

The Effects of Pollen and Seed Migration on Nuclear-Dicytoplasmic Systems. II. A New Method for Estimating Plant Gene Flow From Joint Nuclear-Cytoplasmic Data

Maria E. Orive^{*,†} and Marjorie A. Asmussen[†]

^{*}Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas 66045-2106 and [†]Department of Genetics, University of Georgia, Athens, Georgia 30602-7223

Manuscript received November 25, 1998

Accepted for publication February 14, 2000

ABSTRACT

A new maximum-likelihood method is developed for estimating unidirectional pollen and seed flow in mixed-mating plant populations from counts of joint nuclear-cytoplasmic genotypes. Data may include multiple unlinked nuclear markers with a single maternally or paternally inherited cytoplasmic marker, or with two cytoplasmic markers inherited through opposite parents, as in many conifer species. Migration rate estimates are based on fitting the equilibrium genotype frequencies under continent-island models of plant gene flow to the data. Detailed analysis of their equilibrium structures indicates when each of the three nuclear-cytoplasmic systems allows gene flow estimation and shows that, in general, it is easier to estimate seed than pollen migration. Three-locus nuclear-dicytoplasmic data only increase the conditions allowing seed migration estimates; however, the additional dicytonuclear disequilibria allow more accurate estimates of both forms of gene flow. Estimates and their confidence limits for simulated data sets confirm that two-locus data with paternal cytoplasmic inheritance provide better estimates than those with maternal inheritance, while three-locus dicytonuclear data with three modes of inheritance generally provide the most reliable estimates for both types of gene flow. Similar results are obtained for hybrid zones receiving pollen and seed flow from two source populations. An estimation program is available upon request.

THE juxtaposition of biparental and uniparental inheritance gives joint nuclear-cytoplasmic data special utility in decomposing plant gene flow and estimating the rates of pollen and seed migration. Moreover, plants offer two major cytoplasmic genomes, mitochondrial (mt) and chloroplast (cp) DNA, and three different combinations of uniparental inheritance patterns for this purpose. In most plant species, mtDNA and cpDNA are both inherited maternally, but in some species the two organelles are both inherited paternally or through opposite parents (Kirk and Tilney-Bassett 1978; Birky 1988; Boblenz *et al.* 1990; Harrison and Doyle 1990; Mogensen 1996). This gives us three main types of joint nuclear-cytoplasmic data with varying degrees of power for estimating plant gene flow.

The first is two-locus cytonuclear data with a biparentally inherited nuclear marker and a maternally inherited cytoplasmic marker. This is the most common, but may be the least powerful form of nuclear-cytoplasmic data for estimating pollen flow, since pollen then only carries the nuclear marker. The second is two-locus cytonuclear data with a paternally inherited cytoplasmic marker; although less common, this type of data can be much more informative, since pollen flow will now be

reflected in both the nuclear and cytoplasmic markers. The third and final class is three-locus, nuclear-mitochondrial-chloroplast data combining biparental nuclear inheritance with both forms of uniparental cytoplasmic inheritance. Such dicytonuclear data with three distinct modes of inheritance are currently available from conifer species in the family Pinaceae, which inherit their mitochondria maternally and chloroplasts paternally (Neale *et al.* 1986; Wagner *et al.* 1987; Neale and Sederoff 1989; Dong and Wagner 1994), and should be the most informative system of all.

The theoretical foundation for using these three types of data for the estimation of plant gene flow has been laid out in a series of continent-island migration models for mixed-mating populations. These have fully delimited the effects of unidirectional pollen and seed flow upon both standard two-locus cytonuclear systems, with a single maternally or paternally inherited cytoplasmic marker (Asmussen and Schnabel 1991; Schnabel and Asmussen 1992), and three-locus dicytonuclear systems with both modes of cytoplasmic inheritance (Asmussen and Orive 2000). This extensive analytic framework now allows us to estimate rates of pollen and seed migration from any of the three main types of nuclear-cytoplasmic data in plants by fitting the expected equilibrium frequencies to the observed joint genotypic counts.

Although cytonuclear disequilibria are not necessary

Corresponding author: Maria E. Orive, Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045-2106. E-mail: orive@ukans.edu

for estimation *per se*, permanent nonrandom associations nonetheless increase the chances that the equilibrium state will depend on, and thus allow estimates of, the rates of pollen and seed migration. In this regard, Asmussen and Schnabel (1991) found that, with maternal cytoplasmic inheritance, nonzero cytonuclear disequilibria are maintained only if migrant seeds carry nonrandom cytonuclear associations; pollen dispersal has only a small effect on the disequilibria caused by seed migration. However, with paternal cytoplasmic inheritance (Schnabel and Asmussen 1992), pollen migration significantly affects the equilibrium cytonuclear structure of the resident population through nonrandom cytonuclear associations in the migrant pollen. In addition, the presence of both types of gene flow can then generate permanent cytonuclear disequilibria via intermigrant admixture effects, such as differences in nuclear and chloroplast allele frequencies in migrant pollen and seeds.

For both two-locus systems, the factors necessary for permanent cytonuclear associations can arise in many biological situations. Disequilibria will be present in migrant pollen or seeds, for example, when suitable selection or other nonrandomizing forces act on the source population(s), as well as when gene flow is contributed by multiple, genetically distinct sources, as might be expected in hybrid zones and other areas of admixture. Similarly, allele frequency differences between the two forms of gene flow can be caused by the presence of distinct sources for migrant pollen and seeds, by multiple pollen and seed sources whose relative contributions vary with the form of gene flow, or by selection or other evolutionary forces acting during the life cycle of the source population(s).

Finally, dicytonuclear systems with opposite uniparental inheritance of the two cytoplasmic markers (Asmussen and Orive 2000) can produce permanent cytonuclear associations via all of the pathways for the two-locus systems, as well as cytoplasmic and three-locus associations through new interactions involving all three genomes. The latter require allelic cytonuclear disequilibria for both cytoplasmic markers in migrants, or different allele frequencies in migrant pollen or seeds for one of the three markers plus nonrandom associations in migrant seeds between the other two markers. These results suggest that three-locus, nuclear-mtDNA-cpDNA data juxtaposing both forms of uniparental inheritance should be especially powerful for estimating pollen and seed migration rates, since they provide the greatest number of avenues for the accumulation of nonrandom associations.

Here we develop a formal maximum-likelihood procedure for estimating both types of plant gene flow based on this three-part theoretical framework, using the general, dicytonuclear migration model developed in a companion article (Asmussen and Orive 2000). Unlike previous methods (Petit *et al.* 1993; Ennos 1994; Hu

and Ennos 1999), this new approach allows separate estimates of seed and pollen gene flow rates rather than their ratio or the product of effective population size and migration rate, Nm . The method developed here utilizes joint genotype counts from either a two-locus cytonuclear system (with either a maternally or paternally inherited cytoplasmic marker) or a full three-locus dicytonuclear system with both modes of uniparental transmission. In each case, the maximum-likelihood estimates are the parameter values whose equilibrium genotypic frequencies best fit the data. This method thus requires that the cytonuclear or dicytonuclear structure of the population under analysis be at equilibrium and no longer changing.

We begin by deriving the general conditions under which each system can be used in this way for estimation, illustrating these through consideration of some important special cases. In addition, we test the relative utility of these three types of data for decomposing plant gene flow into pollen and seed migration via simulated data. Finally, since cytonuclear and dicytonuclear data also offer a valuable tool for studying gene flow into a hybrid zone and other areas of admixture (Asmussen *et al.* 1989; Avise *et al.* 1990; Paige *et al.* 1991; Sites *et al.* 1996; Goodisman and Asmussen 1997; Goodisman *et al.* 1998), we extend our analysis to the case of a hybrid plant population receiving pollen and seeds from two source populations or species.

CONDITIONS FOR GENE FLOW ESTIMATION

We wish to determine the conditions under which the cytonuclear and dicytonuclear models of unidirectional pollen and seed migration developed previously (Asmussen and Schnabel 1991; Schnabel and Asmussen 1992; Asmussen and Orive 2000) can be used to estimate plant gene flow. As in Asmussen and Orive (2000), mtDNA here represents a maternally inherited and cpDNA a paternally inherited cytoplasmic marker. The underlying models assume each locus is diallelic with alleles A and a at the nuclear marker, alleles (cytotypes) M and m at the mitochondrial marker, and C and c at the chloroplast marker. There are thus 6 possible joint n-mtDNA and n-cpDNA genotypes (Table 1) and 12 possible joint n-mtDNA-cpDNA genotypes (Table 2). Definitions and notation for all variables (Table 3) follow those given in a companion article (Asmussen and Orive 2000). Plain upper case letters denote variables in adults of the resident (study) population, with a caret (^) indicating an equilibrium value; the corresponding values in the two migrant pools are indicated by overbars, with upper case letters denoting values in migrant pollen and lower case letters denoting the values in migrant seeds. For instance, P represents the nuclear allele frequency in the resident population, \hat{P} the equilibrium nuclear allele frequency in this population, and \bar{P} and \bar{p} the nuclear allele frequencies in mi-

TABLE 1
Joint n-mtDNA and n-cpDNA genotype frequencies and decomposition in terms of disequilibria

Cytotype	Nuclear genotype			Total
	AA	Aa	aa	
<i>M</i>	$U_{1M} = UX_M + D_{AA/M}$	$V_{1M} = VX_M + D_{Aa/M}$	$W_{1M} = WX_M + D_{aa/M}$	X_M
<i>m</i>	$U_{2M} = UY_M - D_{AA/M}$	$V_{2M} = VY_M - D_{Aa/M}$	$W_{2M} = WY_M - D_{aa/M}$	Y_M
Total	U	V	W	1
<i>C</i>	$U_{1C} = UX_C + D_{AA/C}$	$V_{1C} = VX_C + D_{Aa/C}$	$W_{1C} = WX_C + D_{aa/C}$	X_C
<i>c</i>	$U_{2C} = UY_C - D_{AA/C}$	$V_{2C} = VY_C - D_{Aa/C}$	$W_{2C} = WY_C - D_{aa/C}$	Y_C
Total	U	V	W	1

grant pollen and seeds, respectively. The frequencies of the diallelic cytonuclear combinations carried by migrant pollen are denoted as in Table 4.

All three migration models apply to mixed-mating populations with nonoverlapping generations (Clegg 1980) in which reproduction is a combination of selfing (probability *s*) and of outcrossing (probability $1 - s$), where $0 < s < 1$, as well as to populations that reproduce solely by outcrossing ($s = 0$) or solely by selfing ($s = 1$). The study population is censused among adults, and gene flow occurs each generation via pollen and seeds following the generation cycle shown in Figure 1. Here, the pollen migration rate *M* represents the fraction of outcrossed pollen derived each generation from migrants, with the remaining fraction, $1 - M$, of local origin, where $0 \leq M < 1$; the overall fraction of migrant pollen per generation is thus the product, $M(1 - s)$. Similarly, each generation a fraction *m* of the total seed pool is assumed to be derived from the source population(s) and the remaining fraction $1 - m$ from the resident population, where $0 \leq m < 1$. The migration rates and mating system, as well as the genetic composition of the source(s), are assumed constant over time; this corresponds to a continent-island model of migration with unidirectional migration (Figure 2). In addition, we assume that the markers are unaffected by selection, mutation, or random genetic drift in the resident population.

General considerations: A basic consideration for any

estimation procedure is that the number of parameters to be estimated not exceed the degrees of freedom (*i.e.*, number of independent classes) in the data. This means that at most 5 parameters can be estimated from the 6 joint genotypes in two-locus cytonuclear data (Table 1), and at most 11 from the 12 joint genotypes in three-locus nuclear-mitochondrial-chloroplast (n-mtDNA-cpDNA) data (Table 2). Tests for goodness-of-fit between observed and expected frequencies will require a further degree of freedom, leaving at most 4 parameters that may be estimated from each two-locus system and 10 from the full, three-locus system. In each system, additional degrees of freedom and estimation power are possible by assaying multiple, unlinked nuclear markers.

The estimation method developed here additionally requires that the equilibrium state of the relevant cytonuclear or dicytonuclear migration model depend on the parameter(s) being estimated. This requires that the parameter(s) independently affect the equilibrium value of at least one variable out of the set of independent variables used to describe the system. Sufficient sets of equilibrium values for each of the three possible types of data are provided by the theory developed in Asmussen and Schnabel (1991), Schnabel and Asmussen (1992), and Asmussen and Orive (2000). Using the definitions and notation in Table 3, a set of five independent variables for the equilibrium of a two-locus n-mtDNA system with maternal inheritance includes the final frequencies of nuclear alleles (\hat{P}) and nuclear het-

TABLE 2
Joint n-mtDNA-cpDNA genotype frequencies and decomposition in terms of disequilibria

Joint cytotype	Nuclear genotype		
	AA	Aa	aa
<i>MC</i>	$U_{11} = UX_{11} + D_{AA/MC}$	$V_{11} = VX_{11} + D_{Aa/MC}$	$W_{11} = WX_{11} + D_{aa/MC}$
<i>Mc</i>	$U_{12} = UX_{12} - D_{AA/MC} + D_{AA/M}$	$V_{12} = VX_{12} - D_{Aa/MC} + D_{Aa/M}$	$W_{12} = WX_{12} - D_{aa/MC} + D_{aa/M}$
<i>mC</i>	$U_{21} = UX_{21} - D_{AA/MC} + D_{AA/C}$	$V_{21} = VX_{21} - D_{Aa/MC} + D_{Aa/C}$	$W_{21} = WX_{21} - D_{aa/MC} + D_{aa/C}$
<i>mc</i>	$U_{22} = UX_{22} + D_{AA/MC} - D_{AA/M} - D_{AA/C}$	$V_{22} = VX_{22} + D_{Aa/MC} - D_{Aa/M} - D_{Aa/C}$	$W_{22} = WX_{22} + D_{aa/MC} - D_{aa/M} - D_{aa/C}$

To complete the decomposition, $U = P - \frac{1}{2}V$, $W = 1 - P - \frac{1}{2}V$, $D_{AA/*} = D_{A/*} - \frac{1}{2}D_{Aa/*}$, $D_{aa/*} = -D_{A/*} - \frac{1}{2}D_{Aa/*}$, where * indicates *M*, *C*, or *MC*; and $X_{ij} = X_{iM}X_{jC} + (-1)^{i+j}D_{M/C}$ for $i, j = 1, 2$, where $X_{1M} = X_M$, $X_{2M} = Y_M$, $X_{1C} = X_C$, and $X_{2C} = Y_C$.

TABLE 3
Key frequency and disequilibrium variables in cytonuclear and dicytonuclear systems

Definition of variable ^a	Description
$P = \text{freq}(A) = \text{freq}(AA) + 1/2\text{freq}(Aa)$	Nuclear allele frequency
$V = \text{freq}(Aa)$	Nuclear heterozygote frequency
$X_M = \text{freq}(M)$	Mitochondrial frequency
$X_C = \text{freq}(C)$	Chloroplast frequency
$D_{N/M} = \text{freq}(N/M) - \text{freq}(N)\text{freq}(M)$	Allelic ($N = A$) or genotypic ($N = AA, Aa, aa$) cytonuclear disequilibrium for the mitochondrial marker
$D_{N/C} = \text{freq}(N/C) - \text{freq}(N)\text{freq}(C)$	Allelic ($N = A$) or genotypic ($N = AA, Aa, aa$) cytonuclear disequilibrium for the chloroplast marker
$D_{M/C} = \text{freq}(MC) - \text{freq}(M)\text{freq}(C)$	Cytoplasmic disequilibrium between the two cytoplasmic markers
$D_{N/MC} = \text{freq}(N/MC) - \text{freq}(N)\text{freq}(MC)$	Three-locus joint allelic ($N = A$) or joint genotype ($N = AA, Aa, aa$) disequilibrium for the joint cytotype MC
$D_{N/M/C} = \text{freq}(N/M/C) - f_N X_M X_C - f_N D_{M/C} - X_M D_{N/C} - X_C D_{N/M}$	Three-way allelic ($N = A$) or three-way genotype ($N = AA, Aa, aa$) disequilibrium for the cytotype M/C

^a freq denotes frequency and $f_N = \text{freq}(N)$.

erozygotes (\hat{V}), the final mitochondrial frequency (\hat{X}_M), and the final allelic and genotypic cytonuclear disequilibria giving associations between mtDNA cytotypes and nuclear alleles ($\hat{D}_{A/M}$) and heterozygotes ($\hat{D}_{Aa/M}$). The corresponding set of variables for a two-locus n-cpDNA system with paternal inheritance includes the final frequencies of the nuclear alleles and nuclear heterozygotes (\hat{P} and \hat{V}), the equilibrium chloroplast frequency (\hat{X}_C), and the final allelic ($\hat{D}_{A/C}$) and heterozygote ($\hat{D}_{Aa/C}$) cytonuclear disequilibria between the nuclear and chloroplast markers. The formulas for these are given in appendix a as (A1–A8).

