidence of allele frequencies among the T sites takes place in each subpopulation. Numerical results based on Eq. (7) show that a pattern similar to that in Fig. 3 can be obtained. The increment percentage of $F_{st(M)}$ at the neutral organelle site increases with the number of selective nuclear sites (results are not shown here). Tight linkage among the T sites facilitates the cumulative effects. For instance, about 50% of increment in $F_{st(M)}$ can be reached, with parameters $T = 10$, the recombination rate 0.02 between adjacent selective nuclear sites, $m_S = 0.04$, $m_P = 0.08$, $h = 0.0$, $s = 0.03$, $Q_A = Q_E = 0.5$, LD between adjacent selective nuclear sites $\bar{D} = 0.15$, the cytonuclear disequilibrium $\bar{D}_{AE} = 0.15$, $L = 30$, and $N_e = 50$.

MC simulations confirm the pattern about the cumulative effects from multiple selective nuclear sites and the effects of effective population size on $F_{st(M)}$. For example, the average relative increment in $F_{st(M)}$ increases from 6.65% to 15.85% when the number of selective nuclear sites changes from two to three sites (Fig. 6a and b; the steady-state $F_{st(M)} = 0.2133 \pm 0.047$, 0.2317 ± 0.05, 0.2 for the two, three, and pure neutral case, respectively). The average relative increments in $F_{st(M)}$ from MC simulations are greater than

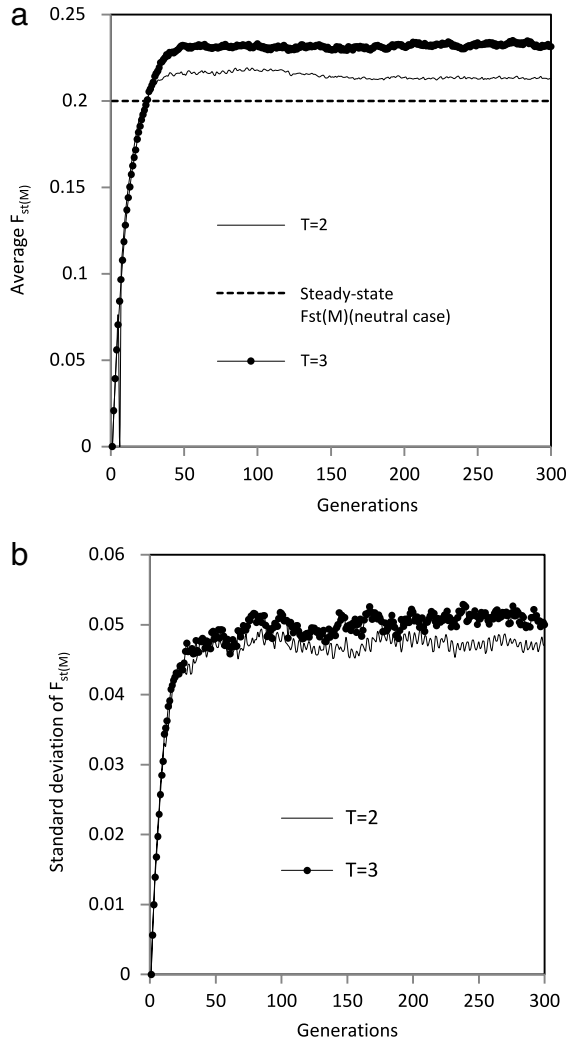

Fig. 6. Cumulative effects from two and three selective nuclear sites on F_{st} for a neutral maternal organelle allele: (a) the change of average F_{st} with time; (b) the standard deviation of F_{st}. The results are obtained from 2000 independent simulations, with the migration rates of seeds $m_S = 0.04$ and pollen $m_P = 0.08$, the degree of dominance $h = 0.0$, the selection coefficient $s = 0.03$, the adjacent recombination rate $= 0.02$, the number of subpopulations $L = 30$, the effective population size $N_e = 50$. In the two-site case, the selective allele frequencies in the initial genotypic frequencies (or in migrants) $Q_{A_1} = Q_{A_2} = 0.5$, the neutral allele frequency $Q_E = 0.3$, the genotypic cytonuclear disequilibria $\bar{D}_{A_1A_2E} = \bar{D}_{a_1a_2e} = 0.09$ and all other cytonuclear disequilibria $= -0.03$. In the three-site case, $Q_{A_1} = Q_{A_2} = Q_{A_3} = 0.5$, $Q_E = 0.3$, and $\bar{D}_{A_1A_2A_3B} = \bar{D}_{a_1a_2a_3b} = 0.105$ and all other cytonuclear disequilibria $= -0.015$.

those predicted from Eq. (7), 0.96% and 1.67% for the two and three nuclear sites, respectively, owing to the limited number of subpopulations used in simulations ($L = 30$). Again, the predicted results are within the range of one standard deviation of the simulation results.

5. Nuclear and paternally and maternally inherited organelle genome system

In this section, the influences of selection on organelle genomes on $F_{st(B)}$ for a neutral nuclear allele are concentrated, and three plant genomes with contrasting modes of inheritance are jointly considered. The results can be applied to some gymnosperms where chloroplast genomes are paternally inherited while mitochondrial genomes are maternally inherited (Mogenson, 1996). Biologically, the interaction among the three genomes can be introduced by various mechanisms (Rand et al., 2004; Wolf, 2009). Here, their interactions come from the cytonuclear disequilibria that are mainly generated by seed and pollen dispersal. In the following, I first examine the effects of partial selfing on $F_{st(B)}$ and then formulate the analytical expression for $F_{st(B)}$ in the random mating case.

5.1. Mixed mating system

Let N and n be the two neutral nuclear alleles, C and c be the two selective alleles on the organelle genomes with maternal inheritance, and D and d be the two selective alleles on the genomes with paternal inheritance. There are 12 cytonuclear genotypes, i.e. $NNDC, \ldots,$ and $nndc$, with genotypic frequencies $p_{NNDC(l)}, \ldots,$ and $p_{nndc(l)}$ in the lth subpopulation, respectively. In the pollen grains, the four gametic frequencies are $p_{ND(l)}$, $p_{Nd(l)}$, $p_{nD(l)}$, and $p_{nd(l)}$. Let $D_{ND(l)} (= p_{ND(l)} - p_{N(l)}p_{D(l)})$ be the gametic cytonuclear disequilibrium between N and D alleles. Four gametic frequencies in ovules are $p_{NC(l)}, p_{Nc(l)}, p_{nC(l)}$, and $p_{nc(l)}$, different from those in pollen. Similarly, let $D_{NC(l)} (= p_{NC(l)} - p_{N(l)}p_{C(l)})$ be the gametic cytonuclear disequilibrium between N and C alleles.

Let the fitness be $1 + s_{C(l)}/2$ for C, $1 - s_{C(l)}/2$ for c, $1 + s_{D(l)}/2$ for D, and $1 - s_{D(l)}/2$ for d in the lth subpopulation. The joint fitness for any two-site genotype can be calculated using the multiplicative-viability model and the assumption of weak selection. The average population fitness is derived as $1 + s_{D(l)}(p_{D(l)} - p_{d(l)})/2 + s_{C(l)}(p_{C(l)} - p_{c(l)})/2$, independent of the selfing rate α. The steady-state frequencies for the B and C alleles in the lth subpopulation at the balance between migration and selection are respectively expressed as

$$0 = (m_S + (1-\alpha)m_P)(Q_D - p_{D_{(l)}}) + s_{D(l)}p_{D_{(l)}}p_{d_{(l)}} + \alpha s_{C(l)}D_{DC(l)}, \tag{8}$$

$$0 = m_S(Q_C - p_{C_{i(l)}}) + s_{C(l)}p_{C_{(l)}}p_{c_{(l)}} + \alpha s_{D(l)}D_{DC(l)}, \tag{9}$$

where Q_D and Q_C are the frequencies of D and C in migrants, and $D_{DC(l)} (= p_{DC(l)} - p_{D(l)}p_{C(l)})$ is the disequilibrium between the two organelle sites. The steady-state equation for the disequilibrium between the two organellele genomes, $D_{DC(l)}$, is

$$D_{DC(l)} = m_S(p_{D(l)} - Q_D)(p_{C(l)} - Q_C) + m_S\bar{D}_{DC} + \alpha\left(1 - m_S + s_{D(l)}(p_{d(l)} - p_{D(l)}) + s_{C(l)}(p_{c(l)} - p_{C(l)})\right)D_{DC(l)}, \tag{10}$$

where $\bar{D}_{DC}$ is the disequilibrium between D and C in migrants. Eq. (10) indicates that $D_{DC(l)}$ is maintained by partial selfing and seed dispersal. The above three joint equations can be used to calculate the frequencies of C and D.

