

The Number of Self-Incompatibility Alleles in a Finite, Subdivided Population

Mikkel Heide Schierup

Department of Ecology and Genetics, Institute of Biology, University of Aarhus, DK-8000 Aarhus C, Denmark

Manuscript received November 10, 1997
Accepted for publication February 27, 1998

ABSTRACT

The actual and effective number of gametophytic self-incompatibility alleles maintained at mutation-drift-selection equilibrium in a finite population subdivided as in the island model is investigated by stochastic simulations. The existing theory founded by Wright predicts that for a given population size the number of alleles maintained increases monotonically with decreasing migration as is the case for neutral alleles. The simulation results here show that this is not true. At migration rates above $Nm = 0.01-0.1$, the actual and effective number of alleles is lower than for an undivided population with the same number of individuals, and, contrary to Wright's theoretical expectation, the number of alleles is not much higher than for an undivided population unless $Nm < 0.001$. The same pattern is observed in a model where the alleles display symmetrical overdominant selection. This broadens the applicability of the results to include proposed models for the major histocompatibility (MHC) loci. For a subdivided population over a large range of migration rates, it appears that the number of self-incompatibility alleles (or MHC-alleles) observed can provide a rough estimate of the total number of individuals in the population but it underestimates the neutral effective size of the subdivided population.

IN a controversial paper, Wright (1939) presented a mathematical model for the number of alleles in a single-locus multi-allelic gametophytic self-incompatibility (GSI) system that can be maintained at equilibrium between selection, mutation and drift in a finite population. He used a diffusion approximation to derive the stationary frequency distribution from which the number of alleles can be determined as the reciprocal of the mean of the distribution integrated from $1/(2N)$ to 1, where N is the number of individuals. His motivation for the work was to explain the finding by Emerson (1939) of 37 alleles in a sample from a population of only about 500 individuals. His results showed that a large number of alleles is indeed maintained due to the very strong frequency-dependent selection. The work was, however, criticized first by Fisher (1958) and later by Moran (1962). They maintained that the use of the diffusion approximation was invalid because selection is strong enough to change the frequency of an allele drastically in one generation, *e.g.*, from $1/(2N)$ to $1/2$ in one generation with three alleles, and because the frequency change of an allele in one generation depends not only on its own frequency but also on the frequency of all other alleles, *i.e.*, the problem has a dimension equal to the total number of self-incompatibility (S -) alleles in the population. Wright (1960, 1964) replied to these criticisms and showed, by comparison to Fisher's derivations, that with the addition of a

minor correction to the variance component his formulae predict the number of S -alleles very well despite their violation of the assumptions behind the diffusion approximation. The accuracy of his formulae was later confirmed by computer simulations (Ewens and Ewens 1966; Mayo 1966; Crosby 1966), and the debate was laid to rest. A more tractable and slightly improved formula to estimate the number of S -alleles in a finite population was later given by Yokoyama and Nei (1979), and Yokoyama and Hetherington (1982) investigated the expected number of S -alleles in a sample.

Although Wright's treatment showed that a large number of alleles is maintained under GSI, the number of alleles observed by Emerson in an isolated population of 500 individuals could not be explained under biologically reasonable values of the mutation rate ($u < 10^{-5}$). Therefore, Wright extended his theory in the second part of the 1939 paper to investigate the effect of population subdivision. A basic observation from his theory for an undivided population is that the relation between population size and the number of S -alleles is concave. Therefore, in a finite subdivided population, the number of S -alleles maintained in the population will be greatest when the population is divided into completely isolated subpopulations because subpopulations under the infinite alleles model will contain, after sufficient time, a unique set of alleles each. Migration between subpopulations will oppose this effect and reduce the total number of alleles, but Wright posited that the number of S -alleles maintained in any subdivided population will still exceed that of an undivided population of equal total size. This conjecture was later shown to be correct for neutral alleles (Nagylaki 1985). Wright

Address for correspondence: Mikkel Heide Schierup, Department of Ecology and Genetics, University of Aarhus, Building 540, Ny Munkegade, DK-8000 Aarhus C., Denmark.
E-mail: mikkel.schierup@biology.aau.dk

(1939) produced formulae that quantitatively describe the effect of migration using a finite island model with pollen migration. He concluded that subdivision with small, but reasonable pollen migration rates ($m_p = 0.002-0.02$) can greatly increase the number of *S*-alleles in the population. But his conclusions were based on four numerically solved examples of a set of formulae derived using a number of crude approximations. Their accuracy has never been checked in detail. Ewens and Ewens (1966) performed a single simulation of the finite island model of *S*-alleles and concluded that the effect of subdivision is small. However, they used a very high migration rate and did not relate their results to Wright's.