Last, for a full three-locus nuclear-dicytoplasmic system, a sufficient set of 11 variables includes the 8 variables listed above for the two cytonuclear systems, together with the final cytoplasmic disequilibrium between the two cytoplasmic markers ($D_{M/C}$) and the joint allelic and genotypic disequilibria for the three markers, measuring associations between the joint mitochondrial-chloroplast cytotypes and the nuclear alleles ($D_{A/MC}$) and heterozygotes ($D_{Aa/MC}$) [(A9–A11) in appendix a]. An alternate parameterization of the three-locus system replaces the joint disequilibria ($D_{A/MC}$ and $D_{Aa/MC}$) with the final three-way allelic ($D_{A/M/C}$) and heterozygote ($D_{Aa/M/C}$) disequilibria [(A12–A13) in appendix a], which measure the associations among the three markers (nu-

clear, mitochondrial, and chloroplast) after taking into account all of the possible two-way associations between them (nuclear-mitochondrial, nuclear-chloroplast, and mitochondrial-chloroplast).

General estimation conditions: Examination of these equilibrium formulas readily gives the general conditions under which each of the three systems can be used to estimate the rate of seed (m) and pollen (M) migration. These are given in Table 5 for the general case of mixed-mating populations. The necessary conditions include (i) nuclear polymorphism in migrant seeds ($0 < \bar{p} < 1$), (ii) unequal nuclear or chloroplast allele frequencies in migrant pollen and seeds ($\bar{P} \neq \bar{p}$ or $\bar{X}_C \neq \bar{x}_c$), and (iii) the existence of two-locus disequilibria between chloroplast and nuclear alleles in migrant pollen ($\bar{D}_{A/C} \neq 0$) or seeds ($\bar{d}_{A/C} \neq 0$), or between chloroplast and mitochondrial alleles in migrant seeds ($\bar{d}_{M/C} \neq 0$). Allele frequency differences between the two types of migrant pools (ii) could be caused by distinct sources for migrant pollen and seeds, or by selection or other evolutionary forces acting during the life cycle of the source population. Such intermigrant differences could also be caused by having unequal pollen and seed dispersal from multiple sources (*e.g.*, source 1 may contribute a greater fraction of migrant pollen than migrant seeds, while source 2 contributes a greater fraction of

TABLE 4
Joint cytonuclear frequencies in migrant pollen

Cytotype	Nuclear allele		Total
	A	a	
C	$\bar{P}_{1C} = \bar{P}\bar{X}_C + \bar{D}_{A/C}$	$\bar{Q}_{1C} = \bar{Q}\bar{X}_C - \bar{D}_{A/C}$	\bar{X}_C
c	$\bar{P}_{2C} = \bar{P}\bar{Y}_C - \bar{D}_{A/C}$	$\bar{Q}_{2C} = \bar{Q}\bar{Y}_C + \bar{D}_{A/C}$	\bar{Y}_C
Total	\bar{P}	\bar{Q}	1

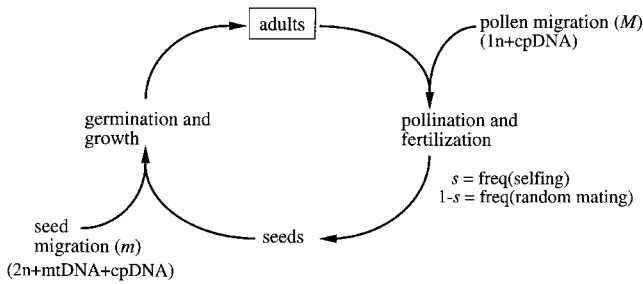


Figure 1.—Adult census model showing the two types of gene flow within a generation cycle. Pollen migration occurs first, at rate M , followed by seed migration at rate m . Mating is a mixture of selfing, which occurs at rate s , and outcrossing, which occurs at rate $1 - s$. Here, mtDNA and cpDNA represent a maternally inherited and a paternally inherited cytoplasmic marker, respectively.

migrant seeds than pollen). Migrant pollen or seeds would be expected to carry associations between chloroplast and nuclear or mitochondrial alleles (iii) when selection or other evolutionary forces act on the source population, as well as when the pollen or seeds are contributed by multiple, genetically distinct sources, as might be expected in hybrid zones and other areas of admixture.

The results given in Table 5 show that it is harder to estimate pollen flow than seed flow; whenever the data allow estimation of pollen migration, the rate of seed migration can also be estimated. This is not surprising, given that only two of the three markers experience movement via pollen migration, while all three move during seed migration. Moreover, an inspection of the conditions in Table 5 reveals that there are at least three situations when only seed migration can be estimated. Two of these arise when there are equal nuclear allele frequencies in the two migrant pools ($\bar{P} = \bar{p}$). In such cases, the rate of pollen flow cannot be estimated using cytonuclear data with maternal cytoplasmic inheritance (n-mtDNA), while estimation of the seed migration rate requires only that the migrant seeds be polymorphic at the nuclear marker ($0 < \bar{p} < 1$). The latter condition also allows estimates of the seed migration rate from cytonuclear data with paternal cytoplasmic inheritance (n-cpDNA) and n-mtDNA-cpDNA data with both forms of cytoplasmic inheritance in cases where none of the three systems allows estimation of the pollen migration rate (*i.e.*, no intermigrant allele frequency differences in the nuclear or paternally inherited marker, or allelic cytonuclear disequilibria for the paternally inherited marker in migrant pollen and seeds). Finally, nonrandom associations between the two cytoplasmic markers ($\bar{d}_{M/C} \neq 0$) allow seed migration estimates from n-mtDNA-cpDNA data in the absence of all four conditions that allow pollen migration estimates.

Three-locus data allow estimation of the gene flow parameters whenever data can be used from at least one

of the two-locus cytonuclear systems and also sometimes when neither of these allows estimation. However, dicytonuclear data with three modes of inheritance increase only the conditions for estimating the seed migration rate; their conditions for estimating the pollen migration rate (M) are the same as for two-locus data with paternal inheritance (n-cpDNA). Three-locus data may nevertheless increase the power to detect the two forms of gene flow and/or give more accurate estimates because the rates of pollen and seed flow enter into more of the terms that determine the final state of the dicytonuclear system via cytonuclear and dicytonuclear disequilibria in migrant seeds (appendix a). These migrant associations are not listed explicitly in the estimation conditions given in Table 5, since they require that migrant seeds be polymorphic at the nuclear marker, which is itself sufficient for estimation of the rate of seed migration.

Estimation conditions for special cases: To further explore the conditions under which the two gene flow rates may be estimated, we consider the important special cases of the general continent-island model of unidirectional pollen and seed migration presented in a companion article (Asmussen and Orive 2000). These include (1) seed migration alone ($0 < m < 1$, $M = 0$), (2) pollen migration alone ($0 < M < 1$, $m = 0$), (3) complete random mating ($s = 0$), (4) complete self-fertilization with seed migration ($s = 1$, $0 < m < 1$), (5) equal nuclear allele frequencies in the two migrant pools ($\bar{P} = \bar{p}$), (6) equal frequencies of the paternally inherited cytoplasmic marker in the two migrant pools ($\bar{X}_C = \bar{x}_C$), (7) equivalent migrant pools ($\bar{P} = \bar{p}$, $\bar{X}_C = \bar{x}_C$, $\bar{D}_{A/C} = \bar{d}_{A/C}$), and (8) no migrant disequilibria ($\bar{D} = \bar{d} = 0$ for all migrant pollen and seed disequilibria). The equilibrium structure in these eight simpler cases provides valuable insight into the conditions under which the equilibria for the dicytonuclear system and the two two-locus cytonuclear systems lose their dependence on, and their utility for estimating, either of the two migration rates. The details for these are given in Asmussen and Orive (2000).

1. *Seed migration alone* ($0 < m < 1$, $M = 0$): The equilibria for all three systems (n-mtDNA, n-cpDNA, and n-mtDNA-cpDNA) depend on, and allow the estimation of, the seed migration rate when this is the sole form of gene flow, as long as the migrant seeds are polymorphic for the nuclear marker ($0 < \bar{p} < 1$). Data from the n-mtDNA-cpDNA system can be used in the additional case where migrant seeds carry nonrandom associations between the two cytoplasmic markers ($\bar{d}_{M/C} \neq 0$).

2. *Pollen migration alone* ($0 < M < 1$, $m = 0$): The absence of seed migration places significant constraints on the estimation of the pollen migration rate, M . The equilibrium for cytonuclear systems with maternal inheritance (n-mtDNA) is independent of M and thus cannot be used to estimate the rate of pollen migration if this is the only form of gene flow. The n-cpDNA and

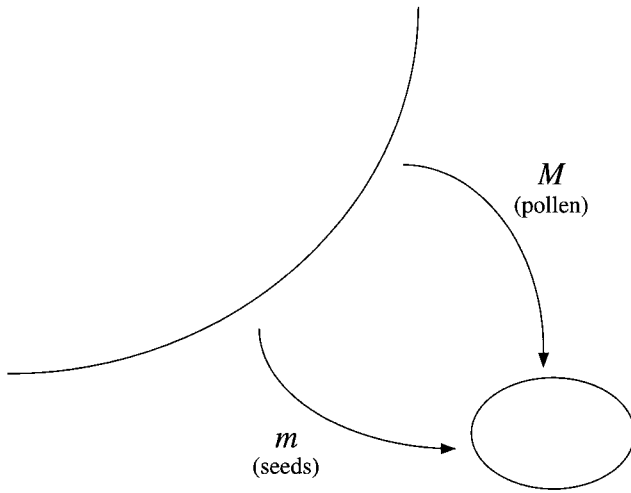


Figure 2.—Continent-island migration model with unidirectional pollen (M) and seed (m) migration.

n-mtDNA-cpDNA systems with a paternally inherited cytoplasmic marker do allow estimation of the pollen migration rate, provided that the migrant pollen carry nonrandom cytonuclear allelic associations ($\bar{D}_{A/C} \neq 0$).

3. *Complete random mating ($s = 0$):* Complete random mating places no additional constraints on estimation for the rates of pollen and seed flow beyond those given in Table 5 for the general case of mixed-mating populations.

4. *Complete self-fertilization with seed migration ($s = 1$, $0 < m < 1$):* The opposite extreme of complete selfing is distinctive since the lack of outcrossing means such populations are also closed to pollen flow. Both types of two-locus cytonuclear systems allow estimation of the seed migration rate (m) in purely selfing populations as long as the migrant seed pool includes either heterozygous seeds ($\bar{v} \neq 0$) or cytonuclear disequilibria for that two-locus system ($\bar{d}_{N^*} \neq 0$ for $N = A, AA, Aa$, or aa and $*$ = M or C). Dicytonuclear data may also be used in such cases, as well as when migrant seeds carry joint or three-way disequilibria ($\bar{d}_{N^*} \neq 0$ for $N = A, AA, Aa$, or aa and $*$ = MC or M/C).

5. *Equal nuclear allele frequencies in migrant pollen and seeds ($\bar{P} = \bar{p}$):* Although we have noted a number of situations that can produce unequal frequencies in the two forms of gene flow, in other cases they will be the same. When this holds for the nuclear marker, the residents' nuclear allele frequency reaches the common migrant value ($\hat{P} = \bar{P} = \bar{p}$). As a result, the cytonuclear system with maternal cytoplasmic inheritance (n-mtDNA) loses its dependence on the pollen migration rate and thus cannot be used to estimate this type of gene flow (Asmussen and Schnabel 1991). The n-cpDNA system with paternal cytoplasmic inheritance and the full n-mtDNA-cpDNA system can still be used to estimate both types of gene flow under all of the other conditions given for the general case (Table 5),

although their estimation power may be reduced due to the elimination of several intermigrant admixture effects that contribute to permanent disequilibria.

6. *Equal frequencies of the paternally inherited cytoplasmic marker in the two migrant pools ($\bar{X}_C = \bar{x}_C$):* If the paternally inherited marker has equal frequencies in the two migrant pools, its frequency in the resident population approaches the common migrant value ($\hat{X}_C = \bar{X}_C = \bar{x}_C$). As with equal nuclear allele frequencies, this reduces the power for estimating gene flow rates for both the n-cpDNA and n-mtDNA-cpDNA systems with a paternally inherited marker by eliminating many of the intermigrant factors that generate permanent disequilibria.

7. *Equivalent migrant pools ($\bar{P} = \bar{p}$, $\bar{X}_C = \bar{x}_C$, $\bar{D}_{A/C} = \bar{d}_{A/C}$):* If the two migrant pools are equivalent, cytonuclear data with maternal inheritance (n-mtDNA) are restricted to estimating the rate of seed migration (and then only if migrant seeds are polymorphic for the nuclear marker, as is true whenever $\bar{P} = \bar{p}$). The two- and three-locus systems with a paternally inherited cytoplasmic marker (n-cpDNA and n-mtDNA-cpDNA) still allow estimation of both types of gene flow as long as the migrant pollen and seeds carry nonrandom associations between nuclear and paternally transmitted cytoplasmic alleles ($\bar{D}_{A/C}, \bar{d}_{A/C} \neq 0$).

8. *No migrant disequilibria ($\bar{D} = \bar{d} = 0$, for all migrant pollen and seed disequilibria):* With no migrant disequilibria, it should be possible to estimate both types of gene flow from any of the three systems as long as the two migrant pools differ in their nuclear allele frequencies ($\bar{P} \neq \bar{p}$, see special case 5 above). Further, the cytonuclear system with paternal transmission (n-cpDNA) and the full three-locus system (n-mtDNA-cpDNA) can also provide both estimates if migrant pollen and seeds have distinct frequencies of the paternally inherited marker ($\bar{X}_C \neq \bar{x}_C$).