Following the life cycle of hermaphroditic plants, the frequency of the neutral nuclear allele before genetic drift in the lth subpopulation is derived as

$$p^{***}_{N(l)} = (1 - (1-\alpha)m_P/2 - m_S)\,p_{N(l)} + ((1-\alpha)m_P/2 + m_S)\,Q_N + s_{C(l)}\left((1-\alpha)D_{NC(l)}/2 + \alpha\Delta_1\right) + s_{D(l)}\left((1-\alpha)D_{ND(l)}/2 + \alpha\Delta_2\right), \tag{11}$$

where $\Delta_1 = p_{NNDC(l)} + p_{NNdC(l)} - p_{NNDc(l)} - p_{NNdc(l)} + \left(p_{NnDC(l)} + p_{NndC(l)} - p_{NnDc(l)} - p_{Nndc(l)}\right)/2$, $\Delta_2 = p_{NNDC(l)} + p_{NNDc(l)} - p_{NNdC(l)} - p_{NNdc(l)} + \left(p_{NnDC(l)} + p_{NnDc(l)} - p_{NndC(l)} - p_{Nndc(l)}\right)/2$, and Q_N is the frequency of N in the migrants. Here, Δ_1 and Δ_2 are related to the genotypic disequilibria introduced by partial selfing. Eq. (11) indicates that both selective organelle sites can change the frequency of the neutral nuclear allele when selection coefficients are not equal to zero.

The procedure similar to the preceding two cytonuclear systems in MC simulation is used to look at the effects of a mixed mating system on $F_{st(B)}$ for the neutral nuclear allele. Constant

selection coefficients at the organelle sites are considered in all subpopulations, i.e. $s_{C(1)} = \cdots = s_{C(L)} = S_C$ and $s_{D(1)} = \cdots = s_{D(L)} = S_D$. Fig. 7a shows the increase in F_{st} with the selfing rate. Compared with the steady-state $F_{st(B)}$ under a pure neutral process ($=(1+4N_e\tilde{m})^{-1} = 0.0588$ with the same parameters as in Fig. 7; Ennos, 1994), the steady increment can reach about 17.08%, 23.84%, and 99.84% at the 100th generation when α is 0.0, 5%, and 50%, respectively. The steady-state standard deviation increases with the selfing rate (Fig. 7b).

To assess the effects of seed and pollen dispersal on $F_{st(B)}$, extensive simulations are conducted (not given here). In the three-genome system with contrasting modes of inheritance, the results indicate that an increase in migration rate of seeds can increase the relative increment in $F_{st(B)}$ under certain conditions (when the migration rate is comparable or smaller than the selection coefficients). An increase in migration rate of pollen tends to reduce the relative increment in $F_{st(B)}$ in the three-genome system. For example, let $\alpha = 5\%, s_C = s_D = 0.04, L = 30, N_e = 50$, and the same setting of initial genotypic frequencies as in Fig. 7. Simulation results (5000 independent runs) show that the average steady-state increment in $F_{st(B)}$ is 23.98% with $m_S = 0.04$ and $m_P = 0.08 (F_{st(B)} = 0.0729 \pm 0.0158, F_{st(B)} = 0.0588$ for the pure neutral case). The average relative increment is about 10.0% with $m_S = 0.01$ and $m_P = 0.08$ ($F_{st(B)} = 0.100 \pm 0.0205, F_{st(B)} = 0.0909$ for the pure neutral case). The average steady-state increment becomes 27.39% with $m_S = 0.04$ and $m_P = 0.02$ ($F_{st(B)} = 0.1158 \pm 0.0230, F_{st(B)} = 0.0909$ in the pure neutral case). These results indicate that effects of pollen dispersal are not as high as the effects of seed dispersal in changing $F_{st(B)}$. This is because pollen dispersal only affects the cytonuclear disequilibrium between nuclear and paternally inherited organelle alleles (D_{NC}) while seed dispersal simultaneously affects both types of cytonuclear disequilibria (D_{NC} and D_{ND}).

5.2. Random mating system

At the steady state, the allele frequencies at these two organelle sites in the lth subpopulation can be readily obtained by setting $\alpha = 0$ in Eqs. (8) and (9),

$$0 = \tilde{m}'\left(Q_D - p_{D(l)}\right) + s_{D(l)}p_{D(l)}(1 - p_{D(l)}), \tag{12}$$

$$0 = m_S\left(Q_C - p_{C(l)}\right) + s_{C(l)}p_{C(l)}(1 - p_{C(l)}). \tag{13}$$

It can be viewed that the two organelle sites evolve independently and their allele frequencies can be separately calculated from the above two equations.

Appendix D details the derivation of the steady-state population differentiation for the neutral nuclear allele,

$$F_{st(B)} = \frac{1}{1+4N_e\tilde{m}}\left(1+\rho_3\right), \tag{14}$$

where ρ_3 is the change due to cytonuclear disequilibrium. ρ_3 is a complex function of migration rates, selection coefficients, effective population size, and cytonuclear allele frequencies (Appendix D). Compared with the result under a pure neutral process, the relative increment in $F_{st(B)}$ is equal to $100\rho_3\%$.

Similarly, consider a specific case where equal selection coefficients are present among subpopulations, i.e. $s_{C(1)} = \cdots = s_{C(L)} = s_C$ and $s_{D(1)} = \cdots = s_{D(L)} = s_D$. The frequencies of selective organelle alleles are equal among subpopulations, i.e. $p_{C(1)} = \cdots = p_{C(L)} = p_C$ and $p_{D(1)} = \cdots = p_{D(L)} = p_D$. Appendix D derives the method for calculating $F_{st(B)}$ in this specific case.

Numerical results based on Eq. (14) show that seed dispersal facilitates the increment in $F_{st(B)}$ under certain conditions (when the migration rate is not greater than selection coefficients; Fig. 8a). This is because seed dispersal can simultaneously affect two types of cytonuclear disequilibria. The increment in $F_{st(B)}$ increases

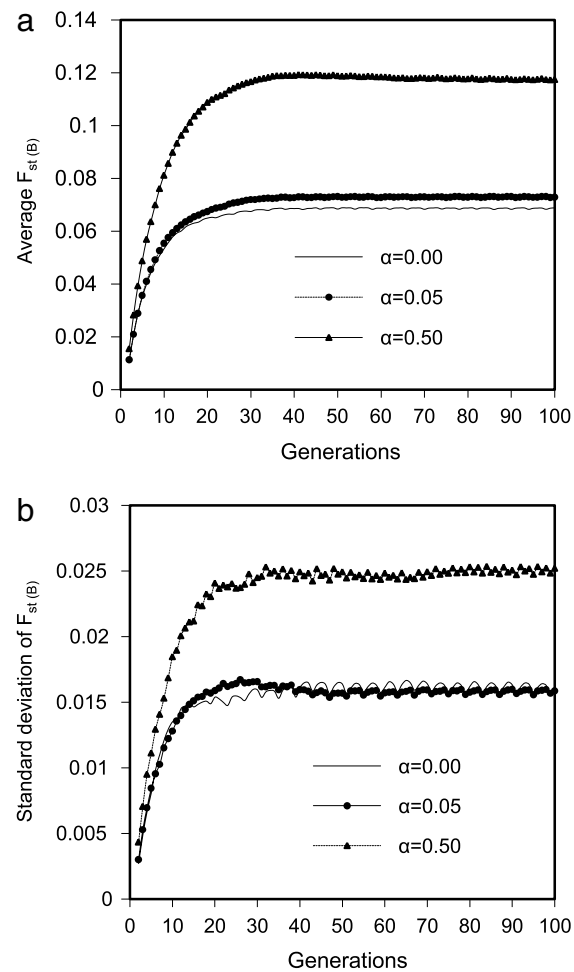

Fig. 7. Effects of selfing (α) on F_{st} for a neutral nuclear allele: (a) the change of average F_{st} with time; (b) the standard deviation of F_{st}. The results are obtained from 5000 independent simulations, with the migration rates of seeds $m_S = 0.04$ and pollen $m_P = 0.08$, the selection coefficients $s_B = s_C = 0.04$, the number of subpopulations $L = 30$, and the effective population size $N_e = 50$. The initial genotypic frequencies (the same as those in migrants) are set as $Q_{NNBC} = 0.44$, $Q_{NNBc} = 0.01$, $Q_{NNbC} = 0.01$, $Q_{NNbc} = 0.010$, $Q_{NnBC} = 0.015$, $Q_{NnBc} = 0.015$, $Q_{NnbC} = 0.015$, $Q_{Nnbc} = 0.015$, $Q_{nnBC} = 0.01$, $Q_{nnBc} = 0.01$, $Q_{nnbC} = 0.01$, and $Q_{nnbc} = 0.44$.

with the effective population size since ρ_3 is proportional to the effective population size. Unlike seed dispersal, pollen dispersal mainly changes the cytonuclear disequilibrium between nuclear and the organelle site with paternal inheritance. An increase in pollen dispersal can reduce $F_{st(B)}$ increment, given the constant setting of cytonuclear disequilibria in migrants (Fig. 8b). MC simulation results confirm these patterns showed in Fig. 8 about the roles of seed and pollen dispersal (results not shown here). This pattern can also be viewed from the MC simulations in the mixed mating system.