In recent years, the loci believed to be responsible for the incompatibility reaction have been cloned in the single-locus, multi-allelic GSI systems of Solanaceae (Anderson *et al.* 1986) and Papaveraceae (*Papaver rhoeas*; Foote *et al.* 1994), which revealed that these systems evolved independently. The frequency dependent selection is theoretically expected to give extremely long coalescence times of alleles (Vekemans and Slatkin 1994). This prediction is consistent with analyses of nucleotide sequences in the Solanaceae (Ioerger *et al.* 1990) where alleles show some transspecific polymorphisms and coalescence times of all *S*-alleles within a species have been estimated to be 40–80 million years. Recently, Richman *et al.* (1995, 1996a,b) found highly divergent nucleotide sequences between alleles in natural populations of two Solanaceous species. Using the theory by Wright (1939) and Vekemans and Slatkin (1994), they estimated the coalescence time of all alleles to infer the long-term effective population size, and they estimated the total number of *S*-alleles from their sample to derive the short-term effective population size for each species. They concluded that the two species had very different population histories and size and argued that these differences might partly be explained by differences in their ecology (Richman *et al.* 1996b; Richman and Kohn 1996). However, these estimates assume that each species forms a single panmictic population. This is unlikely, because individuals of both species inhabit large fragmented areas.

In Papaveraceae, the number of *S*-alleles and their frequencies have been estimated from controlled crosses in four populations of *Papaver rhoeas* (Lawrence and O'Donnell 1981; Lawrence *et al.* 1993; Lane and Lawrence 1993). Contrary to the expectation for a panmictic population, these studies showed that the frequencies of *S*-alleles are highly unequal. Furthermore, even widely separated populations share a large fraction of the alleles. The two alleles so far sequenced only show 55% similarity in the deduced amino acid sequence, but two copies of one of the alleles, from a British and a Spanish population, were found to be identical at the deduced amino acid level (Walker *et al.* 1996).

The purpose of this article is to revisit the qualitative

and quantitative effects of population subdivision on the number of *S*-alleles maintained at equilibrium in a finite population. To date, only Wright's theory has modelled the number of alleles maintained by balancing selection in a subdivided population, and therefore his theory has also been referred to in discussions of symmetrical overdominant selection (Takahata 1993). The (unknown) mechanism responsible for the high polymorphism and allelic diversity of several loci in the major histocompatibility complex (MHC, and the homologous HLA system in humans) is mimicked by symmetrical overdominance or the closely related model of frequency-dependent selection by minority advantage (Takahata 1990; Takahata and Nei 1990), assuming selection coefficients in the range of 0.1–10% (Maruyama and Nei 1981; Takahata *et al.* 1992; Satta *et al.* 1994). Inferences about long-term human evolution based on MHC allele surveys have been made assuming symmetrical overdominance in an undivided population (Klein *et al.* 1990; Takahata 1993; Ayala 1995; Ayala and Escalante 1996), so it is also important to study the effect of subdivision on the number of alleles maintained under symmetrical overdominance.

I have used stochastic simulations to characterize the dynamics of *S*-alleles in a finite island model at mutation-selection-drift equilibrium. Following Wright (1939), a finite island model of equally sized demes with pollen migration is considered, and the results are compared to values obtained from his formulae. The quantitative effects of subdivision of a fixed number of individuals are considered, and the stationary frequency distributions of alleles under different migration rates and population sizes are analyzed. Furthermore, the effect of relaxing the assumption of a fixed migration rate is studied. For *S*-alleles, the effective migration rate after mating exceeds the proportion of migrating pollen if migrants are more likely to be compatible than nonmigrants. A model of symmetrical overdominance is also studied, and the qualitative results of both models are compared to the theory for neutral variation in a subdivided population.