ESTIMATING GENE FLOW

Here we present a new method to estimate rates of unidirectional pollen and seed migration using joint cytonuclear or dicytonuclear frequencies. We focus specifically on the case of uniparentally inherited cytoplasmic markers where, in the dicytonuclear case, these are inherited through opposite parents. This method uses data in the form of counts in adults of joint n-mtDNA, n-cpDNA, or n-mtDNA-cpDNA genotypes, where mtDNA here represents a maternally inherited marker and cpDNA a paternally inherited marker. We first outline the general approach for the simplest cases with a single diallelic marker from each genome and then indicate two initial extensions to data involving multiple unlinked nuclear markers and/or multiallelic markers. The utility of this method is illustrated with simulated data, which also allows us to compare estimates from dicytonuclear data with opposite cytoplasmic inheritance with estimates obtained from two-locus

TABLE 5

Conditions under which n-mtDNA, n-cpDNA, and n-mtDNA-cpDNA data can be used to estimate the rate of seed (*m*) and pollen (*M*) migration for the general mixed-mating model

Migration parameter	Cytonuclear system ^a		
	n-mtDNA	n-cpDNA	n-mtDNA-cpDNA
<i>m</i>	$0 < \bar{p} < 1$ or $\bar{P} \neq \bar{p}$	$0 < \bar{p} < 1$ or $\bar{P} \neq \bar{p}$ or $\bar{X}_C \neq \bar{x}_C$ or $\bar{D}_{A/C} \neq 0$	$0 < \bar{p} < 1$ or $\bar{P} \neq \bar{p}$ or $\bar{X}_C \neq \bar{x}_C$ or $\bar{D}_{A/C} \neq 0$ or $\bar{d}_{M/C} \neq 0$
<i>M</i>	$\bar{P} \neq \bar{p}$	$\bar{P} \neq \bar{p}$ or $\bar{X}_C \neq \bar{x}_C$ or $\bar{D}_{A/C} \neq 0$ or $\bar{d}_{M/C} \neq 0$	$\bar{P} \neq \bar{p}$ or $\bar{X}_C \neq \bar{x}_C$ or $\bar{D}_{A/C} \neq 0$ or $\bar{d}_{M/C} \neq 0$
<i>m</i> and <i>M</i>	Same as for <i>M</i> alone	Same as for <i>M</i> alone	Same as for <i>M</i> alone

^a mtDNA and cpDNA represent maternally and paternally inherited cytoplasmic markers, respectively.

cytonuclear data with a single maternally or paternally inherited cytoplasmic marker. A program implementing this estimation procedure is available from the authors upon request.

Maximum-likelihood estimation: We use maximum likelihood (Edwards 1992) in conjunction with the general continent-island model of pollen and seed dispersal (Figures 1 and 2) developed in Asmussen and Orive (2000) to find the rates of seed and pollen migration whose equilibrium cytonuclear or dicytonuclear genotype frequencies best fit the observed genotypic counts. The equilibrium frequencies of the two-locus cytonuclear genotypes (n-mtDNA and n-cpDNA, Table 1) or the three-locus genotypes (n-mtDNA-cpDNA, Table 2) are specified in terms of the migration and selfing rates (*M*, *m*, *s*) and the known migrant seed genotype frequencies (e.g., $\bar{u}_{11}, \dots, \bar{w}_{22}$) and migrant pollen frequencies (e.g., $\bar{P}_{1G}, \dots, \bar{Q}_{2G}$, Table 4), using the equilibrium formulas in (A1–A13) for the general case in conjunction with the decompositions shown in Tables 1 and 2. We then jointly estimate the three parameters *M*, *m*, and *s* by finding their values that maximize the log-likelihood function for the observed joint genotypic counts, using the “simulated annealing” minimization routine ameba from *Numerical Recipes in C*, Ed. 2 (Press *et al.* 1992).

The full log-likelihood function for dicytonuclear data is given by

$$\ln L(N_{AA/M/C}, \dots, N_{aa/m/c} | M, m, s) = N_{AA/M/C} \ln(\hat{U}_{11}) + \dots + N_{aa/m/c} \ln(\hat{W}_{22}) \quad (1)$$

(ignoring the constant multinomial coefficient), where, for example, $N_{AA/M/C}$ gives the observed count of adults with the *AA/M/C* joint genotype, and \hat{U}_{11} gives the equilibrium frequency of that joint genotype for the specified parameter values under our model. If unknown, the pro-

gram may be modified so as to simultaneously estimate the migrant frequencies from the source population. In this case, the migrant frequencies become parameters for the likelihood function, in addition to the migration and selfing rate parameters *M*, *m*, and *s*. The 95% confidence limits are found by bootstrapping the data set a user-specified number of times and dropping the lowest 2.5% and highest 2.5% of each estimate. The program gives the lower and upper bounds on the confidence interval as well as the length of the interval, obtained by subtracting the lower from the upper bound.

Multiple nuclear markers: An extension for estimating rates of seed and pollen migration from data with multiple unlinked nuclear markers is straightforward. In this case, the overall likelihood is calculated as a product of the likelihoods for each of the nuclear markers considered separately (Asmussen *et al.* 1989); the overall log likelihood is thus a sum of *n* factors like (1), one for each of the *n* unlinked nuclear markers. For example, if $N_{ij,k,tx}$ gives the observed count for a particular joint genotype involving nuclear marker *x* (where the nuclear alleles are now denoted by numerical indices *i, j*, the mitochondrial cytotype by the index *k*, and the chloroplast cytotype by the index *l*) and $\hat{F}_{ij,k,tx}$ gives the expected frequency of that joint genotype under our model, the log-likelihood function is given by

$$\ln L(N_{ij,k,tx} \text{ for all } i \leq j, k, l, x | M, m, s) = \sum_x \sum_{i \leq j} \sum_k \sum_l N_{ij,k,tx} \ln(\hat{F}_{ij,k,tx}), \quad (2)$$

where the outer sum is over each nuclear marker, $x = 1, \dots, n$. Estimates for the parameters are obtained by maximizing this composite function via the optimization routine described above.

Consideration of linked nuclear markers requires simultaneous consideration of the various recombination

frequencies (Barton and Turelli 1991; Turelli and Barton 1994) and is beyond the scope of this initial treatment.

Multiallelic data: An extension to multiple alleles greatly increases the number and complexity of associations considered by this model. As a first step to addressing multiallelic data, such data may be converted to diallelic form by grouping one allele *vs.* all others at a locus. This approach was chosen for the estimation program developed here, because it is the simplest and allows the user to specify the grouping for the locus, rather than arbitrarily averaging over all possible groupings, which could be difficult to interpret. Further theory must be developed to allow a more comprehensive multiallelic analysis.

Results from simulated data: We tested this method with simulated diallelic data sets containing a single nuclear marker; counts of the 12 possible joint three-locus genotypes ($N_{AA/M/C} \dots, N_{aa/m/c}$) in Table 2 were generated as random samples from the equilibrium genotypic distributions under specified migration and selfing rates and specified migrant pollen and seed frequencies. Due to the many possible combinations of parameter values, the examples chosen are not meant to be exhaustive, but instead illustrate some of the factors that affect gene flow estimation using this method. To assess the relative utility of the three possible types of data, we also extracted counts for the two-locus genotypes in each two-locus cytonuclear system. For each run, all three data sets were bootstrapped 200 times to construct 95% confidence intervals for each of the three estimated parameters (M , m , and s).

We consider two different populations (designated A and B) whose migrant compositions differ in the conditions that allow estimation of the two rates of gene flow (Table 5). For population A, the nuclear allele frequency is equal in migrant pollen and seeds ($\bar{P} = \bar{p} = 0.7$), while the frequency of the paternally inherited marker differs in the two migrant pools ($\bar{X}_C = 1.0$, $\bar{x}_C = 0.7$), allowing us to examine the effect of having only one type of intermigrant frequency difference on estimating the two migration rates. The resident population receives migrant seeds of two types, $AA/M/C$ and $aa/m/c$ ($\bar{u}_{11} = 0.7$, $\bar{w}_{22} = 0.3$), which produces cytoplasmic as well as all possible allelic and homozygote disequilibria in the migrant seeds ($\bar{d}_{M/C} = \bar{d}_{A/*} = \bar{d}_{aa/*} = -\bar{d}_{aa/*} = 0.21$ and $\bar{d}_{aa/*} = \bar{d}_{AA/M/C} = 0$, where * indicates M , C , or MC ; $\bar{d}_{A/M/C} = \bar{d}_{AA/M/C} = -\bar{d}_{aa/M/C} = -0.084$). The migrant pollen are also of two types, A/C and a/C ($\bar{P}_{1C} = 0.7$, $\bar{Q}_{1C} = 0.3$, $\bar{D}_{A/C} = 0$), but carry no nonrandom associations. Population B receives the same types of migrant seeds as population A and thus has the same disequilibria in the migrant seed pool. The migrant pollen also still carry no allelic cytonuclear disequilibrium, but are now monomorphic for both the nuclear and paternally transmitted cytoplasmic markers ($\bar{Q}_{1C} = 1.0$). As a result, population B has intermigrant frequency differences for both these markers ($\bar{P} = 0$, $\bar{p} = 0.7$, $\bar{X}_C = 1.0$, $\bar{x}_C =$

0.7), allowing more ways to generate permanent disequilibria and estimate gene flow than population A.

These two situations, with disequilibria in migrant seeds but not in migrant pollen, could arise if a population receives migrant seeds from multiple sources, but migrant pollen from only one source, distinct from the sources for migrant seeds. For instance, a central population might receive animal-dispersed seeds from two nearby populations but only receive wind-dispersed pollen in one prevailing direction from a more distant population. The gene flow estimates in the numerical examples below would then represent composite estimates of the overall rates of pollen and seed flow from all sources. The hybrid zone model in the subsequent section outlines how to estimate the separate contributions from each source.

For each migrant composition, we examined various combinations of the seed migration ($m = 0, 0.05, 0.1$), pollen migration ($M = 0, 0.05, 0.1, 0.2$), and selfing ($s = 0.1, 0.5, 0.9$) rates. We estimated these parameters from three simulated data sets with representative sample sizes of $N = 100, 300$, and 500 . The results for m and M are given in appendix c, where the symbol $\langle z \rangle$ indicates the estimated value for each parameter z . The estimates are normalized relative to their deviation from the true value,

$$\Delta(z) = \left| \frac{\langle z \rangle - z}{z} \right|.$$

When the parameter value is zero, the estimate itself is given instead. Also presented in appendix c are the lengths of the 95% confidence intervals for the estimates of m (I_m) and M (I_M), each of which has a maximum possible length of 1.0.

When only the frequency of the paternally inherited cytoplasmic marker differs between the two migrant pools (population A, Table C1), the n-mtDNA data, with a maternally inherited cytoplasmic marker, cannot estimate the pollen migration rate (Table 5) because the equilibrium state for this cytonuclear system is then independent of the pollen migration rate (Asmussen and Schnabel 1991). In such cases, however, the n-cpDNA data set, with a paternally inherited cytoplasmic marker, and the n-mtDNA-cpDNA data set, with both forms of cytoplasmic inheritance, generally give good estimates of both seed and pollen migration rates, considering accuracy together with the size of the confidence interval ($\Delta(m) < 0.5$ for 83.3% of estimates, $I_m < 0.23$ for all estimates; $\Delta(M) < 0.5$ for 75% of estimates, $I_M < 0.4$ for 71% of estimates).

All three systems consistently give good estimates when the allele frequencies of both the nuclear and the paternally inherited cytoplasmic marker differ in the migrant pools (population B, Table C2), so that permanent disequilibria are generated by intermigrant admixture effects. Such differences in allele frequencies would be expected, for example, if multiple, genetically dis-

tinct sources each contributed differentially to the migrant pollen and seed pools. When there is no seed migration ($m = 0$), none of the three systems can estimate the pollen migration rate in our population B examples because the absence of allelic cytonuclear disequilibrium in migrant pollen ($\bar{D}_{A/C} = 0$) makes the equilibrium state independent of the pollen migration rate (see the special case $m = 0$ above). In other cases with no seed migration, where migrant pollen carry cytonuclear allelic disequilibrium ($\bar{D}_{A/C} \neq 0$), estimation of the pollen migration rate is possible from two- and three-locus cytonuclear data containing a paternally inherited marker (n-cpDNA and n-mtDNA-cpDNA). This, however, requires knowledge of the initial frequency for the maternally inherited marker, X_M^0 , which is now the marker's expected frequency since its value is not affected by pollen migration alone (appendix a); if necessary, this may be treated as an unknown parameter and jointly estimated along with M and s .

In general, increased sample sizes appear to improve the accuracy of the estimates and decrease the size of the confidence intervals for both migrant compositions (A and B), although there is a great deal of variability among different simulated data sets. High selfing rates ($s = 0.9$) generally worsen estimates of pollen migration, although again there is a great deal of variation across runs. These poor estimates may be because high selfing reduces the total fraction of migrant pollen, $M(1 - s)$, so that such pollen contribute less to the genetic composition of the population at equilibrium. The equilibrium genotype frequencies are thus less useful for estimating the pollen migration rate under high selfing.

Figures 3 and 4 give results from two particular test runs for these same migrant compositions (A and B). In each case, the estimates of the seed migration rate $\langle m \rangle$ and the pollen migration rate $\langle M \rangle$ for the three systems (n-mtDNA, n-cpDNA, and n-mtDNA-cpDNA) are given for the three sample sizes: $N = 100, 300,$ and 500 (estimates of selfing rate not shown). The dashed lines give the actual values of the migration rates, the solid boxes give the estimates, and the open boxes indicate the upper and lower bounds for the 95% confidence limits.

Figure 3 gives an example from population A ($\bar{P} = \bar{p}, \bar{X}_C \neq \bar{x}_C$) with a selfing rate of $s = 0.1$. Estimates of seed migration rates are close to the true value and confidence intervals are small for all three systems ($I_m \leq 0.154$), with the exception of n-cpDNA for $N = 100$ ($I_m = 0.227$). For this example, the n-mtDNA system could not estimate the pollen migration rate because, as noted in the special case $\bar{P} = \bar{p}$ above, the equilibrium state for cytonuclear systems with a maternally inherited cytoplasmic marker does not depend on the pollen migration rate with equal nuclear frequencies in the migrant pools. The n-mtDNA-cpDNA system gave smaller confidence intervals for all of the estimates.

An example from population B ($\bar{P} \neq \bar{p}, \bar{X}_C \neq \bar{x}_C$) where the selfing rate is $s = 0.5$ is given in Figure 4. All three

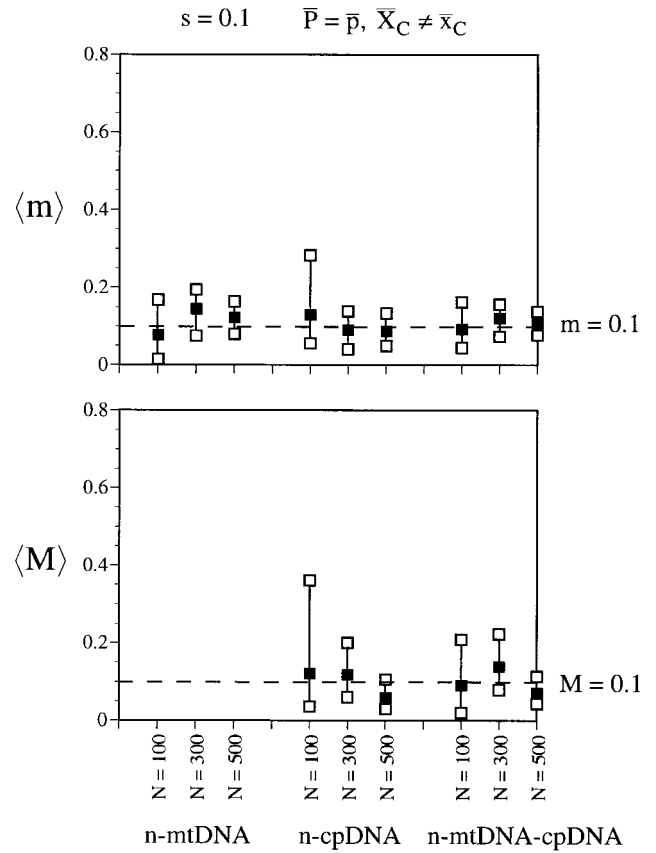


Figure 3.—Estimates of seed migration rate (m) and pollen migration rate (M), where $\bar{P} = \bar{p}, \bar{X}_C \neq \bar{x}_C$, and $s = 0.1$. Values for migrant pools were $\bar{u}_{11} = 0.7, \bar{w}_{22} = 0.3, \bar{P}_{1C} = 0.7,$ and $\bar{Q}_{1C} = 0.3$. Disequilibria in the migrant pools were $\bar{d}_{M/C} = \bar{d}_{A^*/} = \bar{d}_{AA^*/} = -\bar{d}_{aa^*/} = 0.21, \bar{d}_{Aa^*/} = \bar{d}_{Aa/M/C} = 0$ (where * indicates $M, C,$ or MC), $\bar{d}_{A/M/C} = \bar{d}_{AA/M/C} = -\bar{d}_{aa/M/C} = -0.084,$ and $\bar{D}_{A/C} = 0$. Solid boxes give estimates while open boxes give the 95% confidence limits. The cytonuclear system with maternal inheritance (n-mtDNA) cannot estimate M for this case; details are given in text.

systems give good estimates for the seed migration rate ($\Delta(m) \leq 0.335$), and both n-mtDNA and n-mtDNA-cpDNA give consistently small confidence intervals for this estimate ($0.035 \leq I_m \leq 0.125$). For the pollen migration rate, only the n-mtDNA system with maternal cytoplasmic inheritance and the n-mtDNA-cpDNA system with three forms of inheritance gave good results for $N = 100$ and 500 ($\Delta(M) \leq 0.345$), with the n-mtDNA-cpDNA system giving slightly smaller confidence intervals for all of the estimates. All three systems consistently overestimated the pollen flow rate for the run at $N = 300$.