Numerical results based on Eq. (14) indicate that the selection strength at organelle sites facilitates $F_{st(B)}$ increment (Fig. 9), similar to the results in the preceding two cytonuclear systems. This pattern is consistent with MC simulation results. For example, let $m_S = 0.03$, $m_P = 0.06$, $L = 30$, $N_e = 50$, $Q_{AAB} = 0.45$, $Q_{AaB} = 0.03$, $Q_{aaB} = 0.02$, $Q_{AAb} = 0.02$, $Q_{Aab} = 0.03$, and $Q_{aab} = 0.45$, $Q_C = Q_D = 0.65$, $Q_N = 0.5$, $\bar{D}_{NNCD} = \bar{D}_{nncd} = 0.1$, $\bar{D}_{NnCD} = \bar{D}_{nnCD} = \bar{D}_{Nncd} = \bar{D}_{nncd} = -0.05$, and all other genotypic cytonuclear disequilibria $= 0.0$. Simulation (2000 independent runs) results show that the average relative increment in $F_{st(B)}$ at the steady state is 2.21% with selection coefficients $s_C = s_D = 0.02$ ($F_{st(B)} = 0.0786 \pm 0.0184$; $F_{st(B)} = 0.0769$ for the pure neutral

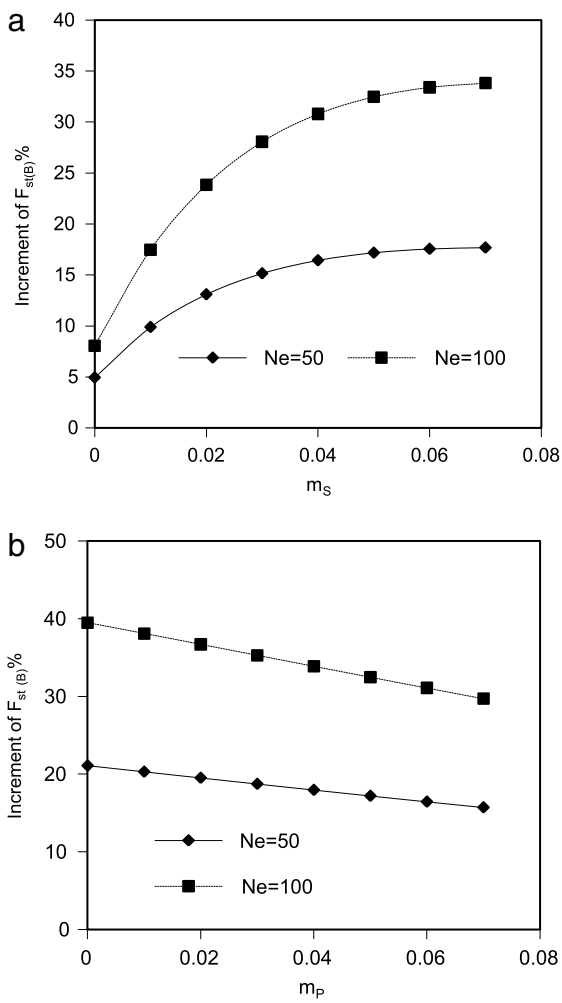

Fig. 8. Effects of seed and pollen dispersal on the increment in F_{st} for a neutral nuclear allele: (a) seed dispersal; (b) pollen dispersal. Calculations are based on Eq. (14) with the allele frequencies in migrants $Q_C = Q_D = Q_N = 0.5$, the cytonuclear disequilibrium $\bar{D}_{NC} = \bar{D}_{ND} = 0.10$, the selection coefficient $s_C = s_D = 0.08$, and the number of subpopulations $L = 30$. In (a), the migration rate of pollen is fixed at $m_P = 0.05$. In (b), the migration rate of seeds is fixed at $m_S = 0.05$.

case), 8.58% with $s_C = s_D = 0.04$ ($F_{st(B)} = 0.0835 \pm 0.0185$), and 11.31% with $s_C = s_D = 0.06$ ($F_{st(B)} = 0.0856 \pm 0.0167$). The relative increments predicted from Eq. (14) are 0.70%, 3.17%, and 6.51% for $s_C = s_D = 0.02$, 0.04, and 0.06, respectively, with the same parameters settings. Again, these predicted values are within the range of one standard deviation in each simulation case.

6. Discussion

The purpose of this study is to investigate the effects of transient cytonuclear disequilibrium on population differentiation of a neutral nuclear or organelle allele, complementary to the previous studies in the nuclear system for animal and plant species (Barton, 2000; Hu and He, 2005). This is also an alternative approach to looking at the spread of neutral alleles in space in the cytonuclear system (Hu, 2008). The results show that population differentiation for a neutral nuclear (or organelle) allele increases owing to the effects of selection on organelle (or nuclear) genomes via transient cytonuclear disequilibrium. This effect is often neglected in interpreting empirical F_{st}'s with cytonuclear markers. In a mixed mating system, selfing can substantially increase the transient increment in F_{st} for a neutral nuclear or organelle allele. An increment in F_{st} for a neutral organelle allele can also be substantial in

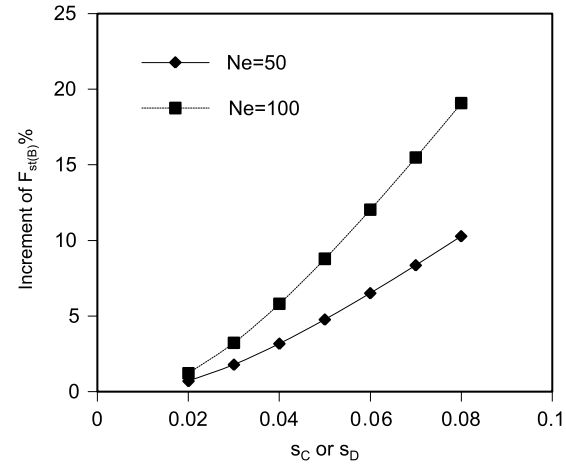

Fig. 9. Effects of selection on paternal and maternal organelle genomes on F_{st} for a neutral nuclear allele. Calculations are based on Eq. (14) with the allele frequencies in migrants $Q_C = Q_D = Q_N = 0.5$, the cytonuclear disequilibrium $\bar{D}_{NC} = \bar{D}_{ND} = 0.1$, the migration rates of seeds $m_S = 0.03$ and pollen $m_P = 0.06$, and the number of subpopulations $L = 30$.

the presence of cumulative selective nuclear sites. It is speculated that such an increment can be furthered when the joint effects of multiple selective nuclear sites and partial selfing occur although this is not examined here. In general, the transient increment in F_{st} for a neutral nuclear or organelle allele is not only related to the mode of its genetic inheritance but also to the vector of its dispersal (seeds and pollen). These theoretical insights highlight the significance of transient cytonuclear disequilibrium in changing F_{st} for neutral alleles. The results also expand our understanding of empirical studies when cytonuclear markers are used to investigate population structure.

Note that the transient cytonuclear disequilibrium is maintained by the joint effects of migration, selection, and genetic drift. Relative effects of each component may be different and vary with structured populations, which can be deduced from the expectations of cytonuclear disequilibria explored here. Although cytonuclear disequilibrium has been stressed for inferring different evolutionary and ecological processes (Asmussen and Schnabel, 1991; Schnabel and Asmussen, 1992; Harrison, 1989; Ennos et al., 1999; Orive and Barton, 2002), its impacts on population differentiation have only been investigated in a limited way (Datta et al., 1996; Datta and Arnold, 1998). The transient increment in F_{st} is important for a subdivided population with a short history, such as those recently expanded populations owing to climate changes where transient cytonuclear disequilibrium exists (Latta et al., 2001). However, the eventual consequence for F_{st} of a neutral allele remains the same as those obtained for individual genomes since the cytonuclear disequilibrium decays with time.

It is important to distinguish the presently addressed cytonuclear systems from two alternative cytonuclear systems. One is the system between neutral nuclear and organelle alleles where cytonuclear disequilibrium is generated by migration and/or genetic drift. This type of cytonuclear disequilibrium can be stable when it is present in immigrants (Asmussen and Schnabel, 1991; Asmussen and Arnold, 1991; Schnabel and Asmussen, 1992; Dakin, 2006), or transient when it is generated by genetic drift (Fu and Arnold 1992; Datta and Arnold 1998; Latta et al. 2001). One distinct feature is that the cytonuclear disequilibrium between neutral sites does not change their F_{st}'s and the two neutral sites evolve independently. This provides the genetic basis for independently using F_{st}'s of nuclear and organelle genomes to estimate the ratio of pollen to seed dispersal (Petit et al., 1993; Ennos, 1994; Hu and Ennos, 1997, 1999), or to infer the asymmetric introgression between cytonuclear genomes in hybrid zones (Arnold, 1993; Avise, 1994; Ennos et al., 1999).