THEORY

Single population: Gametophytic self-incompatibility: Consider a population of N diploid individuals with a gametophytic self-incompatibility system with n_a alleles, S_1, S_2, \dots, S_{n_a} . In this system the compatibility reaction is determined by matching either of the two alleles in the style with the allele in the pollen. If either allele matches, the combination is incompatible. No other selection occurs. Mutation is at the rate u , with new *S*-alleles generated according to the infinite alleles model, *i.e.*, no back-mutation is allowed.

At least three alleles are necessary for the system to work, and all individuals are heterozygous. Selection under GSI is therefore stronger than with extreme overdominant selection where the fitness of each homozygote is zero, because GSI produces direct selection for rare genotypes in that they have a larger mating pool than common ones (Clark 1993; Veke-

mans and Slatkin 1994). The stationary frequency spectrum $\phi(q)$ of a given allele was derived by Wright (1939) using diffusion approximation. At equilibrium, the mutation rate balances the loss of alleles due to drift:

$$u = \frac{n(n-3)}{4N(n-1)} \frac{1}{2N} \phi\left(\frac{1}{2N}\right).$$

(Wright 1939). The number of alleles can be found by numerical solution of this equation and the condition

$$\int_{1/2N}^{1/2} \phi(q) dq = 1$$

on the gene frequency spectrum ϕ . Yokoyama and Nei (1979) provide a simpler formula that gives slightly different results because it is based on a different variance component more appropriate for a self-incompatibility system (Fisher 1958; Wright 1960). However, to allow a direct comparison, Wright's formulae are used in this study.

Neutrality and symmetrical overdominance: The expectation for the effective number of alleles n_E ($n_E = 1/F$, where F is the population fixation index) maintained in a population of N individuals was derived by Kimura and Crow (1964) for a neutral locus and a locus with symmetrical overdominance. For the neutral locus $n_E = 1/(4Nu + 1)$. For symmetrical overdominance n_E is found by a diffusion approximation approach very similar to the one described for GSI. Ewens (1964) obtained expressions for the actual number of alleles n_a for both neutrality and symmetrical overdominance.

Subdivided population: Population subdivision is modelled using the finite island model with P equally sized subpopulations. Migration between the subpopulations occurs exclusively via pollen, *i.e.*, a pollen cloud of migrants, at a constant rate m_p . The migration rate m by both sexes that is usually used in the island model differs from m_p but to a good approximation,

$$m = \frac{1}{2} m_p$$

(Wright 1939). I refer to this migration rate as m_p except when discussing the number of migrants, which is usually denoted by

$$Nm = \frac{1}{2} Nm_p$$

Gametophytic self-incompatibility: Wright (1939) extended the theory to a subdivided population by constructing the stationary frequency distribution for an allele in the total population. To be able to use the diffusion approximation for this problem, Wright assumed the following: (1) The proportion of demes where a given allele is present is proportional to its frequency; (2) a given allele has roughly the same frequency in all the demes in which it occurs; and (3) an incoming migrant allele is equally likely to be of any extant type.

If the mutation rate is small, these assumptions imply that the stationary frequency distribution of an allele in the total population is expected to be approximately symmetric, as in a single population.

These approximations allow the problem to be reduced to one dimension, where the process is determined by the balance between migration/mutation and drift in the local demes and by the balance between mutation and drift in the total population. The number of alleles maintained can then be determined by simultaneous solutions of the equations at equilibrium between mutation, migration and loss of alleles in

the deme, and the corresponding equations at equilibrium between mutation and loss of alleles in the total population.

Neutrality and symmetrical overdominance: Neutrality was analyzed by Maruyama (1970) who obtained explicit expressions for the effective number of alleles under the finite island and stepping stone models, and Nagylaki (1985) proved that n_E for a subdivided population under any migration pattern always exceeds the value for an undivided population, though the increase is significant only when $Nm < 1$. Symmetrical overdominance in a subdivided population has not been treated analytically.