HYBRID ZONE MODEL

We next consider explicitly the estimation of unidirectional migration rates from multiple sources, such as would be found in a hybrid zone or other areas of admixture. Here, we estimate the migration rates for pollen and seed from each source separately, in contrast to the previous framework that allowed only composite

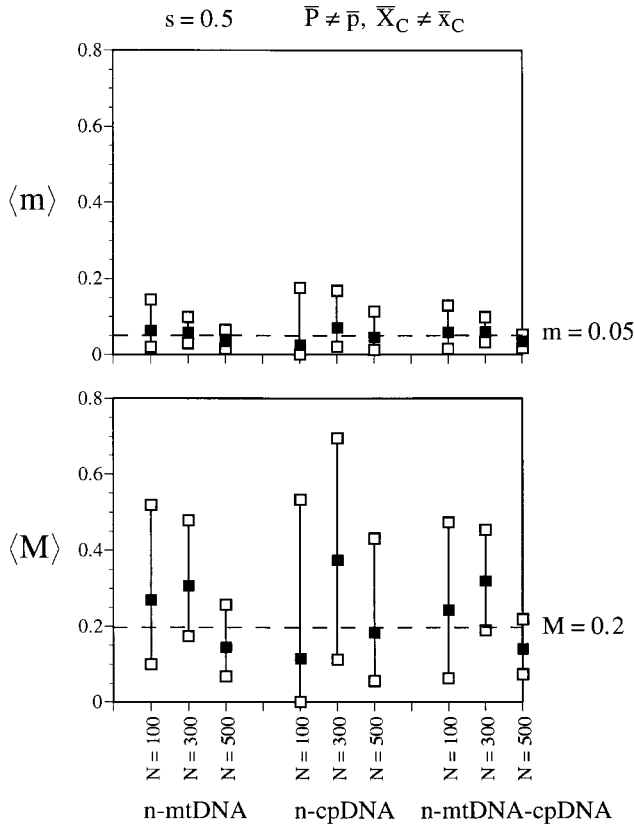


Figure 4.—Estimates of seed migration rate (m) and pollen migration rate (M) where $\bar{P} \neq \bar{p}$, $\bar{X}_C \neq \bar{x}_C$, and $s = 0.5$. Values for migrant pools were $\bar{u}_{11} = 0.7$, $\bar{w}_{22} = 0.3$, and $\bar{Q}_{1C} = 1.0$. Migrant disequilibria are as in Figure 3. Solid boxes give estimates while open boxes give the 95% confidence limits.

estimates for the total amount of each form of gene flow. The migration model for this application is depicted in Figure 5. We assume that each generation a fixed fraction M_1 of outcrossed pollen in the hybrid population is derived from source population 1 (species 1), and a fraction M_2 is derived from source population 2 (species

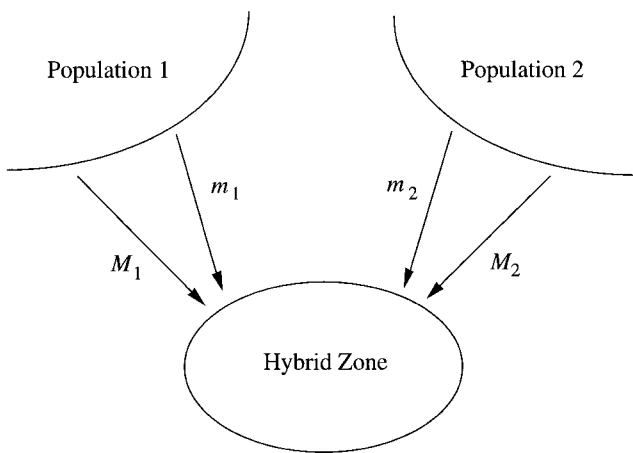


Figure 5.—Migration model with pollen (M_1 , M_2) and seed (m_1 , m_2) migration from two populations into a hybrid zone.

2), where both sources have constant genetic compositions. Similarly, a fixed fraction m_1 of the seeds migrate from source population 1 and a fixed fraction m_2 from source population 2. The total pollen and seed migration rates are then $M = M_1 + M_2$ and $m = m_1 + m_2$, respectively, with the remaining fractions contributed by the resident population. The general model described in Asmussen and Orive (2000) applies here with the migration rates (M , m) now given by these decompositions.

Composition of total migrant pools: Frequencies in each total migrant pool are now the weighted averages of the corresponding frequencies in the two source populations. For example, the nuclear allele frequency in migrant pollen will be

$$\bar{P} = \frac{M_1 \bar{P}^{(1)} + M_2 \bar{P}^{(2)}}{M},$$

where $\bar{P}^{(i)}$ is the nuclear allele frequency in migrant pollen from source population i . The disequilibria in the total migrant pools will be the result of admixture between the contributions of the two sources and can be calculated using (15–18) of the companion article (Asmussen and Orive 2000). For instance, the cytoplasmic disequilibrium in the total migrant seed pool is

$$\bar{d}_{M/C} = \frac{m_1 \bar{d}_{M/C}^{(1)} + m_2 \bar{d}_{M/C}^{(2)}}{m} + \frac{m_1 m_2}{m^2} (\bar{x}_M^{(1)} - \bar{x}_M^{(2)}) (\bar{x}_C^{(1)} - \bar{x}_C^{(2)}),$$

where, once again, the superscript indicates the source population. We can use this new model to jointly estimate the four migration rates M_1 , M_2 , m_1 , m_2 in hybrid zones in much the same way we used the original model to estimate the two (composite) migration rates (M , m).

For our numerical test of the estimation procedure for hybrid zones, we focus on the case where the two source populations show fixed differences at all three loci. This corresponds to having diagnostic nuclear, mitochondrial, and chloroplast markers for the two source populations, as might be found in a hybrid zone where two genetically diverged taxa come into contact. Such a situation is presumably the optimal case for estimation; the utility in other cases can be determined via simulations in the same way as we have done here. We assume that source population 1 is fixed for $AA/M/C$ and source population 2 is fixed for $aa/m/c$. The migration rates from the two sources then uniquely determine the frequencies in migrant seeds,

$$\bar{u}_{11} = \bar{p} = \bar{x}_M = \bar{x}_C = \frac{m_1}{m} \quad \text{and} \quad \bar{w}_{22} = \frac{m_2}{m}, \quad (3)$$

and in migrant pollen (Table 4),

$$\bar{P}_{1C} = \bar{P} = \bar{X}_C = \frac{M_1}{M} \quad \text{and} \quad \bar{Q}_{2C} = \frac{M_2}{M} \quad (4)$$

as well as their disequilibria,

$$\bar{d}_{M/C} = \bar{d}_{A^*} = \bar{d}_{AA^*} = -\bar{d}_{aa^*} = \frac{m_1 m_2}{M^2} \quad (5)$$

$$\bar{d}_{A/M/C} = \bar{d}_{AA/M/C} = -\bar{d}_{aa/M/C} = \frac{m_1 m_2 (m_2 - m_1)}{M^3} \quad (6)$$

$$\bar{d}_{Aa^*} = \bar{d}_{Aa/M/C} = 0 \quad (7)$$

$$\bar{D}_{A/C} = \frac{M_1 M_2}{M^2}, \quad (8)$$

where * indicates *M*, *C*, or *MC*. Note that this shows that the signs of the three-way disequilibria can serve as useful indicators of asymmetry in seed migration from two genetically distinct sources, since (6) will be positive if $m_2 > m_1$, negative if $m_1 > m_2$, and zero only when $m_1 = m_2$.

Results from simulated data: Several different combinations of migration and selfing rates (M_1 , M_2 , m_1 , m_2 , s) were tested, selected from the same range used in the general case. The numerical results are given in appendix d. As for the original model, due to the many possible combinations of parameter values, the examples given in appendix d are not meant to be exhaustive but are illustrative of some of the factors that impact this estimation method. The cytonuclear system with a maternally inherited cytoplasmic marker (n-mtDNA) consistently had trouble estimating pollen migration rates, generally giving the largest confidence intervals for M_1 and M_2 and often having very poor estimates as well. With equal rates of pollen and seed migration ($m_1 = m_2 = M_1 = M_2 = 0.1$), the cytonuclear system with a paternally inherited cytoplasmic marker (n-cpDNA) was also often poor at estimating the pollen migration rate, and, for low or intermediate selfing rates, had larger confidence intervals for estimates of seed migration rates than either the n-mtDNA or n-mtDNA-cpDNA systems. Note that, for this case, the migrant pools are equivalent, with equal allele frequencies for the nuclear and paternally inherited cytoplasmic markers ($\bar{P} = \bar{p} = 0.5$, $\bar{X}_C = \bar{x}_C = 0.5$) so that there are no simple intermigrant effects.

In the examples where only one source population contributes migrant pollen ($M_1 = 0.0$), the n-cpDNA system with paternal cytoplasmic inheritance tended to estimate the seed migration rate from the other source population as zero instead of the true value of 0.01 ($\langle m_2 \rangle = 0.0$, $\Delta(m_2) = 1.0$). This may be because, with diagnostic markers and only one source of migrant pollen, the migrant pollen pool is fixed at both the nuclear and paternally inherited loci ($\bar{P} = \bar{X}_C = 0$) and has no allelic cytonuclear disequilibrium ($\bar{D}_{A/C} = 0$). In contrast, the three-locus system with both modes of cytoplasmic inheritance was usually quite successful at estimating all four migration parameters, and, in cases where one of the two-locus systems failed to estimate migration rates ($\Delta(z) > 0.5$), the n-mtDNA-cpDNA system generally succeeded.

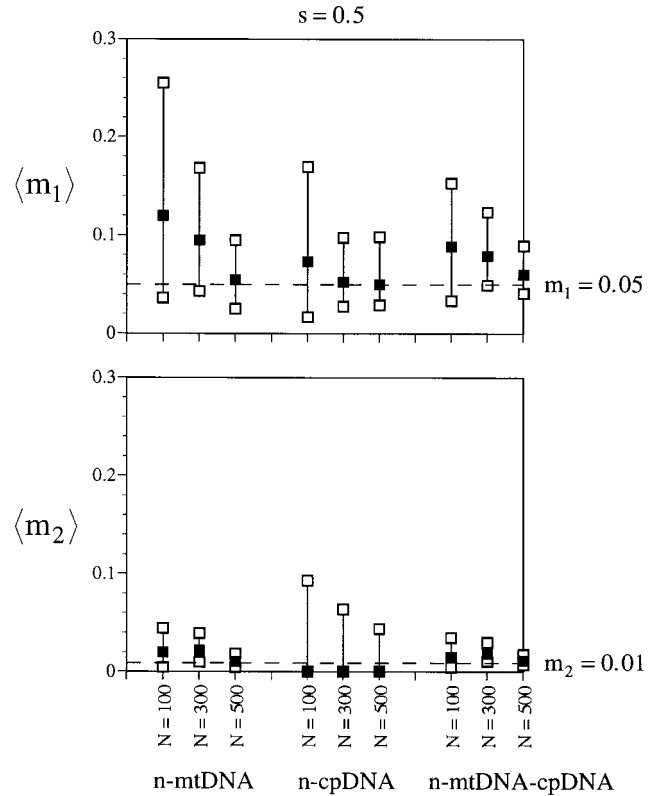


Figure 6.—Estimates of the two seed migration rates in a hybrid zone when the two seed and two pollen migration rates differ ($m_1 = 0.05$, $m_2 = 0.01$, $M_1 = 0.1$, $M_2 = 0.2$) and where $\bar{P} \neq \bar{p}$, $\bar{X}_C \neq \bar{x}_C$, and $s = 0.5$. Each source population was fixed for diagnostic markers, so the migration rates determined the migrant composition; see Equations 3–8 in text. Solid boxes give estimates while open boxes give the 95% confidence limits.

With smaller sample sizes ($N = 100$), there were occasionally runs where none of the three systems could estimate a parameter. For example, for $m_1 = 0$, $m_2 = 0.05$, $M_1 = 0.1$, $M_2 = 0$, $s = 0.9$, and $N = 100$, by chance the simulated data set did not include any heterozygous individuals ($N_{Aa/m/c} = N_{Aa/m/c} = 0$) although the equilibrium heterozygote frequency was $\hat{V} = 0.0388$ for that parameter set. All three systems estimated the seed migration rate to be zero in this case ($\langle m_2 \rangle = 0.0$, $\Delta(m_2) = 1.0$). This is indicative of the problem that all iterative maximum-likelihood methods encounter when there are missing observations, as can occur with small sample sizes (Weir 1996). Paralleling the original model, there was a tendency for high selfing rates to decrease the accuracy of estimates for pollen migration rates.

Figures 6–9 give examples from two particular test runs. As before, the estimates of the seed migration rates (Figures 6 and 8, $\langle m_1 \rangle$ and $\langle m_2 \rangle$) and the pollen migration rates (Figures 7 and 9, $\langle M_1 \rangle$ and $\langle M_2 \rangle$) for the three systems (n-mtDNA, n-cpDNA, and n-mtDNA-cpDNA) are given for the three sample sizes: $N = 100$, 300, and 500. Again, the dashed lines give the actual

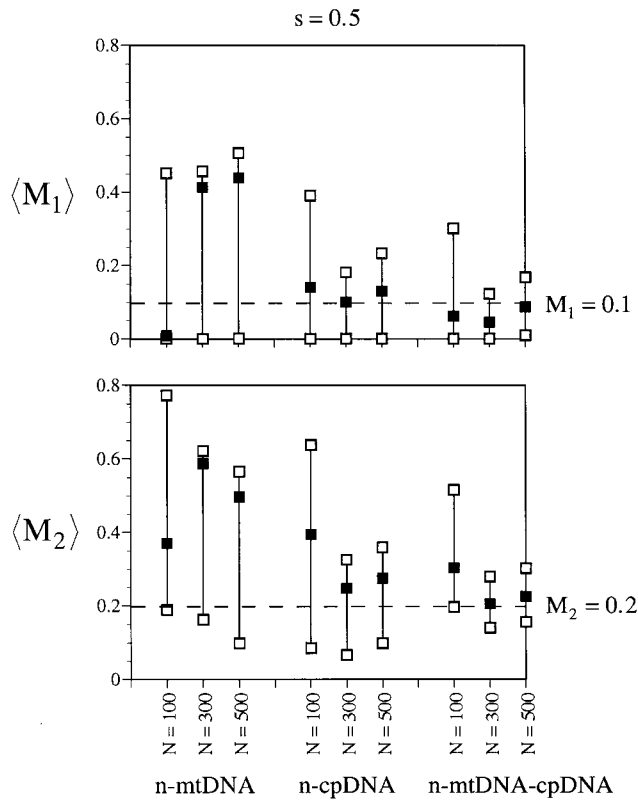


Figure 7.—Estimates of the two pollen migration rates in a hybrid zone when the two seed and two pollen migration rates differ ($m_1 = 0.05$, $m_2 = 0.01$, $M_1 = 0.1$, $M_2 = 0.2$) and where $\bar{P} \neq \bar{p}$, $\bar{X}_C \neq \bar{x}_C$, and $s = 0.5$, as in Figure 6. Each source population was fixed for diagnostic markers, so the migration rates determined the migrant composition; see Equations 3–8 in text. Solid boxes give estimates while open boxes give the 95% confidence limits. Note the difference in scale between this figure and Figure 6, which gives estimates for the two seed migration rates.

values for each of the migration rates, the solid boxes give the estimates, and the open boxes indicate the upper and lower bounds for the 95% confidence limits.