The other system is the one between selective cytonuclear sites where a stable cytonuclear disequilibrium can be maintained by the joint effects of selection and migration in structured populations (Hu and Li, 2002; Monsen et al., 2007), or by the cytonuclear epistatic effects on fitness (Asmussen et al., 1987). This stable cytonuclear disequilibrium can also be implied from the results in a strictly nuclear system where stable linkage disequilibrium (LD) between selective sites can be maintained in structured populations (Li and Nei, 1974; Slatkin, 1975). As a consequence, their F_{st}'s are interdependent although F_{st} remains to be formulated in either the nuclear or cytonuclear system (Wright, 1969), owing to various patterns of spatial selection.

Similar to the cytonuclear cline situation (Hu, 2008), a comparison between the nuclear and cytonuclear systems is related to an array of combinations of different conditions (Barton, 2000; Hu and He, 2005). It cannot be excluded that cytonuclear disequilibrium may be stronger than LD between linked selective and neutral nuclear sites under some combinations of the number of selective sites, the migration rate of seeds and pollen, the magnitude of cytonuclear disequilibrium in migrants, and the level of selection strengths. Under this situation, cytonuclear disequilibrium can have larger effects than LD in the nuclear system on shaping F_{st} for a neutral allele. However, for a pair of cytonuclear sites (one neutral and the other selective) in a more general condition, cytonuclear disequilibrium could not be as large as the LD between linked neutral and selective nuclear sites because the "recombination rate" between cytonuclear sites is 1/2, the maximum recombination rate. Cytonuclear disequilibrium dissipates faster with time than does the LD between linked selective and neutral nuclear sites. Thus, a frequent outcome is that the magnitude (or the persistent time) for the transient increment of F_{st} for a neutral allele is greater (or longer) in the nuclear system than in the cytonuclear system. Empirical comparisons await appropriate data collections.

Unlike the effects of seed dispersal, effects of selection on the transient increment in F_{st} are realized by changing cytonuclear disequilibrium via altering selective allele frequencies in each subpopulation. High selection strength increases the frequency of adaptive alleles and facilitates cytonuclear disequilibrium. The general case is that selection coefficient and degree of dominance vary in space, displaying various adaptive alleles and population fitness. The expectations of allele frequencies, cytonuclear disequilibrium, and their second-order functions, such as $E(D_{A_iB})$, $E(p_B)$, $E(p_B^2)$, $E(p_B D_{AB})$, and $E(D_{AB}^2)$ in the first cytonuclear system examined in this study, can vary in space as well. Under this situation, the transient increment in F_{st} may be different in magnitude from the results obtained under equal spatial selection strengths. Nevertheless, the qualitative results remain the same.

Concerning the transient increment in F_{st} for a neutral nuclear allele, the present study only addresses a simple cytonuclear system. One more complex system is that a neutral nuclear allele is affected by the selective sites from both organelle and nuclear genomes. Inclusion of the effects from linked selective nuclear sites can reinforce the transient increment in F_{st} (Barton, 2000; Hu and He, 2005). A challenge is to assess the relative contributions from nuclear and organelle genomes under a variety of conditions, which requires a more sophisticated analysis.

Similarly, concerning the transient increment in F_{st} for a neutral organelle allele, the selective sites on the same organelle genomes can also bring about selective sweep effects (Maruyama and Birky, 1991; Ennos et al., 1999). Selective sweep effects can be strong and rapidly spread through whole population (Mazur and Falco, 1989; Hanson, 1991), producing the same genetic background across all subpopulations since recombination rarely occurs between organelle genomes. Under this condition, it is important to understand that the present results refer to the additional effects from cumulative selective nuclear sites.

Concerning individual fitness, this study only addresses a simple cytonuclear system where cytonuclear epistasis is excluded. Evidence is recorded about the interdependence in function and some DNA sequence homology between cytonuclear genomes (Rand et al., 2004; Wolf, 2009). Cytonuclear disequilibrium can also be generated by epistatic effects in addition to the mechanisms addressed here (Asmussen et al., 1987). It is of interest for future study to address the epistatic effects on transient increment in F_{st} for a neutral nuclear or organelle allele.

The results here have several implications for the use of cytonuclear markers. The first implication is that a caution is required in interpreting population differentiation observed with molecular markers, especially in the presence of disequilibria between nuclear and organelle markers. The precondition for such analyses is the neutrality test when molecular markers from nuclear or organelle genomes are separately employed for F_{st} analysis (Kimura, 1983; Li, 1997), and/or the cytonuclear disequilibria test when cytonuclear markers are simultaneously used (Basten and Asmussen, 1997; Castro et al., 1999). Use of those neutral markers without cytonuclear disequilibria can effectively remove the transient impacts on population structure analysis.

The second implication is related to the comparison of using F_{st} to estimate seed and pollen dispersal with using cytonuclear disequilibrium, which has been discussed by Ennos (1994). Cytonuclear disequilibrium is independent of F_{st} when neutral nuclear and organelle markers are employed, implying that both of them can be separately used to infer migration parameters in structured populations (Asmussen and Schnabel, 1991; Schnabel and Asmussen, 1992; Ennos, 1994). This independent property also implies their similarity for inferring population genetic structure under a strict neutral condition. The present study furthers the comparison by formulating the relationships between cytonuclear disequilibrium and F_{st} in a different model. Both F_{st} and cytonuclear disequilibrium are the function of selection, migration, and genetic drift. This also implies an additional condition for $F_{st} \neq 1/(1 + 4N_e m)$ for a neutral nuclear allele when cytonuclear system is considered (Whitlock and McCauley, 1999). When the cytonuclear disequilibrium exists between selective nuclear and neutral organelle sites or the reverse case, effects of seed and pollen dispersal are compounded with those of selection and genetic drift. Under this situation, additional data are required for using F_{st} or cytonuclear disequilibrium to estimate seed and pollen dispersal. Thus, clarification of cytonuclear disequilibrium for mechanism is required in prior for further explorations.

The third implication is the prediction of strong effects on F_{st} from nonrandom mating in the cytonuclear system. It is commonly held that selfing facilitates population differentiation in addition to causing inbreeding depression in many plant species in the nuclear system (Wright, 1969; Mitton, 1992). The importance of nonrandom mating in generating cytonuclear disequilibrium has also been stressed (Asmussen et al., 1987; Maroof et al., 1992). The present simulations further imply that the joint effects of selection and partially selfing can substantially increase the transient increment of F_{st}. This may enhance the persistence of transient increment over times and allow more opportunity to detect high F_{st}. This also increases misinterpretation of the transient increment in F_{st} as evidence for evolutionary forces other than the effects of the mixed mating system (Overath and Asmussen, 2000).

The fourth implication is about the roles of seed and pollen dispersal in shaping the transient increment of F_{st} in the cytonuclear system. Early studies indicate that effectiveness of seed and pollen dispersal as vectors of gene flow in changing F_{st} is related to the mode of inheritance in the nuclear or organelle genome system (Petit et al., 1993; Ennos, 1994; Hu and Ennos, 1997, 1999). This is also the case in spreading a selective or neutral allele in cytonuclear clines in hybrid zones (Hu and Li, 2002; Hu, 2008). The present theory furthers the function of seed and pollen dispersal in changing

the transient increment of F_{st} in addition to shaping the steady-state F_{st}. Seed dispersal can generate two types of cytonuclear disequilibria and facilitate the transient increment of F_{st} for a neutral allele from either nuclear or organelle genomes. Pollen dispersal only directly changes the cytonuclear disequilibrium between nuclear and paternal organelle sites. Such a divergence predicts that plant species with various ratios of pollen to seed dispersal may have unequal transient increments in F_{st}.

Finally, it is of interest to mention briefly the implication of the present theory to genome-wide F_{st} mapping at a fine scale. The present theory implies that transient increment of F_{st} at a neutral nuclear site could be misinterpreted as a F_{st} outlier (false positive), especially when the subdivided population has a short history. This problem is likely not serious when the transient increment mainly comes from effects of seed and pollen dispersal or genetic drift because the whole genome-wide F_{st}'s will be simultaneously altered. However, when the transient increment comes from selection at organelle genomes, together with genetic hitchhiking effects from nuclear genomes, regional variation among F_{st}'s at different neutral sites exists. These compounded sources of perturbations complicate our insights into genomic evolution at the population level.

Acknowledgments

I sincerely appreciate Nicholas H. Barton, Marcus W. Feldman, and two reviewers for helpful comments that substantially improved earlier versions of this article.