COMPUTER SIMULATIONS

The evolutionary dynamics of S -alleles were simulated for a diploid population of P demes each with N individuals, giving a total population size of $N_i = PN$. The mating process for a single generation was as follows: For each deme a random individual among the N present was chosen as the maternal parent. A pollen genotype was then randomly chosen with probability $(1 - m_p)$ from the same deme and with probability m_p from the total population. For the GSI model the pollen is then checked for compatibility with the maternal plant by matching the haploid pollen genotype with each of the alleles of the maternal parent. If the combination is compatible, one of the two alleles of the maternal parent is then picked randomly and combined with the pollen allele to form a new zygote. If the combination is incompatible, the pollen is discarded and the ovule retained, *i.e.*, the pollen elimination model of Finney (1952). Depending on the migration model, a new pollen donor can either be chosen to have the same migratory status ("pollen packet" migration model), or else the migratory status of the pollen donor is again randomly determined ("single pollen" migration model). The "pollen packet" migration model was assumed by Wright and can be thought of as migration mediated by a pollinator carrying pollen grains from several sources, whereas "single pollen" migration assumes that pollen arrives to a given stigma independently, *e.g.*, by wind. The difference between the migration models is that, with the "pollen packet" model, the proportion of successful matings by migrating pollen is m_p whereas with "single pollen" migration the proportion of successful matings by migrating pollen is larger than m_p if a pollen grain is less likely to get a fertile mating in a within-deme cross than in a between-deme cross. The mating process is repeated until N new zygotes are produced in each deme. Each gene is then subjected to mutation with probability u , with each mutation giving rise to a new allelic type (the infinite alleles model). The maximal migration rate is $m_p = 1$, in which case all pollen grains are drawn randomly from the population, but ovules remain within demes (equivalent to $m = 0.5$).

For the symmetric overdominance model, a combination of maternal and paternal gametes is retained if the gametes have different genotypes; otherwise the homozygous zygote is discarded with probability s , the selection coefficient against homozygotes. In this case, only the "pollen packet" migration model is used.

Number of alleles: Each run was started with either $2N$, different alleles or, in the case of large population sizes, with 20 (GSI) or 7 (overdominance) alleles, in order to reach equilibrium faster. The population was allowed to evolve for the larger of $100/(Nu)$ and 60,000 generations before starting to record the average number of alleles per deme, the total number of alleles, and the corresponding effective number of alleles. These statistics were then recorded every $1/(1000u)$ generations for $40/(1000u)$ generations and the average numbers over this interval constituted one replicate. For each

TABLE 1
Comparison of Wright's theory with simulations

Case	Individuals (N)	Migration (m_p)	Mutation (u)	Number of alleles maintained		
				Wright's formulae	Pollen packet migration	Single pollen migration
Case 1						
Deme	50				8.29 (0.01)	8.60 (0.01)
Total (10 demes)	500	0.005	0.000045		18.09 (0.04)	17.74 (0.04)
Deme	50			11.6	10.56 (0.01)	10.88 (0.01)
Total (10 demes)	500	0.02	0.000045	20	16.78 (0.03)	16.77 (0.03)
Undivided	500		0.000045	17.5	16.96 (0.02)	
Case 2						
Deme	50			13.4	13.01 (0.01)	13.57 (0.01)
Total (10 demes)	500	0.02	0.001	50	43.11 (0.06)	41.97 (0.07)
Undivided	500		0.001	33.8	33.14 (0.05)	
Case 3						
Deme	50			8.2	7.78 (0.01)	8.08 (0.01)
Total (10 demes)	500	0.002	0.0002	50	34.88 (0.06)	32.62 (0.05)
Undivided	500		0.0002	22.1	21.54 (0.03)	
Case 4						
Deme	10				3.99 (0.01)	4.43 (0.01)
Total (50 demes)	500	0.005	0.00009		34.72 (0.07)	26.88 (0.06)
Deme	10			5.3	4.87 (0.01)	5.42 (0.01)
Total (50 demes)	500	0.02	0.00009	50	23.75 (0.04)	21.73 (0.05)
Undivided	500		0.00009	19.2	18.64 (0.03)	

The number of alleles obtained from Wright's theory and the number of alleles at equilibrium (SE) from simulations using the "pollen packet" and "single pollen" migration models are compared for the four cases previously reported by Wright (1939). For each case, the number of alleles in an undivided population, the average number of alleles within demes and the number in the total population for each subdivided population are given. For Case 1 and 4, results from simulations are also given for a fourfold reduction in migration rate. The number of alleles calculated by Wright are also included. The pollen migration rate is $m_p = 2m$, where m is the migration rate for both sexes.

parameter set 30 (for $N_i = 2,000$) or 100 (otherwise) such replicates were performed and standard deviations calculated over these replicates.