In the example given by Figures 6 and 7, all four migration rates differed ($m_1 = 0.05$, $m_2 = 0.01$, $M_1 = 0.1$, $M_2 = 0.2$) and the selfing rate was $s = 0.5$. Increasing sample size generally decreased the size of the confidence intervals for all three systems, except for the pollen migration rates for $N = 500$. The three-locus system (n-mtDNA-cpDNA) gave consistently smaller confidence intervals for pollen migration rates than the other two systems. The n-cpDNA system with paternal inheritance gave very poor estimates for the smaller of the seed migration rates (Figure 6, m_2), while the n-mtDNA system with maternal inheritance always performed poorly in estimating pollen migration (Figure 7). In contrast, the n-mtDNA-cpDNA system gives reasonable estimates for all four migration rates.

For the example shown in Figures 8 and 9, all four migration rates were the same ($m_1 = m_2 = M_1 = M_2 = 0.1$) and the selfing rate was $s = 0.1$. Again, increasing

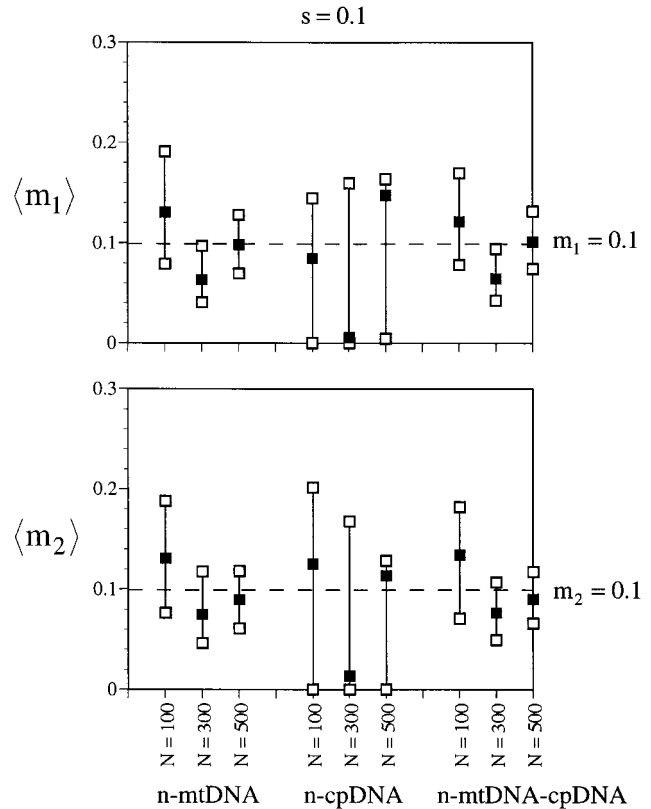


Figure 8.—Estimates of the two seed migration rates in a hybrid zone when the two seed and two pollen migration rates are the same ($m_1 = 0.1$, $m_2 = 0.1$, $M_1 = 0.1$, $M_2 = 0.1$) and where $\bar{P} = \bar{p}$, $\bar{X}_C = \bar{x}_C$, and $s = 0.1$. Each source population was fixed for diagnostic markers, so the migration rates determined the migrant composition; see Equations 3–8 in text. Solid boxes give estimates while open boxes give the 95% confidence limits.

sample size decreased the confidence intervals for most of the estimates, although the effect was not as great as in the previous example. The cytonuclear system with paternal inheritance (n-cpDNA) gave large confidence intervals and often poor estimates for all four migration rates. The cytonuclear system with maternal inheritance (n-mtDNA) again gave very poor estimates for the pollen migration rates (Figure 9). All three systems underestimated the seed migration rates and overestimated the pollen migration rates for the $N = 300$ run, indicating that a particular data set may poorly reflect the “true” genotypic frequencies for a population, leading to inaccurate estimates. Once again, the n-mtDNA-cpDNA system performed best overall.

DISCUSSION

The juxtaposition of biparental and uniparental inheritance in the same individual makes joint cytonuclear data particularly useful for decomposing plant gene flow and estimating the pollen (haploid) and seed (diploid)

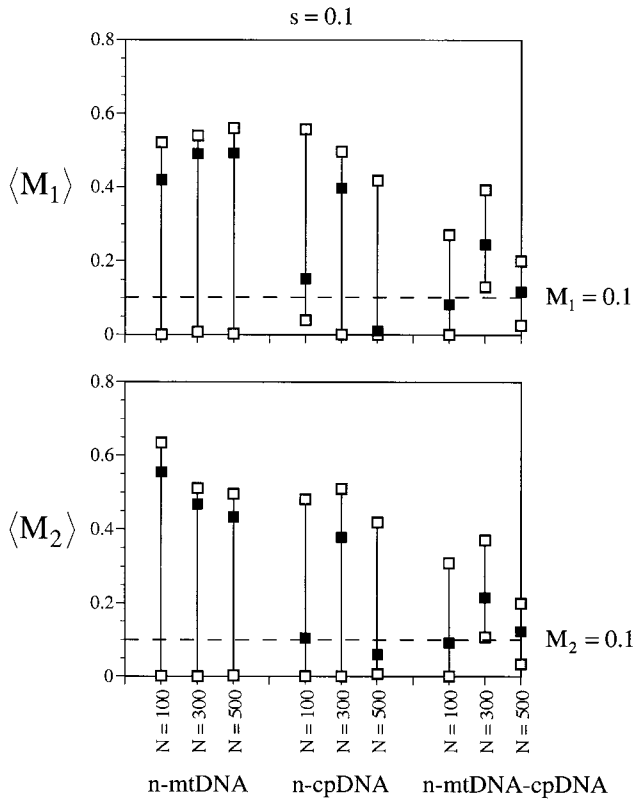


Figure 9.—Estimates of the two pollen migration rates in a hybrid zone when the two seed and two pollen migration rates are the same ($m_1 = 0.1$, $m_2 = 0.1$, $M_1 = 0.1$, $M_2 = 0.1$) and where $\bar{P} = \bar{p}$, $\bar{X}_c = \bar{x}_c$, and $s = 0.1$, as in Figure 8. Each source population was fixed for diagnostic markers, so the migration rates determined the migrant composition; see Equations 3–8 in text. Solid boxes give estimates while open boxes give the 95% confidence limits. Note the difference in scale between this figure and Figure 8, which gives estimates for the two seed migration rates.

components. Here we have used previously developed continent-island models of unidirectional migration (Asmussen and Schnabel 1991; Schnabel and Asmussen 1992; Asmussen and Orive 2000) to develop a new method for estimating both forms of plant gene flow in mixed-mating populations. A program implementing this approach is available from the authors by request. This procedure can utilize joint two-locus cytonuclear data, with either maternal or paternal cytoplasmic inheritance, as well as joint nuclear-mitochondrial-chloroplast data with both forms of cytoplasmic inheritance, for any number of unlinked nuclear markers. The underlying theory is based on diallelic loci, with multiallelic data accommodated in this initial implementation by grouping one user-specified allele *vs.* all others. Three-locus data from species where both organelles are transmitted through the same parent have not been considered since these are equivalent to a two-locus cytonuclear system with a multiallelic cytoplasmic marker. Such three-locus systems have estimation benefits over data from either two-locus cytonuclear system, but presum-

ably only from the increased degrees of freedom in the data. Biparental inheritance of cytoplasmic markers has been reported for several species of plants (Metzlaff *et al.* 1981; Medgyesy *et al.* 1986), but we have not included it here since such data lack the special utility of uniparentally inherited cytoplasmic markers.

We have focused here on censusing adults, both for convenience and since, in general, assaying three markers from adult tissues will be easier than doing so from seeds, especially for species whose seeds are small. However, seeds from conifers may be particularly easy to assay, especially for nuclear allozymes (Conkle 1971). The method presented here can be readily extended to populations censused at the seed stage, by converting the equilibrium genotype frequencies for adults (\hat{Z}) in the log-likelihood equations (1) and (2) to those for seeds (\hat{z}). The corresponding seed equilibria are given by

$$\hat{z} = \frac{\hat{Z} - m\bar{z}}{1 - m},$$

where \bar{z} is the corresponding frequency in migrant seeds and m is the overall rate of seed migration. Note that this conversion follows directly from the equilibrium relationship

$$\hat{Z} = m\bar{z} + (1 - m)z,$$

which applies to all frequency variables in the two life stages (see Equation 22, Asmussen and Orive 2000). Researchers interested in a version of the estimation program for seed-censused populations should contact the authors.

The estimation procedure developed in this article provides joint estimates of both pollen and seed migration rates, as well as selfing rates, from observed counts of either joint cytonuclear or joint dicytonuclear genotypes. In each case, estimation is based on fitting the expected genotype frequencies under the appropriate pollen and seed dispersal model (Asmussen and Orive 2000) to the data. This requires that the equilibrium state for the system depend on the parameter(s) in question and that the censused population be at equilibrium, with a constant dicytonuclear and/or cytonuclear structure over successive generations. Ideally, the latter condition would be verified by assaying the target population in at least two time points separated by one or more generations. Further, basic estimation constraints may be added by the need for a sufficient number of degrees of freedom in the data. In general, two-locus cytonuclear data from a single nuclear marker allow estimation of up to five parameters, or four if a degree of freedom is desired to test for fit between observed and expected genotypic frequencies; three-locus dicytonuclear data can allow the estimation of six additional parameters. The basic estimation power of both types of systems can be substantially increased by assaying multiple unlinked nuclear markers.

Consideration of the equilibrium structure for the two types of cytonuclear systems and the full dicytonuclear system shows that it is harder to estimate pollen flow, which carries only the nuclear and paternally transmitted cytoplasmic marker, than seed flow, which carries, and is therefore reflected by, all three types of markers. Although all conditions allowing estimates of the pollen migration rate also allow estimates of the seed migration rate, the reverse is not true. For example, with cytonuclear data for a maternally inherited cytoplasmic marker, equal nuclear allele frequencies in migrant pollen and seeds ($\bar{P} = \bar{p}$) make it impossible to estimate the pollen migration rate, while the rate of seed migration can still be estimated if there is nuclear polymorphism in the migrant seeds ($0 < \bar{p} < 1$). Further evidence of the greater constraints on estimating pollen flow is found with three-locus data with both forms of uniparental cytoplasmic inheritance; these increase the conditions for estimating the seed migration over those in the two-locus systems, but do not add to the conditions under which pollen migration can be estimated.

When considering the range of conditions under which both migration rates can be estimated, cytonuclear data with a paternally inherited cytoplasmic marker are better than those with a maternally inherited marker, while the three-locus n-mtDNA-cpDNA system with three modes of inheritance is the best of all. The use of three-locus data increases the conditions for estimating the seed migration rate over those for a cytonuclear system with only paternal cytoplasmic inheritance, but both systems have the same conditions for estimating the pollen migration rate, and both systems can estimate both migration rates under a wider set of conditions than can a cytonuclear system with maternal inheritance. Finally, examination of the final cytonuclear structure for each of the three systems shows that the rates of pollen and seed migration enter into more terms that determine the dicytonuclear system, indicating that the use of these types of three-locus data should increase the power to detect these two forms of plant gene flow and/or give more accurate estimates.

Comparisons of estimates from simulated data for cytonuclear systems with single nuclear markers and either maternal or paternal cytoplasmic inheritance to estimates from simulated joint nuclear-mtDNA-cpDNA data show that the three-locus system with three modes of inheritance is generally the most useful for estimating both seed and pollen migration rates. There are several cases where one or the other of the two-locus cytonuclear systems cannot be used to estimate a migration rate while the three-locus, nuclear-dicytoplasmic system succeeds, generally making it more reliable for estimating the two types of gene flow. For example, as indicated above, a cytonuclear system with maternal inheritance cannot be used to estimate pollen migration rates when nuclear allele frequencies are equal in both migrant pools ($\bar{P} = \bar{p}$) or when there is no seed migration ($m =$

0). The utility of cytonuclear systems with paternal inheritance and nuclear-mtDNA-cpDNA systems with both maternal and paternal inheritance is also somewhat reduced in populations without seed migration ($m = 0$), for these require allelic cytonuclear disequilibrium in migrant pollen ($\bar{D}_{A/C} \neq 0$) in order to estimate pollen migration rates if this is the only form of gene flow.

Similar results apply to the case where two populations, fixed for alternate alleles at all three markers, contribute migrants to a hybrid population. Here there can be up to four separate migration rates. Note that this means that either of the two-locus cytonuclear systems can be used to estimate all four migration parameters plus the selfing rate, but (with only one nuclear marker) they could not be used to test the fit of these estimates, since they would have at most five independent data classes. Estimates for the two pollen migration rates from cytonuclear data with maternal inheritance were generally poor in our numerical examples, and they also had large confidence intervals. Cytonuclear data with paternal inheritance are better, since the cytoplasmic marker is then carried by pollen as well as seeds, but such data estimated all four migration rates poorly when all of the rates were equal. For the examples considered, only the full nuclear-mtDNA-cpDNA data set consistently allowed accurate estimates of all four migration rates.

Comparing estimates of genetic subdivision for organelle genes (Birky *et al.* 1983, 1989; Takahata and Palumbi 1985) with estimates for nuclear genes gives another approach for using differences in inheritance patterns (uniparental *vs.* biparental) to infer types of gene flow. Previous theoretical treatments either have used measures such as G_{ST} solely to provide estimates of the product of effective population size and migration rate (effective migration rate, Nm) or have calculated the relative rates of pollen and seed migration by comparing such estimates for markers with contrasting modes of inheritance (Petit *et al.* 1993; Ennos 1994; Hu and Ennos 1999). Such methods also depend on the assumption of genetic equilibrium in the population under consideration, in this case that between genetic drift and migration. In addition, these calculations of the relative rates of the two forms of plant gene flow require the assumption of low pollen migration rates when used with paternally inherited markers, such as cpDNA in conifers (Ennos 1994); this assumption may be commonly violated in such wind-pollinated species (Hamrick *et al.* 1995; El Mousadik and Petit 1996). The increasing availability of joint cytonuclear data sets for conifer species (Dong and Wagner 1994; Latta and Mitton 1997) will allow further comparisons between these methods and the method developed here for estimating pollen- and seed-mediated gene flow.

Other systems with separate maternal and paternal uniparental inheritance may also provide opportunities

for partitioning gene flow via female gametes *vs.* that via male gametes, or gamete-level gene flow from zygote-level gene flow. One intriguing possibility is found in mussels of the genus *Mytilus*, where mitochondrial DNA lineages appear to be gender-specific (Zouros *et al.* 1992, 1994). Although we have framed this current work in the context of plant populations (with pollen and/or seed migration and levels of selfing), the results presented here indicate that a system such as that found in *Mytilus* may show gender-specific cytonuclear associations useful for studies of sexual asymmetries in gene flow. Moreover, ongoing theoretical work indicates that cytonuclear data with sex-linked nuclear markers have great value for detecting and estimating gender-specific dispersal in animals (Goodisman and Asmussen 1997; Goodisman *et al.* 1998), another aspect of gene flow that has been difficult to study by traditional means.