Appendix A

First, consider two selective diallelic nuclear sites (i, j) in one subpopulation, with alleles A_i and a_i at the i site and A_j and a_j at the j site, and one neutral organelle site, with alleles B and b. Let D_{ij} be the LD between the i and j sites at the current adult stage and r_{ij} be the recombination rate between them. The frequencies of four nuclear gametes produced by the adults can be written as follows: $p_{A_iA_j} = p_{A_i}p_{A_j} + (1 - r_{ij})D_{ij}, \ldots,$ and $p_{a_ia_j} = p_{a_i}p_{a_j} + (1-r_{ij})D_{ij}$. The cytonuclear gametic disequilibria are set as $p_{A_iA_jB} = p_{A_iA_j}p_B + D_{A_iA_jB}$, $p_{A_iA_jb} = p_{A_iA_j}p_b + D_{A_iA_jb}$, and similar expressions for the remaining six gemates, which extends the definitions of Asmussen et al. (1987) to the three-site case. Some properties for the three-site cytonuclear disequilibria are readily derived, such as $D_{A_iA_jB} + D_{A_iA_jb} = 0$ and $D_{A_iB} = D_{A_iA_jB} + D_{A_ia_jB}$.

Let $1 + s_i$, $1 + h_is_i$, and $1 - s_i$ be the fitness for A_iA_i, A_ia_i, and a_ia_i, respectively, and $1 + s_j$, $1 + h_js_j$, and $1 - s_j$ for A_jA_j, A_ja_j, and a_ja_j, respectively. Fitness for any two-site genotype can be readily calculated according to the assumptions of multiplicative-viability model with weak selection. Let $\tilde{m} = m_S + m_P/2$ and $\tilde{m}' = m_S + m_P$. Following the life cycle of hermaphroditic plants, the allele frequency at the i site in the next adult is derived as

$$p_{A_i}^{***} = (1 - \tilde{m})p_{A_i} + \tilde{m}Q_{A_i} + s_ip_{A_i}p_{a_i}\left(1 + (p_{a_i} - p_{A_i})h_i\right) + s_j(1 - r_{ij})\left(1 + (p_{a_j} - p_{A_j})h_j\right)D_{ij}, \quad (A.1)$$

where Q_{A_i} is the frequency of A_i in migrants. The allele frequency at the j site in the next adults, $p_{A_j}^{***}$, can be simply obtained by replacing subscript i in Eq. (A.1) with subscript j.

The recursion equation for LD between selective nuclear sites is derived as

$$\begin{aligned} D_{ij}^{**} = {} & (1 - r_{ij})D_{ij} + \tilde{m}(1 - r_{ij})(\bar{D}_{ij} - D_{ij}) \\ & + \tilde{m}(p_{A_i} - Q_{A_i})(p_{A_j} - Q_{A_j}) + (1 - r_{ij}) \\ & \times D_{ij}\left(s_i(p_{a_i} - p_{A_i})(1 + (p_{a_i} - p_{A_i})h_i)\right. \\ & \left. + \ s_j(p_{a_j} - p_{A_j})(1 + (p_{a_j} - p_{A_j})h_j)\right), \end{aligned} \quad (A.2)$$

where $\bar{D}_{ij}$ is the LD between sites i and j in migrants.

The allele frequency at the neutral organelle site in the next adults is derived as

$$p_B^{***} = (1 - \tilde{m}')p_B + \tilde{m}'Q_B + s_i\left(1 + (p_{a_i} - p_{A_i})h_i\right)D_{A_iB} + s_j\left(1 + (p_{a_j} - p_{A_j})h_j\right)D_{A_jB}, \quad (A.3)$$

where Q_B is the frequency of B in migrants. The recursion equation for cytonuclear disequilibrium is derived as

$$\begin{aligned} D_{A_iB}^{***} = {} & \frac{1}{2}D_{A_iB} + \tilde{m}(p_B - Q_B)(p_{A_i} - Q_{A_i}) + \frac{1}{2}\tilde{m}'(\bar{D}_{A_iB} - D_{A_iB}) \\ & + \frac{1}{2}s_i\left((p_{a_i} - p_{A_i})(1 + (p_{a_i} - p_{A_i})h_i) - 2p_{a_i}p_{A_i}h_i\right)D_{A_iB}, \end{aligned} \quad (A.4)$$

where $\bar{D}_{A_iB}$ is the cytonuclear disequilibrium in migrants.

Now, consider the case of multiple nuclear sites. From Eqs. (A.1) and (A.2), the results for each of the T selective nuclear sites can be expressed as

$$p_{A_i}^{***} = (1 - \tilde{m})p_{A_i} + \tilde{m}Q_{A_i} + s_ip_{A_i}p_{a_i}\left(1 + (p_{a_i} - p_{A_i})h_i\right) + \sum_{j\neq i}^{T} s_j(1 - r_{ij})\left(1 + (p_{a_j} - p_{A_j})h_j\right)D_{ij}, \quad (A.5)$$

The last term on the right-hand side of Eq. (A.5) represents the cumulative effects from the $T - 1$ selective sites. Expression for the change in LD between selective nuclear sites remains the same as Eq. (A.2).

According to the assumptions, selective sites are subject to the balance of selection-migration (Ohta and Kimura, 1970; Nordborg et al., 1996; Hu and He, 2005). At the steady state, solutions are difficult to obtain from $T + T(T - 1)/2$ joint equations derived from Eqs. (A.2) and (A.5). Without loss of examining the cumulative effects from multiple selective nuclear sites on the neutral organelle site, a coincidence among T allele frequencies is examined.

Consider equal selection coefficients ($s_1 = s_2 = \cdots = s_T = s$, and $h_1 = h_2 = \cdots = h_T = h$) and equal allele frequencies among the T selective sites ($p_{A_1} = p_{A_2} = \cdots = p_{A_T} = p_A$, and $p_a = 1 - p_A$). From Eq. (A.2), we obtained the steady-state LD in one subpopulation

$$D_{ij} = \frac{\tilde{m}\left((1 - r_{ij})\bar{D}_{ij} + (p_A - Q_A)^2\right)}{r_{ij} + (1 - r_{ij})\left(\tilde{m} - 2s(p_a - p_A)(1 + (p_a - p_A)h)\right)}. \quad (A.6)$$

Substituting Eq. (A.6) into Eq. (A.5) and using the average of T equations yield the following steady-state equation for selective nuclear alleles

$$p_A = Q_A + s(1 + (p_a - p_A)h_{(l)}) \times \left(\frac{p_Ap_a}{\tilde{m}} + 2(R_1\bar{D}_{ij} + R_2(p_A - Q_A)^2)\right), \quad (A.7)$$

where $R_1 = \frac{1}{L}\sum_i\sum_{j<i}\frac{(1-r_{ij})^2}{r_{ij}+(1-r_{ij})(\tilde{m}-2s(p_a-p_A)(1+(p_a-p_A)h))}$, and $R_2 = \frac{1}{L}\sum_i\sum_{j<i}\frac{1-r_{ij}}{r_{ij}+(1-r_{ij})(\tilde{m}-2s(p_a-p_A)(1+(p_a-p_A)h))}$. From Eq. (A.7), the steady-state allele frequency p_A $(0 \leq p_A \leq 1)$ can be numerically solved using the iterative method.

Appendix B

This appendix derives the analytical expression for $F_{st(P)}$ and the method for calculating the transient increment in $F_{st(P)}$. After genetic drift, the neutral allele frequency, denoted by $p_{B(l)}^{****}$ in the lth subpopulation, can be written as $p_{B(l)}^{****} = p_{B(l)}^{***} + \delta_{(l)}$ where $\delta_{(l)}$ is the random change due to sampling and $p_{B(l)}^{***}$ is the neutral allele frequency after the event of natural selection acting on T nuclear sites. $\delta_{(l)}$ has the mean zero and variance $p_{B(l)}^{***}(1 - p_{B(l)}^{***})/N_e$. Let $\sigma_B'^2$

be the variance of allele frequency among L subpopulations after genetic drift. Then, I obtained

$$\sigma_B'^2 = \frac{1}{L}\sum_{l=1}^{L}\left(p_{B(l)}^{****} - Q_B\right)^2 = \left(1-\frac{1}{N_e}\right)\left((1-\tilde{m}')^2\sigma_B^2 + \bar{\Pi}_1 + \bar{\Lambda}_1^2\right) + \frac{1}{N_e}Q_B(1-Q_B), \tag{B.1}$$

where $\sigma_B^2 = \sum_{l=1}^{L}\left(p_{B(l)} - Q_B\right)^2/L$, $\bar{\Pi}_1 = 2(1-\tilde{m}')\sum_{l=1}^{L}\left((p_{B(l)} - Q_B)\Lambda_{1(l)}\right)/L$, and $\bar{\Lambda}_1^2 = \sum_{l=1}^{L}\Lambda_{1(l)}^2/L$.