Stationary frequency distributions: For a few parameter sets, the stationary frequency distribution at equilibrium was recorded by observing the frequencies of each allele (both in the deme and in the total population) at 200 generation intervals over 20,000 generations with five replicates as defined above so that at least 10,000 allele frequencies were recorded for each parameter set investigated. To check Wright's assumption 1, the corresponding values of the frequency of a given allele, and the number of demes where it is present were recorded, and to check Wright's assumption 2, the frequency of each allele in each deme where it is present was also recorded.

The computer program was written so that it could check for its ability to produce results in concordance with the existing theory for neutral alleles by switching off the test for compatibility of two gametes. Furthermore, an independently developed program by X. Vekemans (Université Libre De Bruxelles, Belgium) was found to give identical results. The computer program is written in the C language, and is available from me by request.

RESULTS

Gametophytic self-incompatibility: Table 1 summarizes the results of simulations using the four parameter

sets for which Wright (1939) gave a numerical solution. Results are given with both the "pollen packet" migration model (assumed by Wright) and the "single pollen" migration model and results of simulations with a fourfold reduction in migration rate are also given for two of the parameter sets (Cases 1 and 4). The results for the undivided populations are slightly lower than Wright's predictions in all four cases, but less than 5% different. For the subdivided populations, however, the simulated results differ markedly from Wright's predictions in that the increase in the number of alleles maintained in the total population is far less than predicted, although the number of alleles in the local deme is more accurately predicted by Wright's formula. The fit to the prediction is poorest when the mutation rate is low and subdivision is greatest (Case 4). Even a fourfold decrease in migration rate does not give the total number of alleles predicted by Wright. The "single pollen" model (Table 1, last column) leads to an increase in the number of alleles in the local deme, but the number of alleles maintained in the total population is further decreased. The remaining results are all based on the "pollen packet" migration model.

For Case 4 (Table 1), Figure 1 shows the stationary

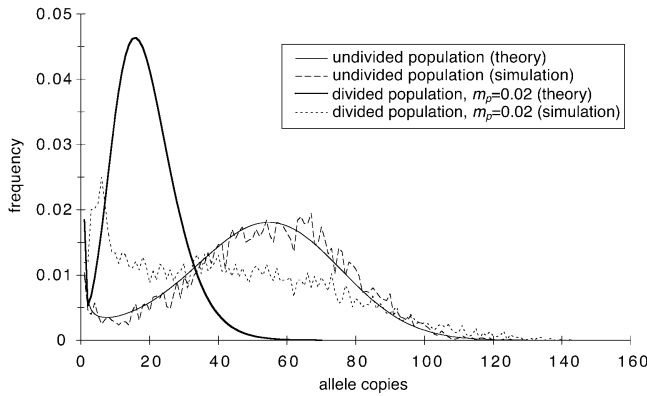


Figure 1.—Stationary frequency distributions for an undivided population ($N = 500$) and a subdivided population (50 demes with 10 individuals) for a pollen migration rate $m_p = 0.02$. Results from simulations and calculated from Wright's theory are shown.

frequency distribution for an undivided population and for a subdivided population with $m_p = 0.02$. The stationary frequency distributions obtained using Wright's formulae are also included in Figure 1. Again, for the undivided population, the theory fits the simulated data well. For the subdivided population, however, a large discrepancy is evident. The simulated distribution has a much larger variance and is highly positively skewed.

The effect of various degrees of subdivision, *i.e.*, changing migration rate, on the total number of alleles is shown in Figure 2 for two demes of 100 individuals with three different mutation rates ($u = 10^{-4}$, 10^{-5} , 10^{-6}). A surprising pattern is evident for $u = 10^{-5}$ and 10^{-6} in that the number of alleles maintained is smallest at intermediate migration rates, which is qualitatively the opposite of the expectation from Wright's theory. The migration rate where a minimum number of alleles is maintained decreases from approximately $m_p = 0.002$ ($Nm = 0.1$) to $m_p = 0.0004$ ($Nm = 0.02$) when u