In conclusion, three-locus, nuclear-dicytoplasmic data from plants such as conifers with both paternally and maternally inherited organelles represent a unique opportunity to more accurately estimate both the absolute and relative rates of pollen- and seed-mediated gene flow. Analyses of joint genotype frequencies and non-random associations among three loci with three distinct modes of inheritance allow a better understanding of the relative roles that the mating system and gene flow play in shaping the genetic structure of such populations than has previously been possible. This information can in turn be used to develop, as we have done here, methods that allow the use of such three-locus data in powerful new ways.

We thank Chris Babcock, Mike Goodisman, Deb Overath, María Sánchez, Janet Hatt, and two anonymous reviewers for their helpful comments and suggestions on this work. Part of this work was completed while M.E.O. was a fellow at the Institute of Cell, Animal and Population Biology, University of Edinburgh, and we thank Nick Barton and his lab for their assistance. This research was supported in part by National Institutes of Health grant GM-48528 and National Science Foundation grant DEB 92-10895 to M.A.A. and by a National Science Foundation NATO postdoctoral fellowship and a University of Kansas New Faculty General Research Fund grant to M.E.O.

LITERATURE CITED

- Asmussen, M. A., and M. E. Orive, 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., and A. Schnabel, 1991 Comparative effects of pollen and seed migration on the cytonuclear structure of plant populations. I. Maternal cytoplasmic inheritance. *Genetics* **128**: 639–654.
- Asmussen, M. A., J. Arnold and J. C. Avise, 1989 The effects of assortative mating and migration on cytonuclear associations in hybrid zones. *Genetics* **122**: 923–934.
- Avise, J. C., W. S. Nelson, J. Arnold, R. K. Koehn, G. C. Williams *et al.*, 1990 The evolutionary genetic status of icelandic eels. *Evolution* **44**: 1254–1262.
- Barton, N. H., and M. Turelli, 1991 Natural and sexual selection on many loci. *Genetics* **127**: 229–255.
- Birky, C. W., 1988 Evolution and variation in plant chloroplast and mitochondrial genomes, pp. 23–53 in *Plant Evolutionary Biology*, edited by L. D. Gottlieb and S. K. Jain. Chapman & Hall, London.
- Birky, C. W., T. Maruyama and P. Fuerst, 1983 An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* **103**: 513–527.
- Birky, C. W., P. Fuerst and T. Maruyama, 1989 Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics* **121**: 613–627.
- Bohlenz, K., T. Nothnagel and M. Metzlauff, 1990 Paternal inheritance of plastids in the genus *Daucus*. *Mol. Gen. Genet.* **220**: 489–491.
- Clegg, M. T., 1980 Measuring plant mating systems. *Bioscience* **30**: 814–818.
- Conkle, M. T., 1971 Isozyme specificity during germination and early growth of knobcone pine. *For. Sci.* **17**: 494–498.
- Dong, J., and D. B. Wagner, 1994 Paternally inherited chloroplast polymorphism in *Pinus*: estimation of diversity and population subdivision, and tests of disequilibrium with a maternally inherited mitochondrial polymorphism. *Genetics* **136**: 1187–1194.
- Edwards, A. W. F., 1992 *Likelihood*. Johns Hopkins University Press, Baltimore.
- El Mousadik, A., and R. J. Petit, 1996 Chloroplast DNA phylogeography of the argan tree of Morocco. *Mol. Ecol.* **5**: 547–555.
- Ennos, R. A., 1994 Estimating the relative rates of pollen and seed migration among plant populations. *Heredity* **72**: 250–259.
- Goodisman, M. A. D., and M. A. Asmussen, 1997 Cytonuclear theory for haplodiploid species and X-linked genes. I. Hardy-Weinberg dynamics and continent-island, hybrid zone models. *Genetics* **147**: 321–338.
- Goodisman, M. A. D., D. D. Shoemaker and M. A. Asmussen, 1998 Cytonuclear theory for haplodiploid species and X-linked genes. II. Stepping-stone models of gene flow and application to a fire ant hybrid zone. *Evolution* **52**: 1423–1440.
- Hamrick, J. L., M. J. W. Godt and S. L. Sherman-Broyles, 1995 Gene flow among plant populations: evidence from genetic markers, pp. 215–232 in *Experimental and Molecular Approaches to Plant Biosystematics*, edited by P. C. Hoch and A. G. Stephenson. Missouri Botanical Garden Press, St. Louis.
- Harrison, R. G., and J. J. Doyle, 1990 Redwoods break the rules. *Nature* **344**: 295–296.
- Hu, X.-S., and R. A. Ennos, 1999 Impacts of seed and pollen flow on population genetic structure for plant genomes with three contrasting modes of inheritance. *Genetics* **152**: 441–450.
- Kirk, J. T. O., and R. A. E. Tilney-Bassett, 1978 *The Plastids: Their Chemistry, Structure, Growth and Inheritance*. Elsevier/North-Holland, Amsterdam.
- Latta, R. G., and J. B. Mitton, 1997 A comparison of population differentiation across four classes of gene marker in limber pine (*Pinus flexilis* James). *Genetics* **146**: 1153–1163.
- Medgyesy, P., A. Pay and L. Marton, 1986 Transmission of paternal chloroplasts in *Nicotiana*. *Mol. Gen. Genet.* **204**: 195–198.
- Metzlauff, M., T. Borner and R. Hagemann, 1981 Variations of chloroplast DNAs in the genus *Pelargonium* and their biparental inheritance. *Theor. Appl. Genet.* **60**: 37–41.
- Mogensen, H. L., 1996 The hows and whys of cytoplasmic inheritance in seed plants. *Am. J. Bot.* **83**: 383–404.
- Neale, D. B., and R. R. Sederoff, 1989 Paternal inheritance of chloroplast DNA and maternal inheritance of mitochondrial DNA in loblolly pine. *Theor. Appl. Genet.* **77**: 212–216.
- Neale, D. B., N. C. Wheeler and R. W. Allard, 1986 Paternal inheritance of chloroplast DNA in Douglas-fir. *Can. J. For. Res.* **16**: 1152–1154.
- Paige, K. N., W. C. Capman and P. Jennetten, 1991 Mitochondrial inheritance patterns across a cottonwood hybrid zone: cytonuclear disequilibria and hybrid zone dynamics. *Evolution* **45**: 1360–1369.
- Petit, R. J., A. Kremer and D. B. Wagner, 1993 Finite island model for organelle and nuclear genes in plants. *Heredity* **71**: 630–641.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling and B. F. Flannery, 1992 *Numerical Recipes in C*. Cambridge University Press, Cambridge.
- Schnabel, A., and M. A. Asmussen, 1992 Comparative effects of pollen and seed migration on the cytonuclear structure of plant

populations. II. Paternal cytoplasmic inheritance. *Genetics* **132**: 253–267.

Sites, J. W. J., C. J. Basten and M. A. Asmussen, 1996 Cytonuclear genetic structure of a hybrid zone in lizards of the *Sceloporus grammicus* complex (Saurai, Phrynosomatidae). *Mol. Ecol.* **5**: 379–392.

Takahata, N., and S. R. Palumbi, 1985 Extranuclear differentiation and gene flow in the finite island model. *Genetics* **109**: 441–457.

Turelli, M., and N. H. Barton, 1994 Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? *Genetics* **138**: 913–941.

Wagner, D. B., G. R. Furnier, M. A. Saghai-Marroof, S. M. Williams, B. P. Dancik *et al.*, 1987 Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proc. Natl. Acad. Sci. USA* **84**: 2097–2100.

Weir, B. S., 1996 *Genetic Data Analysis II*. Sinauer, Sunderland, MA.

Zouros, E., K. R. Freeman, A. Oberhauser Ball and G. H. Pogson, 1992 Direct evidence for extensive paternal inheritance in the marine mussel *Mytilus*. *Nature* **359**: 412–414.

Zouros, E., A. Oberhauser Ball, C. Saavedra and K. R. Freeman, 1994 An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. *Proc. Natl. Acad. Sci. USA* **91**: 7463–7467.

Communicating editor: A. G. Clark

APPENDIX A: EQUILIBRIA FOR FREQUENCIES AND DISEQUILIBRIA

A set of five independent variables for the equilibrium of the n-mtDNA system with maternal inheritance includes the three final nuclear and mitochondrial frequencies,

$$\hat{p} = \frac{2m\bar{p} + (1-m)M(1-s)\bar{P}}{2m + (1-m)M(1-s)} \quad (\text{A1})$$

$$\hat{v} = \frac{2m\bar{v} + 2(1-m)(1-s)[M(\bar{P}\bar{Q} + \bar{Q}\bar{P}) + 2(1-M)\bar{P}\bar{Q}]}{2 - (1-m)s} \quad (\text{A2})$$

$$\hat{X}_M = \begin{cases} \bar{X}_M & \text{if } m > 0 \\ X_M^0 & \text{if } m = 0 \end{cases} \quad (\text{A3})$$

(where X_M^0 is the initial mitochondrial frequency in the resident population), and the two final n-mtDNA cytonuclear disequilibria,

$$\hat{D}_{A/M} = \frac{2m\bar{d}_{A/M}}{2 - (1-m)(1+s)} \quad (\text{A4})$$

$$\hat{D}_{Aa/M} = \frac{2m\bar{d}_{Aa/M} + 2(1-m)(1-s)[M(\bar{Q} - \bar{P}) + (1-M)(\hat{Q} - \hat{P})]\bar{D}_{A/M}}{2 - (1-m)s} \quad (\text{A5})$$

The corresponding set of variables for the n-cpDNA system with paternal inheritance includes the nuclear frequencies \hat{P} and \hat{V} in (A1) and (A2), the equilibrium chloroplast frequency,

$$\hat{X}_c = \frac{m\bar{x}_c + (1-m)M(1-s)\bar{X}_c}{m + (1-m)M(1-s)}, \quad (\text{A6})$$

and the two final n-cpDNA disequilibria,

$$\hat{D}_{A/C} = \frac{2m\bar{d}_{A/C} + (1-m)M(1-s)\bar{D}_{A/C} + K_1(\bar{P} - \bar{p})(\bar{X}_c - \bar{x}_c)}{2 - (1-m)[1+s - M(1-s)]} \quad (\text{A7})$$

$$\begin{aligned} K_2\hat{D}_{Aa/C} &= 2m\bar{d}_{Aa/C} + 2(1-m)(1-s)(\hat{Q} - \hat{P}) \\ &\times [M\bar{D}_{A/C} + (1-M)\hat{D}_{A/C}] \\ &+ \frac{2m(1-m)M(1-s)}{m + (1-m)M(1-s)} (\hat{P}\bar{Q} + \hat{Q}\bar{P} - \bar{v})(\bar{X}_c - \bar{x}_c), \quad (\text{A8}) \end{aligned}$$

where

$$K_1 = \frac{2m(1-m)M(1-s)}{2m + (1-m)M(1-s)}$$

and $K_2 = 2 - (1-m)s$.

A sufficient set of 11 variables for the full, three-locus system includes the 8 two-locus variables in (A1–A8) and the final cytoplasmic disequilibrium,

$$\hat{D}_{M/C} = \frac{m\bar{d}_{M/C}}{1 - (1-m)s} \quad (\text{A9})$$

together with either the two final, three-locus joint disequilibria,

$$\begin{aligned} \hat{D}_{A/MC} &= c_1\bar{d}_{A/MC} + c_2\bar{d}_{A/M} + c_3\bar{d}_{A/C} + c_4\bar{D}_{A/C} \\ &+ c_5(\bar{P} - \bar{p})\bar{d}_{M/C} + c_6(\bar{P} - \bar{p})(\bar{X}_c - \bar{x}_c) \quad (\text{A10}) \end{aligned}$$

$$\begin{aligned} \hat{D}_{Aa/MC} &= c_1\bar{d}_{Aa/MC} + c_2\bar{d}_{A/M} + c_3\bar{d}_{A/C} + c_4\bar{D}_{A/C} + c_5\bar{d}_{A/M}\bar{d}_{A/C} \\ &+ c_6\bar{d}_{A/M}\bar{D}_{A/C} + c_7\bar{d}_{M/C} + c_8(\bar{X}_c - \bar{x}_c) \\ &+ c_9(\bar{P} - \bar{p})(\bar{X}_c - \bar{x}_c) + c_{10}(\bar{P} - \bar{p})(\bar{X}_c - \bar{x}_c)\bar{d}_{A/M} \quad (\text{A11}) \end{aligned}$$

or the two final three-way disequilibria,

$$\hat{D}_{A/M/C} = c_1\bar{d}_{A/M/C} + c_2(\bar{P} - \bar{p})\bar{d}_{M/C} + c_3(\bar{X}_c - \bar{x}_c)\bar{d}_{A/M} \quad (\text{A12})$$

$$\begin{aligned} \hat{D}_{Aa/M/C} &= c_1\bar{d}_{Aa/M/C} + c_2\bar{d}_{M/C} + c_3\bar{d}_{A/M}\bar{d}_{A/C} + c_4\bar{d}_{A/M}\bar{D}_{A/C} \\ &+ c_5(\bar{X}_c - \bar{x}_c)\bar{d}_{A/M} + c_6(\bar{P} - \bar{p})(\bar{X}_c - \bar{x}_c)\bar{d}_{A/M} \\ &+ c_7(\bar{X}_c - \bar{x}_c)\bar{d}_{Aa/M}. \quad (\text{A13}) \end{aligned}$$

The c_i coefficients for the three-locus disequilibria in (A10–A13) are complicated functions of the migrant values and the migration and mating system parameters, whose values are given by (B1–B4) in appendix b. Details for the derivation of (A1–A13) are given in Asmussen and Schnabel (1991), Schnabel and Asmussen (1992), and Asmussen and Orive (2000).

APPENDIX B: EQUILIBRIA FOR THE THREE-LOCUS DISEQUILIBRIA

We give here the details of the full forms for the final three-locus disequilibria used for the decomposition of the three-locus system's equilibrium state. These are given in terms of the six constant factors,

$$k_1 = 1 - (1-m)s$$

$$k_2 = 2 - (1-m)(1+s)$$

$$\begin{aligned}
k_3 &= m + (1 - m)M(1 - s) \\
k_4 &= 2 - (1 - m)[1 + s - M(1 - s)] \\
k_5 &= 2m + (1 - m)M(1 - s) \\
k_6 &= 2 - (1 - m)s,
\end{aligned}$$

which are functions of the migration and mating system parameters (m , M , and s). Using these, the constant coefficients of the various terms for the three-locus joint allelic disequilibrium, $\hat{D}_{A/M/G}$, in (A10) are

$$\begin{aligned}
c_1 &= \frac{m}{k_1} \\
c_2 &= \frac{m(1 - m)(1 - s)\{m(1 - M)\bar{x}_C + M[1 - (1 - m)s]\bar{x}_C\}}{k_1 k_2 k_3} \\
c_3 &= \frac{m(1 - m)(1 - M)(1 - s)\bar{x}_M}{k_1 k_4} \\
c_4 &= \frac{(1 - m)M(1 - s)\bar{x}_M}{k_4} \\
c_5 &= -\frac{m(1 - m)M(1 - s)}{k_1 k_5} \\
c_6 &= \frac{2m(1 - m)M(1 - s)\bar{x}_M}{k_4 k_5}. \tag{B1}
\end{aligned}$$

For the heterozygote three-locus joint genotypic disequilibrium, $\hat{D}_{Aa/MG}$, the constants in (A11) are

$$\begin{aligned}
c_1 &= \frac{2m}{k_6} \\
c_2 &= \frac{4m(1 - m)(1 - s)[M(\bar{Q} - \bar{P})\bar{x}_C + (1 - M)(\hat{Q} - \hat{P})\bar{x}_C]}{k_2 k_6} \\
c_3 &= \frac{4m(1 - m)(1 - M)(1 - s)[(\hat{Q} - \hat{P})\bar{x}_M]}{k_4 k_6} \\
c_4 &= \frac{4(1 - m)M(1 - s)k_1(\hat{Q} - \hat{P})\bar{x}_M}{k_4 k_6} \\
c_5 &= -\frac{16m^2(1 - m)(1 - M)(1 - s)}{k_2 k_4 k_6} \\
c_6 &= -\frac{16m(1 - m)M(1 - s)k_1}{k_2 k_4 k_6}
\end{aligned}$$

$$\begin{aligned}
c_7 &= \frac{2m(1 - m)(1 - s)[\bar{v} - MK - 2(1 - M)\hat{P}\hat{Q}]}{k_1 k_6} \\
c_8 &= \frac{2m(1 - m)M(1 - s)(K - \bar{v})\bar{x}_M}{k_3 k_6} \\
c_9 &= \frac{4m(1 - m)^2 M(1 - M)(1 - s)^2(\hat{Q} - \hat{P})\bar{x}_M}{k_4 k_5 k_6} \\
c_{10} &= -\frac{16m^2(1 - m)^2 M(1 - M)(1 - s)^2}{k_2 k_4 k_5 k_6}, \tag{B2}
\end{aligned}$$

where $K = \hat{P}\bar{Q} + \hat{Q}\bar{P}$.