The expression for $\Lambda_{1(l)}$ is given in the main text, Eq. (3). Note that the above derivation is obtained using the same method as that used by Hu and Ennos (1999). At the steady state, population differentiation for the neutral organelle allele is derived as

$$F_{st(P)} = \frac{1}{1+2N_e\tilde{m}'}\left(1+\rho_1\right), \tag{B.2}$$

where $\rho_1 = (N_e-1)(\bar{\Pi}_1 + \bar{\Lambda}_1^2)/Q_B(1-Q_B)$.

Without loss of examining the cumulative effects of multiple selective nuclear sites, consider a specific case where equal selection coefficients exist among subpopulations and the coincidence of allele frequencies among the T sites takes place in each subpopulation. Let $s_{1(l)} = s_{2(l)} = \cdots = s_{T(l)} = s$, $h_{1(l)} = h_{2(l)} = \cdots = h_{T(l)} = h$, $p_{A_{1(l)}} = p_{A_{2(l)}} = \cdots = p_{A_{T(l)}} = p_A$, and $p_a = 1-p_A$. Then, $\bar{\Pi}_1$ becomes $\bar{\Pi}_1 = 2(1-\tilde{m}')Ts(1+(p_a-p_A)h)(E(p_{B(l)}D_{AB(l)}) - Q_BE(D_{AB(l)}))$, and $\bar{\Lambda}_1^2$ becomes $\bar{\Lambda}_1^2 = T^2s^2(1+(p_a-p_A)h)^2E(D_{AB(l)}^2)$. Note that the expectation E is based on the density distribution of two independent random variables p_B and D_{AB}, i.e. $\phi(p_B, D_{AB})$, the same consideration as Ohta and Kimura (1969b, 1970). Expectation for a single variable can be further simplified with marginal density, e.g. $E(p_B) = \iint p_B\phi(p_B, D_{AB})\mathrm{d}p_B\mathrm{d}D_{AB} = \int p_B\phi(p_B)\mathrm{d}p_B$ and $E(D_{AB(l)}) = \int\int D_{AB(l)}\phi(p_B, D_{AB})\mathrm{d}p_B\mathrm{d}D_{AB} = \int D_{AB(l)}\phi(D_{AB})\mathrm{d}D_{AB}$.

In the next, the Kolmogorov backward equation is used to calculate the following expectations: $E(p_B)$, $E(D_{A_iB})$, $E(p_B^2)$, $E(p_BD_{AB})$, and $E(D_{AB}^2)$. This approach was used by Ohta and Kimura (1970) in addressing associative overdominance effects. Under the coincidence of allele frequencies among the T nuclear sites, individual cytonuclear disequilibria, D_{AB}'s, are equal. According to the results of Ohta and Kimura (1969b, eq. A4, p.238), for a function of two random variables (p_B and D_{AB}) with a stationary distribution, f, the following equation can be derived as

$$\begin{aligned} 0 = E\Bigg\{&\frac{1}{2}\left(\frac{1}{N_e}\left(\frac{1}{2}p_Ap_ap_Bp_b + \frac{1}{4}(1-2p_A)(1-2p_B)D_{AB} - \frac{1}{4}D_{AB}^2\right)\right)\frac{\partial^2 f}{\partial D_{AB}^2} + \frac{1}{2N_e}p_B(1-p_B)\frac{\partial^2 f}{\partial p_B^2} \\ &+ \frac{1}{2N_e}(1-2p_B)D_{AB}\frac{\partial^2 f}{\partial p_B\partial D_{AB}} \\ &+ \left(\tilde{m}'(Q_B-p_B) + Ts(1+(p_a-p_A)h)D_{AB}\right)\frac{\partial f}{\partial p_B} \\ &+ \left(-\frac{1}{2}D_{AB} + \tilde{m}(p_A-Q_A)(p_B-Q_B) + \frac{\tilde{m}'}{2}(\bar{D}_{AB}-D_{AB}) + \frac{1}{2}\Delta D_{AB}\right)\frac{\partial f}{\partial D_{AB}}\Bigg\}, \end{aligned} \tag{B.3}$$

where $\Delta = s\left((p_a-p_A)(1+(p_a-p_A)h) - 2p_ap_Ah\right)$. The deduction of the above equation is based on the life cycle of hermaphroditic plants and the sampling variance for D_{AB} is slightly different from the results without inclusion of the process of pollen and ovule combinations. In deriving Eq. (B.3), the terms with the order of $\tilde{m}'^2, s^2, s\tilde{m}', \tilde{m}'/N_e, s/N_e$, or higher are neglected.

$E(D_{AB})$ and $E(p_B)$ can be calculated by substituting $f = D_{AB}$ and $f = p_B$ in Eq. (B.3). Substitution of f in Eq. (B.3) by D_{AB}^2, p_BD_{AB}, and p_B^2 respectively yields the following three equations for $E(D_{AB}^2)$, $E(p_BD_{AB})$, and $E(p_B^2)$:

$$\begin{pmatrix} a_0 & b_0 & c_0 \\ a_1 & b_1 & c_1 \\ 0 & b_2 & c_2 \end{pmatrix}\begin{pmatrix} E(D_{AB}^2) \\ E(p_BD_{AB}) \\ E(p_B^2) \end{pmatrix} = \begin{pmatrix} g_0 \\ g_1 \\ g_2 \end{pmatrix}, \tag{B.4}$$

where $a_0 = -\frac{1}{4} - N_e(1+\tilde{m}'-\Delta), \ldots,$ and $g_2 = -(1+N_e\tilde{m}'Q_B)E(p_B)$. Eq. (B.4) can be numerically solved using the *Mathematica* tool (Wolfram, 1991). Note that in the absence of cytonuclear disequilibrium in migrants ($\bar{D}_{AB} = 0$), its expectation is absent as well, i.e. $E(D_{AB}) = 0$ from Eq. (B.3), but $E(D_{AB}^2)$ and $E(p_BD_{AB})$ remain present due to the genetic drift effects (Ohta and Kimura, 1970). This can be seen from the solution to Eq. (B.4) in the absence of cytonuclear disequilibrium in migrants. The property is not mentioned in the continent-island model investigated by Asmussen and Schnabel (1991) and Schnabel and Asmussen (1992).

Appendix C

Consider the case of one neutral organelle and T selective nuclear sites in one subpopulation. The results for the T selective nuclear sites are the same as Eqs. (A.2) and (A.5) in Appendix A. Their steady-state values can be calculated using the method described in Appendix A.

The neutral allele frequency before the occurrence of genetic drift in the lth subpopulation can be expressed as Eq. (6) in the main text. The recursion equation for individual pairs of cytonuclear disequilibria in the lth subpopulation, $D_{A_iE(l)}$, can be expressed as

$$\begin{aligned} D_{A_iE(l)}^{**} = &\frac{1}{2}D_{A_iE(l)} + m_S(p_{E(l)} - Q_E)(p_{A_i(l)} - Q_{A_i}) \\ &+ \frac{1}{2}m_S(\bar{D}_{A_iE} - D_{A_iE(l)}) + \frac{1}{2}s_{i(l)}\left((p_{a_i(l)} - p_{A_i(l)}) \times (1+(p_{a_i(l)} - p_{A_i(l)})h_{i(l)}) - 2p_{a_i(l)}p_{A_i(l)}h_{i(l)}\right)D_{A_iE(l)}, \end{aligned} \tag{C.1}$$

where $\bar{D}_{A_iE}$ is the cytonuclear disequilibrium in migrants.

After genetic drift, the variance of allele frequency among L subpopulations, $\sigma_E'^2$, is derived as

$$\sigma_E'^2 = \left(1-\frac{1}{N_e}\right)\left((1-\tilde{m}')^2\sigma_E^2 + \bar{\Pi}_2 + \bar{\Lambda}_2^2\right) + \frac{1}{N_e}Q_E(1-Q_E), \tag{C.2}$$

where $\sigma_E^2 = \sum_{l=1}^{L}\left(p_{E(l)} - Q_E\right)^2/L$, $\bar{\Pi}_2 = 2(1-m_S)\sum_{l=1}^{L}\left((p_{E(l)} - Q_E)\Lambda_{2(l)}\right)/L$, and $\bar{\Lambda}_2^2 = \sum_{l=1}^{L}\Lambda_{2(l)}^2/L$. At the steady state, population differentiation for the neutral organelle allele with maternal inheritance is

$$F_{st(M)} = \frac{1}{1+2N_em_S}\left(1+\rho_2\right), \tag{C.3}$$

where $\rho_2 = (N_e-1)(\bar{\Pi}_2 + \bar{\Lambda}_2^2)/(Q_E(1-Q_E))$.