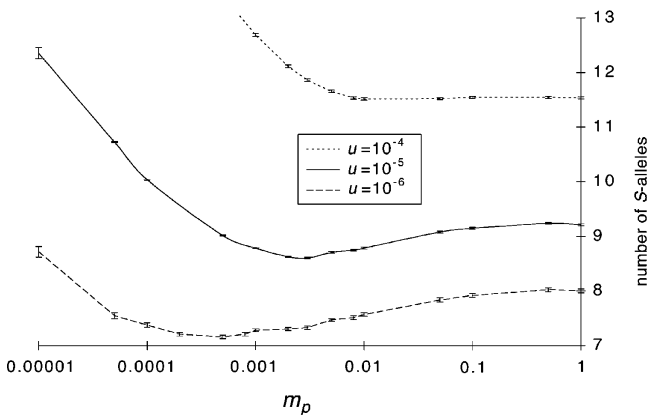


Figure 2.—The number of S -alleles (\pm SE) maintained in a population of 2 demes with 100 individuals as a function of the pollen migration rate m_p . Results are shown for three different mutation rates.

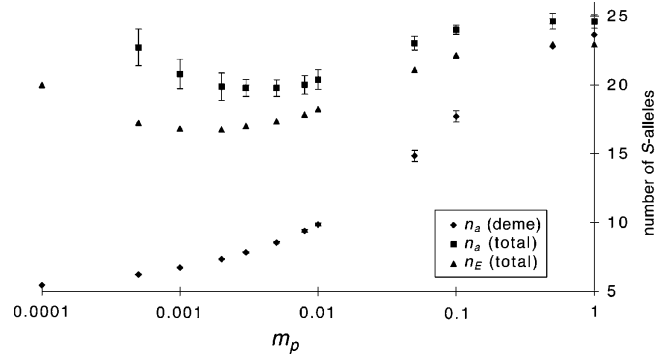


Figure 3.—The number of S -alleles n_s in the total population (\pm SE), the effective number of S -alleles in the total population n_E , and the average number of S -alleles in the demes (\pm SE) as a function of the pollen migration rate m_p . The population is divided into 40 demes with 50 individuals, and $u = 10^{-6}$.

increases from 10^{-5} to 10^{-6} . Furthermore, only when $Nm < 0.02$ ($u = 10^{-5}$) and $Nm < 0.002$ ($u = 10^{-6}$), respectively, is the number of alleles maintained in the total population larger than the value in an undivided population of the same size. To illustrate the quantitative effect in a larger and more subdivided population, Figure 3 shows results for a population of 2000 individuals divided into 40 demes, and with a mutation rate of 10^{-6} . Also included is the pattern of change in the average number of alleles in a single deme, and the effective number of alleles (n_E) in the total population. The pattern observed for the total number of alleles is qualitatively similar to Figure 2, with the minimum number of alleles maintained at a migration rate of approximately $m_p = 0.002$. At this point, only about 20 alleles are maintained, compared to 25 alleles in an undivided population of the same size. The average number of alleles in a single deme decreases monotonically with decreasing migration. The effective number of alleles is included for comparison with neutral theory, where n_E monotonously increases with decreasing m_p . In Figure 3, however, n_E in the total population shows a pattern similar to that for the actual number of alleles, but with the minimum shifted to approximately $m_p = 0.001$.

For the same parameter values as used in Figure 3, Figure 4 shows the stationary frequency distributions of an allele in the total population for four different migration rates, and Figure 5 shows the stationary allele frequency distribution in a deme, for the same migration rates. The four migration rates correspond to the migration rate limit ($m_p = 1$), the migration rate where the decrease in the number of alleles in Figure 3 is steepest ($m_p = 0.05$), a migration rate where the number of alleles is close to its minimum ($m_p = 0.001$), and a migration rate where the number of alleles exceeds that for an undivided population ($m_p = 0.0001$). Figure 4 confirms the results from Figure 1 that decreasing the migration rate dramatically increases the variance of

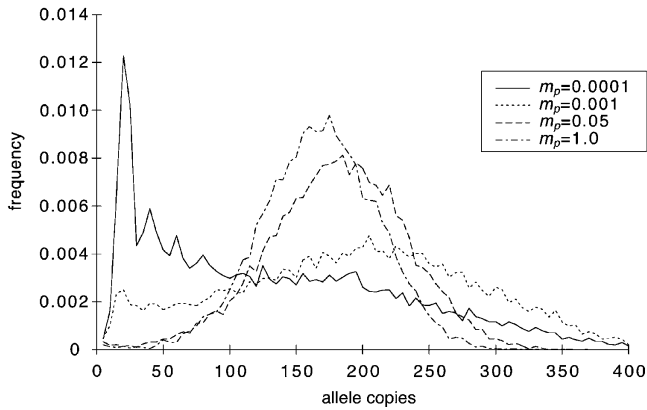


Figure 4.—The simulated stationary frequency spectra of an allele in the total population for four different pollen migration rates m_p . The population is divided into 40 demes with 50 individuals, and $u = 10^{-6}$.