The constants multiplying the various terms for the three-locus, three-way allelic disequilibrium, $\hat{D}_{A/M/G}$, in (A12) are

$$\begin{aligned}
c_1 &= \frac{m}{k_1} \\
c_2 &= -\frac{m(1 - m)M(1 - s)}{k_1 k_5} \\
c_3 &= -\frac{m(1 - m)M(1 - s)}{k_2 k_3}. \tag{B3}
\end{aligned}$$

Finally, the constants for the three-locus, three-way heterozygote disequilibrium, $\hat{D}_{Aa/M/G}$, in (A13) are

$$\begin{aligned}
c_1 &= \frac{2m}{k_6} \\
c_2 &= \frac{2m(1 - m)(1 - s)[\bar{v} - MK - 2(1 - M)\hat{P}\hat{Q}]}{k_1 k_6} \\
c_3 &= -\frac{16m^2(1 - m)(1 - M)(1 - s)}{k_2 k_4 k_6} \\
c_4 &= -\frac{16m(1 - m)M(1 - s)k_1}{k_2 k_4 k_6} \\
c_5 &= \frac{4m^2(1 - m)M(1 - s)(\bar{Q} - \bar{P})}{k_2 k_3 k_6} \\
c_6 &= -\frac{16m^2(1 - m)^2 M(1 - M)(1 - s)^2}{k_2 k_4 k_5 k_6} \\
c_7 &= -\frac{2m(1 - m)M(1 - s)}{k_3 k_6}. \tag{B4}
\end{aligned}$$

APPENDIX C: TABLE C1
Estimates of seed and pollen migration rates from simulated data sets with migrant composition A ($\bar{P} = \bar{p}$, $\bar{X}_C \neq \bar{x}_C$, $\bar{D}_{A/C} = \mathbf{0}$)

m	M	s	System	Δ (m)			Δ (M)		
				N = 100	N = 300	N = 500	N = 100	N = 300	N = 500
0.0	0.05	0.1	n-mt	0.066* (0.148)	0.000* (0.017)	0.008* (0.039)	NA	NA	NA
			n-cp	0.000* (0.000)	0.000* (0.000)	0.000* (0.000)	NA	NA	NA
			n-mt-cp	0.018* (0.094)	0.000* (0.000)	0.000* (0.000)	NA	NA	NA
0.05	0.0	0.1	n-mt	0.143 (0.134)	0.627 (0.060)	0.092 (0.068)	NA	NA	NA
			n-cp	0.907 (0.055)	0.477 (0.065)	0.124 (0.068)	0.002* (0.032)	0.000* (0.004)	0.000* (0.004)
			n-mt-cp	0.428 (0.084)	0.374 (0.050)	0.035 (0.051)	0.010* (0.057)	0.000* (0.007)	0.000* (0.002)
0.05	0.2	0.1	n-mt	0.572 (0.185)	0.452 (0.096)	0.102 (0.069)	NA	NA	NA
			n-cp	1.884 (0.173)	1.190 (0.129)	0.279 (0.063)	2.692 (0.547)	0.854 (0.456)	0.404 (0.187)
			n-mt-cp	0.952 (0.144)	0.827 (0.087)	0.237 (0.045)	1.446 (0.765)	0.646 (0.308)	0.382 (0.147)
		0.5	n-mt	0.679 (0.163)	0.393 (0.057)	0.536 (0.056)	NA	NA	NA
			n-cp	0.218 (0.110)	0.156 (0.072)	0.469 (0.079)	0.086 (0.522)	0.107 (0.253)	0.115 (0.223)
			n-mt-cp	0.184 (0.104)	0.259 (0.048)	0.576 (0.055)	0.211 (0.461)	0.184 (0.191)	0.108 (0.168)
		0.9	n-mt	0.139 (0.092)	0.560 (0.071)	0.388 (0.032)	NA	NA	NA
			n-cp	0.877 (0.229)	0.198 (0.062)	0.252 (0.042)	0.448 (0.799)	0.268 (0.345)	0.365 (0.248)
			n-mt-cp	0.044 (0.083)	0.334 (0.059)	0.328 (0.031)	0.006 (0.404)	0.308 (0.300)	0.205 (0.186)
0.1	0.1	0.1	n-mt	0.240 (0.154)	0.436 (0.120)	0.220 (0.084)	NA	NA	NA
			n-cp	0.296 (0.227)	0.093 (0.099)	0.133 (0.085)	0.208 (0.325)	0.170 (0.141)	0.415 (0.076)
			n-mt-cp	0.087 (0.119)	0.199 (0.083)	0.082 (0.061)	0.107 (0.189)	0.383 (0.144)	0.303 (0.070)
		0.5	n-mt	0.087 (0.182)	0.169 (0.079)	0.226 (0.080)	NA	NA	NA
			n-cp	0.178 (0.159)	0.142 (0.133)	0.158 (0.081)	0.094 (0.336)	0.272 (0.149)	0.185 (0.115)
			n-mt-cp	0.059 (0.127)	0.018 (0.079)	0.006 (0.064)	0.187 (0.292)	0.228 (0.145)	0.409 (0.134)
		0.9	n-mt	0.541 (0.140)	0.355 (0.150)	0.023 (0.094)	NA	NA	NA
			n-cp	0.413 (0.214)	0.447 (0.169)	0.320 (0.050)	0.642 (0.619)	0.404 (0.318)	0.327 (0.170)
			n-mt-cp	0.537 (0.118)	0.318 (0.118)	0.184 (0.064)	0.871 (0.310)	0.255 (0.290)	0.080 (0.215)

Migrant composition A: $\bar{u}_{11} = 0.7$, $\bar{w}_{22} = 0.3$, $\bar{p}_{1C} = 0.7$, $\bar{q}_{1C} = 0.3$; $\Delta(z) = |(\langle z \rangle - z)|$ for each parameter z ; *, $\langle m \rangle$ or $\langle M \rangle$ given instead of Δ (m) or Δ (M); numbers in parentheses are lengths of 95% confidence intervals. mtDNA and cpDNA, maternally inherited and paternally inherited cytoplasmic markers, respectively. NA, not available.

TABLE C2
Estimates of seed and pollen migration rates from simulated data sets with migrant composition B ($\bar{P} = \bar{p}$, $\bar{X}_c \neq \bar{x}_c$, $\bar{D}_{N/c} = \mathbf{0}$)

m	M	s	System	$\Delta(m)$			$\Delta(M)$		
				N = 100	N = 300	N = 500	N = 100	N = 300	N = 500
0.0	0.05	0.1	n-mt	0.000* (0.000)	0.000* (0.000)	0.000* (0.000)	NA	NA	NA
			n-cp	0.000* (0.000)	0.000* (0.000)	0.000* (0.000)	NA	NA	NA
			n-mt-cp	0.000* (0.000)	0.000* (0.000)	0.000* (0.000)	NA	NA	NA
0.05	0.0	0.1	n-mt	1.115 (0.177)	0.581 (0.114)	0.289 (0.077)	0.000* (0.022)	0.000* (0.001)	0.000* (0.003)
			n-cp	2.296 (0.211)	0.314 (0.073)	0.125 (0.071)	0.004* (0.059)	0.000* (0.000)	0.000* (0.003)
			n-mt-cp	1.701 (0.154)	0.300 (0.083)	0.119 (0.047)	0.000* (0.023)	0.000* (0.000)	0.000* (0.002)
0.05	0.2	0.1	n-mt	0.224 (0.135)	0.681 (0.053)	0.438 (0.051)	0.302 (0.609)	0.531 (0.281)	0.365 (0.214)
			n-cp	1.000 (0.125)	1.256 (0.117)	0.137 (0.102)	1.000 (0.450)	1.246 (0.424)	0.133 (0.368)
			n-mt-cp	0.367 (0.102)	0.020 (0.051)	0.317 (0.040)	0.244 (0.463)	0.193 (0.201)	0.264 (0.166)
0.1	0.1	0.1	n-mt	0.252 (0.125)	0.138 (0.069)	0.304 (0.049)	0.345 (0.418)	0.533 (0.306)	0.281 (0.189)
			n-cp	0.506 (0.176)	0.405 (0.147)	0.118 (0.100)	0.432 (0.532)	0.866 (0.583)	0.090 (0.375)
			n-mt-cp	0.139 (0.113)	0.161 (0.065)	0.335 (0.035)	0.212 (0.411)	0.597 (0.264)	0.300 (0.145)
0.1	0.1	0.1	n-mt	1.050 (0.165)	0.171 (0.063)	0.087 (0.046)	0.242 (0.479)	0.274 (0.222)	0.243 (0.161)
			n-cp	0.883 (0.226)	0.456 (0.091)	0.065 (0.053)	0.639 (0.929)	0.111 (0.257)	0.056 (0.203)
			n-mt-cp	0.952 (0.160)	0.212 (0.058)	0.002 (0.037)	0.487 (0.359)	0.088 (0.156)	0.017 (0.128)
0.1	0.1	0.1	n-mt	0.081 (0.185)	0.361 (0.113)	0.387 (0.068)	0.505 (0.133)	0.252 (0.127)	0.455 (0.061)
			n-cp	0.448 (0.197)	0.134 (0.116)	0.129 (0.116)	0.188 (0.165)	0.236 (0.122)	0.163 (0.119)
			n-mt-cp	0.148 (0.153)	0.163 (0.083)	0.305 (0.069)	0.346 (0.109)	0.001 (0.106)	0.318 (0.066)
0.5	0.05	0.1	n-mt	0.296 (0.205)	0.466 (0.063)	0.004 (0.076)	0.358 (0.315)	0.428 (0.085)	0.429 (0.148)
			n-cp	0.312 (0.331)	0.183 (0.129)	0.051 (0.084)	0.464 (0.343)	0.049 (0.139)	0.234 (0.116)
			n-mt-cp	0.107 (0.160)	0.238 (0.071)	0.075 (0.067)	0.251 (0.210)	0.126 (0.085)	0.226 (0.080)
0.9	0.05	0.1	n-mt	0.533 (0.276)	0.814 (0.200)	0.025 (0.083)	1.210 (0.466)	0.211 (0.320)	0.253 (0.235)
			n-cp	0.486 (0.299)	0.200 (0.088)	0.072 (0.086)	0.162 (0.423)	0.763 (0.099)	0.066 (0.142)
			n-mt-cp	0.553 (0.215)	0.190 (0.107)	0.078 (0.066)	0.166 (0.270)	0.219 (0.163)	0.165 (0.097)

Migrant composition B: $\bar{u}_{11} = 0.7$, $\bar{w}_{22} = 0.3$, $\bar{P}_{1c} = 0.0$, $\bar{Q}_{1c} = 1.0$; $\Delta(z) = |(z) - z|/z$ for each parameter z; *, <m> or <M> given instead of $\Delta(m)$ or $\Delta(M)$; numbers in parentheses are lengths of 95% confidence intervals; mtDNA and cpDNA, maternally inherited and paternally inherited cytoplasmic markers, respectively. NA, not available.

APPENDIX D

Estimates of seed and pollen migration rates from simulated data sets for hybrid zone model with fixed population differences

m ₁	m ₂	M ₁	M ₂	s	System	Δ(m ₁)			Δ(m ₂)			Δ(M ₁)			Δ(M ₂)		
						N = 100	N = 300	N = 500	N = 100	N = 300	N = 500	N = 100	N = 300	N = 500	N = 100	N = 300	N = 500
0.1	0.1	0.1	0.1	0.1	n-mt	0.306	0.371	0.022	0.306	0.251	0.104	3.197	3.908	3.925	4.534	3.673	3.334
						(0.112)	(0.056)	(0.058)	(0.111)	(0.071)	(0.057)	(0.521)	(0.533)	(0.558)	(0.632)	(0.510)	(0.493)
						0.154	0.944	0.471	0.253	0.865	0.131	0.523	2.970	0.897	0.038	2.781	0.407
						(0.144)	(0.159)	(0.159)	(0.201)	(0.168)	(0.128)	(0.517)	(0.496)	(0.418)	(0.481)	(0.509)	(0.412)
	0.5	n-mt-cp	0.211	0.357	0.008	0.339	0.237	0.101	0.186	1.445	0.172	0.092	0.092	1.151	0.225		
			(0.091)	(0.052)	(0.057)	(0.111)	(0.058)	(0.051)	(0.272)	(0.263)	(0.174)	(0.309)	(0.263)	(0.165)			
			0.579	0.112	0.072	0.458	0.054	0.125	0.997	0.781	0.576	0.001	0.960	0.606			
			(0.171)	(0.087)	(0.055)	(0.162)	(0.077)	(0.072)	(0.502)	(0.573)	(0.518)	(0.643)	(0.498)	(0.451)			
	0.9	n-cp	0.255	0.456	0.220	0.999	0.271	0.228	1.925	0.977	0.705	5.990	0.690	0.990			
			(0.199)	(0.192)	(0.161)	(0.173)	(0.174)	(0.143)	(0.403)	(0.553)	(0.554)	(0.679)	(0.564)	(0.510)			
			0.704	0.087	0.167	0.519	0.019	0.010	1.000	0.474	1.050	0.916	0.356	0.245			
			(0.147)	(0.082)	(0.049)	(0.130)	(0.069)	(0.052)	(0.263)	(0.328)	(0.254)	(0.469)	(0.333)	(0.224)			
0.05	0.1	0.05	0.1	0.1	n-mt	0.017	0.168	0.265	0.463	0.371	0.295	2.176	0.500	2.830	0.670	1.017	4.018
						(0.206)	(0.117)	(0.100)	(0.265)	(0.140)	(0.109)	(0.717)	(0.552)	(0.500)	(0.590)	(0.603)	(0.627)
						0.360	0.192	0.051	0.026	0.099	0.009	4.594	2.632	0.775	3.406	5.367	0.664
						(0.172)	(0.093)	(0.089)	(0.218)	(0.123)	(0.090)	(0.783)	(0.494)	(0.555)	(0.689)	(0.628)	(0.603)
	0.5	n-cp	0.056	0.033	0.010	0.514	0.205	0.038	0.828	0.652	1.010	1.000	2.590	1.607			
			(0.153)	(0.081)	(0.065)	(0.214)	(0.109)	(0.066)	(0.603)	(0.451)	(0.498)	(0.572)	(0.683)	(0.575)			
			0.507	0.254	0.605	0.367	0.424	0.539	5.265	5.118	4.157	5.766	5.537	5.451			
			(0.066)	(0.045)	(0.026)	(0.129)	(0.074)	(0.058)	(0.360)	(0.334)	(0.285)	(0.741)	(0.679)	(0.714)			
	0.9	n-mt-cp	0.585	0.191	0.126	0.869	0.235	0.122	3.187	0.931	0.800	5.049	0.999	0.707			
			(0.122)	(0.065)	(0.051)	(0.265)	(0.149)	(0.146)	(0.357)	(0.138)	(0.106)	(0.721)	(0.303)	(0.289)			
			0.441	0.260	0.273	0.286	0.020	0.142	2.754	0.435	0.182	3.386	0.043	0.200			
			(0.057)	(0.043)	(0.025)	(0.130)	(0.073)	(0.059)	(0.312)	(0.089)	(0.084)	(0.672)	(0.219)	(0.194)			
0.1	0.1	0.05	0.1	0.1	n-mt	0.116	0.207	0.249	0.186	0.016	0.234	2.175	6.086	0.936	4.279	5.191	0.617
						(0.080)	(0.046)	(0.036)	(0.167)	(0.069)	(0.067)	(0.310)	(0.426)	(0.353)	(0.775)	(0.662)	(0.680)
						0.167	0.568	0.264	0.161	0.076	0.270	0.907	1.000	0.645	2.566	0.214	0.840
						(0.111)	(0.085)	(0.065)	(0.267)	(0.170)	(0.136)	(0.293)	(0.256)	(0.331)	(0.775)	(0.632)	(0.680)
	0.5	n-cp	0.530	0.059	0.074	0.650	0.153	0.089	0.998	0.733	0.402	0.037	1.053	0.182			
			(0.075)	(0.046)	(0.026)	(0.170)	(0.068)	(0.050)	(0.152)	(0.264)	(0.126)	(0.515)	(0.426)	(0.277)			
			1.317	2.405	1.508	1.979	2.305	1.764	0.958	7.014	5.135	2.408	4.942	5.519			
			(0.105)	(0.073)	(0.061)	(0.265)	(0.127)	(0.140)	(0.332)	(0.532)	(0.374)	(0.872)	(0.622)	(0.693)			
	0.9	n-cp	0.024	0.240	0.549	0.455	0.330	0.694	0.021	0.892	0.145	0.995	1.000	1.000			
			(0.111)	(0.050)	(0.068)	(0.303)	(0.084)	(0.144)	(0.595)	(0.361)	(0.349)	(0.865)	(0.644)	(0.662)			
			0.022	0.147	0.323	0.283	0.181	0.462	1.000	0.194	3.304	0.289	1.000	2.375			
			(0.086)	(0.049)	(0.047)	(0.233)	(0.090)	(0.097)	(0.288)	(0.424)	(0.368)	(0.862)	(0.548)	(0.670)			