To calculate ρ_2, consider a specific situation where equal selection coefficients exist among subpopulations and the coincidence of allele frequencies among the T sites takes place in each subpopulation. Then, $\bar{\Pi}_2$ becomes $2(1-m_S)Ts(1+(p_a-p_A)h)(E(p_{E(l)}D_{AE(l)}) - Q_EE(D_{AE(l)}))$, and $\bar{\Lambda}_2^2$ becomes $T^2s^2(1+(p_a-p_A)h)^2E(D_{AE(l)}^2)$. The expressions for $E(D_{A_iE})$, $E(p_E)$, $E(p_E^2)$, $E(p_ED_{AE})$, and $E(D_{AE}^2)$ under the influences of genetic drift can be calculated by replacing $\tilde{m}$, $\tilde{m}'$, p_B and D_{AB} in Eq. (B.3) in Appendix B with m_S, m_S, p_E, and D_{AE}, respectively.

Appendix D

This appendix begins by deriving the expression of $F_{st(B)}$ and then the method for calculating $F_{st(B)}$ in the three-genome system. The neutral allele frequency before the occurrence of genetic drift can be obtained by setting $\alpha = 0$ in Eq. (11) in the main text. The gametic cytonuclear disequilibria after selection in the random mating system are derived as

$$\begin{aligned} D^*_{ND(l)} &= \frac{1}{2}D_{ND(l)} + \frac{\tilde{m}'}{2}\left(\bar{D}_{ND} - D_{ND(l)}\right) \\ &\quad + \tilde{m}\left(p_{N(l)} - Q_N\right)\left(p_{D(l)} - Q_D\right) \\ &\quad + \frac{1}{2}s_{D(l)}\left(p_{d(l)} - p_{D(l)}\right)D_{ND(l)}, \end{aligned} \tag{D.1}$$

$$\begin{aligned} D^*_{NC(l)} &= \frac{1}{2}D_{ND(l)} + \frac{m_S}{2}\left(\bar{D}_{NC} - D_{NC(l)}\right) \\ &\quad + m_S\left(p_{N(l)} - Q_N\right)\left(p_{C(l)} - Q_C\right) \\ &\quad + \frac{1}{2}s_{C(l)}\left(p_{c(l)} - p_{C(l)}\right)D_{NC(l)}, \end{aligned} \tag{D.2}$$

where $\bar{D}_{NC}$ and $\bar{D}_{ND}$ are the cytonuclear disequilibria in migrants. Eqs. (D.1) and (D.2) indicate that seed and pollen dispersal differently contributes to cytonuclear disequilibria between the nuclear and paternal organelle alleles (D_{ND}) and between the nuclear and maternal organelle alleles (D_{NC}).

The variance of allele frequency among subpopulations after genetic drift, ${\sigma'_N}^2$, is derived as

$$\begin{aligned} {\sigma'_N}^2 &= \left(1 - \frac{1}{2N_e}\right)\left((1-m_S)^2\sigma_N^2 + \bar{\Pi}_3 + \bar{\Lambda}_3^2\right) \\ &\quad + \frac{1}{2N_e}Q_N(1-Q_N), \end{aligned} \tag{D.3}$$

where $\sigma_N^2 = \sum_{l=1}^{L}\left(p_{N(l)} - Q_N\right)^2/L$, $\bar{\Pi}_3 = (1-\tilde{m})\sum_{l=1}^{L}\left((p_{N(l)} - Q_N)(s_D D_{ND(l)} + s_{C(l)}D_{NC(l)})\right)/L$, and $\bar{\Lambda}_3^2 = \sum_{l=1}^{L}\left(\frac{1}{4}s_{D(l)}^2 D_{ND(l)}^2 + \frac{1}{2}s_{D(l)}s_{C(l)}D_{ND(l)}D_{NC(l)} + \frac{1}{4}s_{C(l)}^2 D_{NC(l)}^2\right)/L$. Note that $\bar{\Lambda}_3^2$ could be very small and has a small contribution to ${\sigma'_N}^2$. At the steady-state (${\sigma'_N}^2 = \sigma_N^2$), the population differentiation for the neutral nuclear allele is derived as

$$F_{st(B)} = \frac{1}{1 + 4N_e\tilde{m}}\left(1 + \rho_3\right), \tag{D.4}$$

where $\rho_3 = (2N_e - 1)(\bar{\Pi}_3 + \bar{\Lambda}_3^2)/(Q_N(1-Q_N))$.

To calculate ρ_3, consider a specific case where equal selection coefficients are present among subpopulations, i.e. $s_{C(1)} = \cdots = s_{C(L)} = s_C$ and $s_{D(1)} = \cdots = s_{D(L)} = s_D$. The frequencies of selective organelle alleles are equal among subpopulations, i.e. $p_{C(1)} = \cdots = p_{C(L)} = p_C$ and $p_{D(1)} = \cdots = p_{D(L)} = p_D$. Then, $\bar{\Pi}_3$ is expressed as $(1-\tilde{m})(s_C(E(p_N D_{NC}) - Q_N E(D_{NC})) + s_D(E(p_N D_{ND}) - Q_N E(D_{ND})))$, and $\bar{\Lambda}_3^2$ becomes $\frac{1}{4}s_D^2 E(D_{ND}^2) + \frac{1}{2}s_D s_C E(D_{ND}D_{NC}) + \frac{1}{4}s_C^2 E(D_{NC}^2)$. Under this case, each of these expectations ($E(p_N)$, $E(D_{ND})$, ..., and $E(D_{ND}D_{NC})$) is equal among subpopulations. The expectation E is based on the density distribution of three independent random variables p_N, D_{NC}, and D_{ND}, i.e. $\phi(p_N, D_{NC}, D_{ND})$.

The Kolmogorov backward equation is used to calculate the following expectations after genetic drift in one subpopulation: $E(p_N)$, $E(D_{ND})$, $E(D_{NC})$, $E(p_N^2)$, $E(p_N D_{NC})$, $E(p_N D_{ND})$, $E(D_{ND}^2)$, $E(D_{NC}^2)$, and $E(D_{ND}D_{NC})$. According to the results of Ohta and Kimura (1969a,b), for a function of random variables (p_N, D_{NC}, and D_{ND}) with stationary distribution, f, the following equation can be derived as

$$\begin{aligned} 0 = E\Bigg\{ & \frac{1}{4N_e}p_N p_n \frac{\partial^2 f}{\partial p_N^2} + \frac{1}{2}\left(\frac{1}{N_e}\left(\frac{1}{2}p_N p_n p_D p_d\right.\right. \\ & \left.\left. + \frac{1}{4}(1-2p_N)(1-2p_D)D_{ND} - \frac{1}{4}D_{ND}^2\right)\right)\frac{\partial^2 f}{\partial D_{ND}^2} \\ & + \frac{1}{2}\left(\frac{1}{N_e}\left(\frac{1}{2}p_N p_n p_C p_c + \frac{1}{4}(1-2p_N)(1-2p_C)D_{NC}\right.\right. \\ & \left.\left. - \frac{1}{4}D_{NC}^2\right)\right)\frac{\partial^2 f}{\partial D_{NC}^2} + \frac{1}{4N_e}(1-2p_N)D_{ND}\frac{\partial^2 f}{\partial p_N \partial D_{ND}} \\ & + \frac{1}{4N_e}(1-2p_N)D_{NC}\frac{\partial^2 f}{\partial p_N \partial D_{NC}} + \frac{1}{4N_e}D_{ND}D_{NC}\frac{\partial^2 f}{\partial D_{ND}\partial D_{NC}} \\ & + \left(\tilde{m}(Q_N - p_N) + \frac{1}{2}s_D D_{ND} + \frac{1}{2}s_C D_{NC}\right)\frac{\partial f}{\partial p_N} \\ & + \left(-\frac{1}{2}D_{NC} + m_S(p_C - Q_C)(p_N - Q_N) + \frac{m_S}{2}(\bar{D}_{NC} - D_{NC})\right. \\ & \left. + \frac{1}{2}s_C(p_c - p_C)D_{NC}\right)\frac{\partial f}{\partial D_{NC}} \\ & + \left(-\frac{1}{2}D_{ND} + \tilde{m}(p_D - Q_D)(p_N - Q_N) + \frac{\tilde{m}'}{2}(\bar{D}_{ND} - D_{ND})\right. \\ & \left. + \frac{1}{2}s_D(p_d - p_D)D_{ND}\right)\frac{\partial f}{\partial D_{ND}}\Bigg\}. \end{aligned} \tag{D.5}$$

In deriving the above equation, all the terms with the order of $\tilde{m}'^2$, s^2, $s\tilde{m}'$, $\tilde{m}'/N_e$, s/N_e, or higher are neglected.