the stationary frequency distribution. For the smallest migration rate there are therefore alleles of higher frequency than with the highest migration rate, even though the mean frequency of alleles is substantially smaller. The stationary frequency distribution in the local population also displays some striking patterns. For high ($m_p = 1$) and low ($m_p = 0.001$ and 0.0001) migration rates, the distribution appears bell-shaped, as expected under symmetric balancing selection, but there is a smaller mean at the high migration rate because more alleles are maintained in each deme. However, at an intermediate migration rate ($m_p = 0.05$), the shape of the distribution is very different, with a large proportion of alleles at low frequency. This fundamentally violates Wright's assumption 2, that an allele should have the same frequency in all demes where it is present.

Wright's assumption 1, that the number of copies of an allele is linearly proportional to the number of demes in which it is present, is illustrated in Figure 6 for the same migration rates. It appears that, for migra-

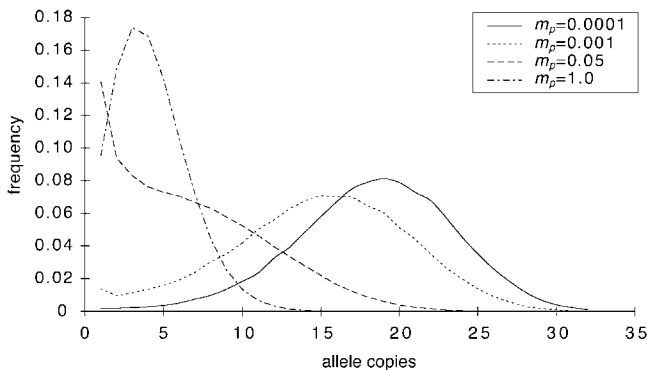


Figure 5.—The simulated stationary frequency spectra of an allele in a deme for four different pollen migration rates m_p . The population is divided into 40 demes with 50 individuals, and $u = 10^{-6}$.

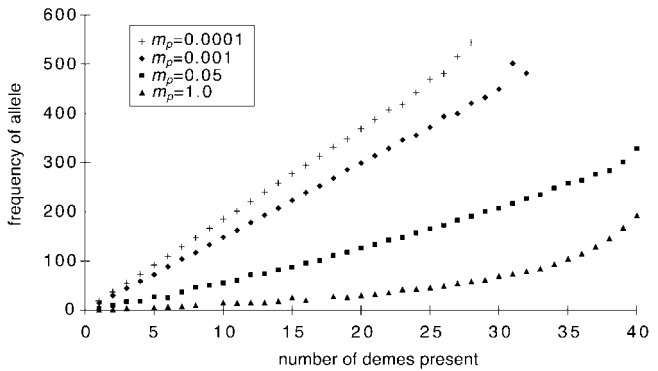


Figure 6.—The relation between the frequency of an allele and the number of demes where it is present for four different pollen migration rates m_p . The population is divided into 40 demes with 50 individuals, and $u = 10^{-6}$.

tion rates above the threshold migration rate yielding the minimum number of alleles ($m_p = 0.002$), there is a bias such that the expected number of occupied demes increases faster when the allele is rare than when it is common.

Symmetrical overdominance and neutrality: Figure 7 shows the relation of the different statistics for the same parameters as Figure 3, but for the symmetrical overdominance model with the selection against homozygotes $s = 0.2$ ($Ns = 10$). The qualitative pattern in this model is very similar to the GSI model, but the number of alleles maintained is generally lower because selection is weaker. The threshold migration rate that yields the minimum number of alleles is of the same magnitude as in the GSI model. The relative decline in number of alleles maintained at the minimum (approximately $m_p = 0.002$), compared to the undivided population case (indistinguishable from $m_p = 1$), is about 20–25% in both models (6 vs. 8 in the overdominance model, 20 vs. 25 in GSI).