(continued)

APPENDIX D
(Continued)

m ₁	m ₂	M ₁	M ₂	s	System	Δ(m ₁)			Δ(m ₂)			Δ(M ₁)			Δ(M ₂)			
						N = 100	N = 300	N = 500	N = 100	N = 300	N = 500	N = 100	N = 300	N = 500	N = 100	N = 300	N = 500	N = 100
0.0	0.05	0.1	0.2	0.1	n-mt	0.000*	0.000*	0.000*	1.073	8.391	1.000	0.074	4.520	0.988	0.999	0.783	0.980	
						(0.000)	(0.000)	(0.000)	(0.484)	(0.536)	(0.246)	(0.901)	(0.887)	(0.299)	(0.739)	(0.680)	(0.782)	
						0.018*	0.000*	0.000*	0.090	0.119	0.817	0.207	0.221	0.287	0.129	0.499	0.138	
						(0.055)	(0.012)	(0.009)	(0.154)	(0.104)	(0.056)	(0.198)	(0.084)	(0.057)	(0.429)	(0.229)	(0.169)	
	0.05	0.1	0.2	0.1	n-mt-cp	0.000*	0.000*	0.000*	0.686	0.128	0.818	0.108	0.222	0.286	0.211	0.504	0.140	
						(0.000)	(0.000)	(0.000)	(0.095)	(0.104)	(0.063)	(0.162)	(0.078)	(0.056)	(0.349)	(0.187)	(0.193)	
						0.000*	0.000*	0.000*	9.944	6.260	7.699	5.024	8.762	8.259	1.000	0.931	0.855	
						(0.000)	(0.000)	(0.000)	(0.648)	(0.368)	(0.410)	(0.991)	(0.746)	(0.710)	(0.742)	(0.534)	(0.376)	
	0.05	0.1	0.2	0.1	n-cp	0.027*	0.000*	0.004*	2.521	0.439	0.459	0.608	0.126	0.094	0.996	0.237	0.456	
						(0.050)	(0.010)	(0.023)	(0.227)	(0.078)	(0.103)	(0.148)	(0.111)	(0.100)	(0.381)	(0.373)	(0.243)	
						0.000*	0.000*	0.000*	0.581	0.443	0.194	0.030	0.123	0.049	0.076	0.237	0.251	
						(0.000)	(0.000)	(0.000)	(0.175)	(0.068)	(0.070)	(0.166)	(0.097)	(0.072)	(0.447)	(0.301)	(0.232)	
0.05	0.1	0.2	0.1	n-mt	0.000*	0.000*	0.000*	0.284	1.788	1.549	0.176	7.766	5.925	0.998	0.489	0.536		
					(0.000)	(0.000)	(0.000)	(0.192)	(0.214)	(0.139)	(0.923)	(0.920)	(0.883)	(0.821)	(0.801)	(0.816)		
					0.000*	0.001*	0.008*	0.488	0.359	1.336	0.224	0.023	0.442	1.347	0.026	1.000		
					(0.000)	(0.007)	(0.015)	(0.105)	(0.130)	(0.143)	(0.315)	(0.235)	(0.144)	(0.904)	(0.827)	(0.722)		
0.05	0.01	0.0	0.2	0.1	n-mt-cp	0.000*	0.000*	0.000*	0.487	0.139	0.126	0.225	0.191	0.613	1.342	0.680	2.931	
						(0.000)	(0.000)	(0.000)	(0.125)	(0.108)	(0.065)	(0.475)	(0.199)	(0.145)	(0.895)	(0.862)	(0.533)	
						0.777	0.594	0.239	0.098	0.513	0.329	0.019*	0.299*	0.135*	0.696	2.196	0.992	
						(0.196)	(0.060)	(0.076)	(0.025)	(0.014)	(0.013)	(0.363)	(0.353)	(0.336)	(0.715)	(0.690)	(0.660)	
0.05	0.01	0.0	0.2	0.1	n-cp	0.403	0.007	0.422	1.000	1.000	1.000	0.000*	0.016*	0.006*	0.494	0.154	0.311	
						(0.095)	(0.043)	(0.028)	(2.0E-6)	(0.014)	(0.017)	(5.0E-6)	(0.065)	(0.029)	(0.425)	(0.191)	(0.129)	
						0.512	0.005	0.359	0.069	0.181	0.435	0.000*	0.008*	0.001*	0.513	0.014	0.359	
						(0.097)	(0.040)	(0.027)	(0.017)	(0.012)	(0.005)	(5.0E-6)	(0.062)	(0.019)	(0.358)	(0.159)	(0.099)	
	0.05	0.01	0.0	0.2	0.1	n-mt	0.723	0.279	0.342	0.653	0.462	0.698	0.342*	0.114*	0.306*	2.139	0.599	2.405
							(0.070)	(0.078)	(0.062)	(0.016)	(0.012)	(0.016)	(0.424)	(0.436)	(0.350)	(0.686)	(0.636)	(0.474)
							0.077	0.240	0.082	1.000	1.000	0.998	0.000*	0.058*	0.000*	0.734	0.506	0.276
							(0.080)	(0.035)	(0.033)	(0.032)	(0.023)	(0.013)	(0.024)	(0.161)	(0.033)	(0.481)	(0.275)	(0.156)
	0.05	0.01	0.0	0.2	0.1	n-cp	0.226	0.158	0.066	0.040	0.376	0.345	0.000*	0.041*	0.000*	0.315	0.360	0.225
							(0.058)	(0.034)	(0.029)	(0.014)	(0.006)	(0.009)	(0.069)	(0.155)	(0.010)	(0.359)	(0.228)	(0.109)
							1.411	0.030	0.136	0.801	0.131	0.226	0.001*	0.366*	0.419	0.754	0.611	0.268
							(0.220)	(0.066)	(0.047)	(0.034)	(0.014)	(0.009)	(0.601)	(0.650)	(0.704)	(0.849)	(0.343)	(0.303)
0.05	0.01	0.0	0.2	0.1	n-cp	0.021	0.249	0.330	1.000	0.417	0.821	0.221*	0.000*	0.171*	0.819	0.004	0.222	
						(0.108)	(0.043)	(0.047)	(0.016)	(0.013)	(0.011)	(0.625)	(0.374)	(0.448)	(0.539)	(0.339)	(0.294)	
						0.658	0.125	0.139	0.234	0.049	0.235	0.000*	0.000*	0.000*	0.438	0.025	0.339	
						(0.131)	(0.050)	(0.035)	(0.024)	(0.011)	(0.007)	(0.434)	(0.082)	(0.159)	(0.381)	(0.209)	(0.117)	

(continued)

APPENDIX D
(Continued)

m_1	m_2	M_1	M_2	s	System	$\Delta(m_1)$			$\Delta(m_2)$			$\Delta(M_1)$			$\Delta(M_2)$		
						N = 100	N = 300	N = 500	N = 100	N = 300	N = 500	N = 100	N = 300	N = 500	N = 100	N = 300	N = 500
0.0	0.05	0.1	0.0	0.1	n-nt	0.000*	0.000*	0.000*	0.032	1.873	0.032	4.786	2.817	0.421*	0.134*	0.281*	
					n-nt	(0.000)	(0.000)	(0.000)	(0.142)	(0.278)	(0.142)	(0.742)	(0.677)	(0.492)	(0.540)	(0.540)	
					n-cp	0.000*	0.000*	0.001*	0.058	0.508	0.239	0.238	0.445	0.012*	0.007*	0.000*	
					n-nt-cp	(0.028)	(0.011)	(0.011)	(0.076)	(0.043)	(0.035)	(0.228)	(0.085)	(0.108)	(0.031)	(0.014)	(0.014)
					n-nt-cp	0.000*	0.000*	0.000*	0.057	0.507	0.247	0.239	0.445	0.012*	0.007*	0.001*	
					n-nt-cp	(0.000)	(0.000)	(0.000)	(0.081)	(0.045)	(0.033)	(0.248)	(0.086)	(0.107)	(0.032)	(0.024)	(0.024)
			0.5	n-nt	0.000*	0.000*	0.000*	4.141	7.058	2.294	8.505	8.989	3.045	0.041*	0.000*	0.010*	
			n-cp	0.000*	0.000*	0.000*	0.000*	0.385	0.284	0.291	0.713	0.536	0.365	0.217	0.231	0.231	
			n-nt-cp	0.006*	0.000*	0.001*	0.017	0.104	0.013	0.013	0.013	0.242	0.086	0.000*	0.025*	0.000*	
			n-nt-cp	(0.030)	(0.016)	(0.011)	(0.093)	(0.063)	(0.038)	(0.038)	(0.210)	(0.119)	(0.085)	(0.153)	(0.087)	(0.026)	
			n-nt-cp	0.000*	0.000*	0.000*	0.186	0.104	0.015	0.015	0.342	0.243	0.103	0.039*	0.025*	0.000*	
			n-nt-cp	(0.000)	(0.000)	(0.000)	(0.085)	(0.027)	(0.038)	(0.081)	(0.256)	(0.117)	(0.080)	(0.237)	(0.109)	(0.056)	
			0.9	n-nt	0.000*	0.000*	0.000*	1.000	0.512	1.933	3.064	4.004	6.507	0.097*	0.003*	0.005*	
			n-cp	(0.000)	(0.000)	(0.000)	(0.785)	(0.142)	(0.177)	(0.890)	(0.936)	(0.924)	(0.193)	(0.742)	(0.281)	(0.670)	
			n-nt-cp	0.000*	0.000*	0.001*	1.000	0.759	0.646	0.646	0.514	0.696	0.272	0.000*	0.050*	0.030*	
			n-nt-cp	(0.000)	(0.001)	(0.005)	(0.000)	(0.027)	(0.081)	(0.081)	(1.000)	(0.064)	(0.208)	(0.218)	(0.203)	(0.542)	
			n-nt-cp	0.000*	0.000*	0.000*	1.000	0.758	0.456	0.456	0.510	0.696	0.441	0.000*	0.049*	0.143*	
			n-nt-cp	(0.000)	(0.000)	(0.000)	(0.000)	(0.029)	(0.090)	(0.090)	(1.000)	(0.065)	(0.193)	(0.742)	(0.281)	(0.670)	
0.05	0.01	0.1	0.2	0.1	n-nt	0.120	0.534	0.437	0.162	0.182	0.440	3.260	0.714	1.623	0.109	0.052	
			n-cp	(0.121)	(0.095)	(0.066)	(0.023)	(0.016)	(0.012)	(0.511)	(0.430)	(0.511)	(0.430)	(0.606)	(0.564)	(0.514)	
			n-nt-cp	0.186	0.415	0.269	1.000	0.196	0.152	1.139	0.322	0.143	0.558	0.006	0.050	0.050	
			n-nt-cp	(0.127)	(0.078)	(0.064)	(0.098)	(0.073)	(0.053)	(0.339)	(0.205)	(0.160)	(0.345)	(0.230)	(0.183)	(0.183)	
			n-nt-cp	0.201	0.054	0.028	0.145	0.262	0.010	0.913	0.122	0.001	0.542	0.115	0.072	0.072	
			n-nt-cp	(0.111)	(0.062)	(0.044)	(0.021)	(0.012)	(0.010)	(0.316)	(0.123)	(0.128)	(0.313)	(0.163)	(0.112)	(0.112)	
			0.5	n-nt	1.383	0.890	0.086	0.939	1.169	0.019	0.917	3.134	3.381	0.847	1.931	1.482	
			n-cp	(0.219)	(0.125)	(0.070)	(0.040)	(0.029)	(0.014)	(0.451)	(0.457)	(0.505)	(0.584)	(0.459)	(0.467)	(0.467)	
			n-nt-cp	0.458	0.043	0.002	1.000	1.000	1.000	0.398	0.001	0.289	0.289	0.235	0.372	0.372	
			n-nt-cp	(0.153)	(0.070)	(0.069)	(0.093)	(0.064)	(0.044)	(0.391)	(0.181)	(0.233)	(0.260)	(0.260)	(0.261)	(0.261)	
			n-nt-cp	0.770	0.572	0.198	0.469	0.842	0.127	0.382	0.555	0.133	0.514	0.021	0.123	0.123	
			n-nt-cp	(0.119)	(0.074)	(0.049)	(0.030)	(0.020)	(0.011)	(0.301)	(0.122)	(0.158)	(0.319)	(0.139)	(0.145)	(0.145)	
			0.9	n-nt	0.120	0.160	0.006	0.032	0.078	0.075	4.344	0.468	4.258	0.801	0.488	0.850	
			n-cp	(0.087)	(0.072)	(0.042)	(0.022)	(0.014)	(0.010)	(0.696)	(0.585)	(0.445)	(0.412)	(0.334)	(0.334)		
			n-nt-cp	0.983	0.493	0.104	2.988	1.025	0.213	1.000	0.998	5.058	0.061	0.971	0.971		
			n-nt-cp	(0.188)	(0.088)	(0.054)	(0.088)	(0.037)	(0.019)	(0.688)	(0.599)	(0.548)	(0.458)	(0.223)	(0.223)		
			n-nt-cp	0.068	0.123	0.032	0.254	0.046	0.104	4.783	2.469	4.007	1.108	1.083	0.716		
			n-nt-cp	(0.085)	(0.054)	(0.051)	(0.031)	(0.012)	(0.011)	(0.746)	(0.561)	(0.498)	(0.442)	(0.331)	(0.233)		

$\Delta(z) = |(\langle z \rangle - z)/z|$ for each parameter z ; *, (z) is given instead of $\Delta(z)$; numbers in parentheses are lengths of 95% confidence intervals; mtDNA and cpDNA, maternally inherited and paternally inherited cytoplasmic markers, respectively.