Substitution of f in Eq. (D.5) by D_{ND}, D_{NC}, and p_N respectively yields three equations that can be used to calculate $E(p_N)$, $E(D_{ND})$, and $E(D_{NC})$. Similarly, substitution of f in Eq. (D.5) by p_N^2, $p_N D_{ND}$, D_{ND}^2, $D_{ND}D_{NC}$, $p_N D_{NC}$, and D_{NC}^2 respectively yields six equations for calculating $E(p_N^2)$, $E(p_N D_{ND})$, $E(p_N D_{NC})$, $E(D_{ND}^2)$, $E(D_{NC}^2)$, and $E(D_{ND}D_{NC})$. The joint equations can be numerically solved with *Mathematica* tool (Wolfram, 1991).

References

Arnold, J., 1993. Cytonuclear disequilibria in hybrid zones. Ann. Rev. Ecol. Syst. 24, 521–554.

Asmussen, M.A., Arnold, J., 1991. The effects of admixture and population subdivision on cytonuclear disequilibrium. Theor. Popul. Biol. 39, 273–300.

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. Effects of associative 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. Chapman and Hall, New York.

Barton, N.H., 1979. Gene flow past a cline. Heredity 43, 333–340.

Barton, N.H., 1986. The effects of linkage and density-dependent regulation on gene flow. Heredity 57, 415–426.

Barton, N.H., 2000. Genetic hitchhiking. Philos. Trans. R. Soc. Lond. B. 355, 1553–1562.

Barton, N.H., Bengtsson, B.O., 1986. The barrier to genetic exchange between hybridising populations. Heredity 57, 357–376.

Basten, C.J., Asmussen, M.A., 1997. The exact test for cytonuclear disequilibria. Genetics 146, 1165–1171.

Birky, C.W., 1995. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. Proc. Natl. Acad. Sci. USA 92, 11331–11338.

Castro, J.A., Ramon, M., Picornell, A., Moya, A., 1999. The genetic structure of *Drosophila subobscura* populations from the islands of Majorca and Minorca (Balearic island, Spain) based on allozymes and mitochondrial DNA. Heredity 83, 271–279.

Charlesworth, B., Morgan, M.T., Charlesworth, D., 1993. The effect of deleterious mutation on neutral molecular variation. Genetics 134, 1289–1303.

Dakin, E., 2006. Cytonuclear disequilibrium in a spatially structured hybrid zone. Theor. Popul. Biol. 70, 82–91.
Datta, S., Arnold, J., 1998. Dynamics of cytonuclear disequilibria in subdivided populations. J. Theor. Biol. 192, 99–111.
Datta, S., Fu, Y.X., Arnold, J., 1996. Dynamics and equilibrium behaviour of cytonuclear disequilibria under genetic drift. Theor. Popul. Biol. 50, 298–324.
Dean, R., Arnold, J., 1996. Small sample properties for estimators of cytonuclear disequilibria. Heredity 77, 396–399.
Dowling, D.K., Friberg, U., Hailer, F., Arnqvist, G., 2007. Intergenomic epistasis for fitness: Within-population interactions between cytoplasmic and nuclear genes in *Drosophila melanogaster*. Genetics 175, 235–244.
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., A Langdon, A., 1999. Using organelle markers to elucidate the history, ecology and evolution of plant populations. In: Hollingsworth, P.M., Bateman, R.M., Gornall, R.J. (Eds.), Molecular Systematics and Plant Evolution. Taylor and Francis, London, pp. 1–19.
Fu, Y.X., Arnold, J., 1992. Dynamics of cytonuclear disequilibria in finite populations and comparison with a two-locus nuclear system. Theor. Popul. Biol. 41, 1–25.
Hanson, M.R., 1991. Plant mitochondrial mutations and male sterility. Ann. Rev. Genet. 25, 461–486.
Harrison, R.G., 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends Ecol. Evol. 4, 6–11.
Hu, X.S., 2008. Barriers to the spread of neutral alleles in the cytonuclear system. Evolution 62, 2260–2278.
Hu, X.S., Ennos, R.A., 1997. On estimation of the ratio of pollen to seed flow among plant populations. Heredity 79, 541–579.
Hu, X.S., Ennos, R.A., 1999. Impacts of seed and pollen flow on population differentiation for plant genomes with three contrasting modes of inheritance. Genetics 152, 441–450.
Hu, X.S., He, F.L., 2005. Background selection and population differentiation. J. Theor. Biol. 235, 207–219.
Hu, X.S., Li, B.L., 2002. Seed and pollen flow and cline discordance among genes with different modes of inheritance. Heredity 88, 212–217.
Jesson, L.K., Barrett, S.P., 2002. Enantiostyly in Wachendorfia (Haemodoraceae): The influence of reproductive systems on the maintenance of the polymorphism. Am. J. Bot. 89, 253–262.
Kimura, M., 1983. The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.
Kruuk, L.E.B., Baird, S.J.E., Gale, K.S., Barton, N.H., 1999. A comparison of multilocus clines maintained by environmental adaption or by selection against hybrids. Genetics 153, 1959–1971.
Latta, R.G., Linhart, Y.B., Mitton, J.B., 2001. Cytonuclear disequilibrium and genetic drift in a natural population of Ponderosa pine. Genetics 158, 843–850.
Lewontin, R.C., Krakauer, J., 1973. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74, 175–195.
Li, W.H., 1997. Molecular Evolution. Sinauer Associates, Inc., Publisher, Sunderland Massachusetts.
Li, W.H., Nei, M., 1974. Stable linkage disequilibrium without epistasis in subdivided populations. Theor. Popul. Biol. 6, 173–183.
Mackay, T.F.C., 2004. Genetic dissection of quantitative traits. In: Singh, R.S., Uyenoyama, M. (Eds.), The Evolution of Population Biology. Cambridge University Press, UK, pp. 51–73.
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.
Maruyama, T., Birky, C.W., 1991. Effects of periodic selection on gene diversity in organelle genomes and other systems without recombination. Genetics 127, 449–451.
Maynard Smith, J., Haigh, J., 1974. The hitchhiking effect of a favourable gene. Genet. Res. 23, 23–35.
Mazur, B.J., Falco, S.C., 1989. The development of herbicide resistant crops. Ann. Rev. Plant Physio. Plant Mol. Biol. 40, 441–470.
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. Am. J. Bot. 83, 383–404.
Monsen, K.J., Honchak, B.M., Locke, S.E., Peterson, M.A., 2007. Cytonuclear disequilibrium in Chrysochus hybrids is not due to patterns of mate choice. J. Heredity 98, 325–330.
Nordborg, M., Charlesworth, B., Charlesworth, D., 1996. The effect of recombination on background selection. Genet. Res. 67, 159–174.
Ohta, T., Kimura, M., 1969a. Linkage disequilibrium due to random genetic drift. Genet. Res. 13, 47–55.
Ohta, T., Kimura, M., 1969b. Linkage disequilibrium at steady state determined by random genetic drift and recurrent mutation. Genetics 63, 229–238.
Ohta, T., Kimura, M., 1970. Development of associative overdominance through linkage disequilibrium in finite populations. Genet. Res. 16, 165–177.
Orive, M.E., Barton, N.H., 2002. Associations between cytoplasmic and nuclear loci in hybridizing populations. Genetics 162, 1469–1485.
Overath, R.D., Asmussen, M.A., 2000. The cytonuclear effects of facultative apomixes: I. Disequilibrium dynamics in diploid populations. Theor. Popul. Biol. 58, 107–121.
Petit, R.J., Kremer, A., Wagner, D.B., 1993. Finite island model for organelle and nuclear genes in plants. Heredity 71, 630–641.
Rand, D.M., Haney, A.J., A J Fry,, 2004. Cytonuclear coevolution: The genomics of cooperation. Trends Ecol. Evolut. 19, 645–653.
Sarchez, M.S., Basten, C.J., Ferrenberg, A.M., Asmussen, M.A., J Arnold, J., 2006. Exact sample sizes needed to detect dependence in 2×3 tables. Theor. Popul. Biol. 69, 111–120.
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., 1985. Gene flow in natural populations. Ann. Rev. Ecol. Syst. 16, 393–430.
Slatkin, M., 1975. Gene flow and selection in a two-locus system. Genetics 81, 787–802.
Whitlock, M.C., McCauley, D.E., 1999. Indirect measures of gene flow and migration: $F_{st} \neq 1/(4Nm + 1)$. Heredity 82, 117–125.
Wolf, J.B., 2009. Cytonuclear interactions can favor the evolution of genomic imprinting. Evolution 63, 1364–1371.
Wolfe, K.H., Li, W.H., Sharpe, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondria, chloroplast and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84, 9054–9058.
Wolfram, S., 1991. Mathematics. Addison Wesley, New York.
Wright, S., 1969. Evolutionary and the Genetics of Populations. In: The Theory of Gene Frequencies, vol. 2. The University of Chicago Press, Chicago.