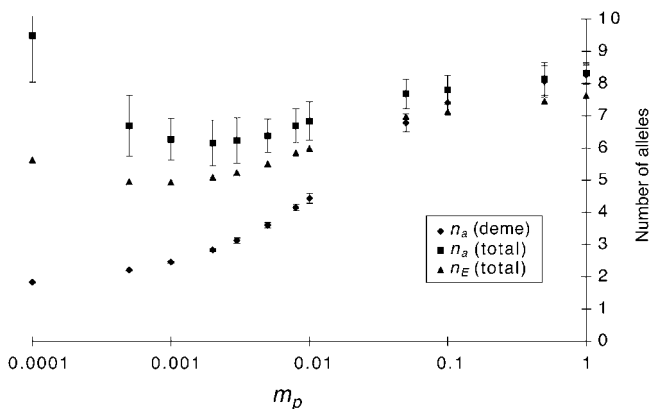


Figure 7.—Symmetrical overdominance model, $s = 0.2$. The number of alleles n_a in the total population (\pm SE), the effective number of alleles in the total population n_E , and the average number of alleles in the demes (\pm SE) as a function of the pollen migration rate m_p . The population is divided into 40 demes with 50 individuals, and $u = 10^{-6}$.

TABLE 2
The effective population size of a subdivided population

Migration m_p	Calculated from n_a		Calculated from n_E		Neutral effective size
	GSI	Overdominance	GSI	Overdominance	
0.0001	3,200	2,600	1,500	1,100	192,000
0.0005	1,720	1,330	1,130	920	97,000
0.002	1,280	1,100	1,060	950	11,500
0.02	1,530	1,600	1,430	1,480	2,950
0.1	1,920	1,800	1,870	1,750	2,190
1.0	2,000	2,000	2,000	1,950	2,000

The population sizes of an undivided population that would maintain the same number of alleles (n_a) and effective number of alleles (n_E) as a subdivided population with 40 demes of 50 individuals each for various pollen migration rates $m_p = 2m$, where m is the migration rate for both sexes. Estimated population sizes are shown for both the effective and actual number of alleles for the GSI and symmetrical overdominance ($s = 0.2$) case, respectively, letting $u = 10^{-6}$. The effective size of the subdivided population with respect to neutral variation is calculated using formula 6 of Nei and Takahata (1993).

The relation between n_E and m for the same u for a neutral locus calculated from Maruyama's (1970) exact formula is almost completely horizontal when $Nm > 1$, but increases quickly with decreasing m when $Nm < 1$ (results not shown).

Effective population size: The population size of an undivided population that would maintain the same number (actual or effective) of alleles as a subdivided population is shown as a function of the migration rate in Table 2, for 40 demes with 50 individuals and $u = 10^{-6}$. Estimated population sizes are included for both the GSI and the symmetrical overdominance model with $s = 0.2$. The population sizes are estimated by comparing the actual and effective number of alleles maintained at various migration rates, as shown in Figures 3 and 7 with corresponding Figures (not shown) of the relation between population size and the number of alleles (and effective alleles) for an undivided population with $u = 10^{-6}$. For instance, with GSI an undivided population of approximately 1280 individuals maintains about 20 alleles which is the same number as the subdivided population with $m_p = 0.002$. These population sizes can thus be viewed as the effective population size of the subdivided population with respect to the SI-system (or to a locus with symmetrical overdominance).

The last column of Table 2 shows the effective population size of a subdivided population, assuming the neutral model of genetic variation as calculated using Nei and Takahata's (1993) formula 6. Table 2 shows that the effective population sizes estimated from n_a for GSI and symmetrical overdominance are lower (20–40%) than the actual number of individuals, at intermediate migration rates. For effective sizes estimated from n_E the decrease at intermediate migration rates is even more pronounced and the effective size is below the number of individuals even at $m_p = 0.0001$. When $Nm < 1$, these effective sizes deviate markedly from the effective

size for a neutral locus, which increases dramatically with decreasing m_p .

DISCUSSION

The simulation study shows that the theory provided by Wright (1939) for a GSI-system in a subdivided population is insufficient. Wright's work supported the intuitively reasonable prediction that the number of S-alleles maintained under the finite island model always increases with decreasing migration rate between demes. Although this is correct for the effective number of neutral alleles (Maruyama 1970; Nagylaki 1985), the simulation results show that there exists an intermediate migration rate at which the number of S-alleles reaches a minimum, and that only below this migration rate is the conjecture correct. The migration rate at which this minimum number of S-alleles occurs decreases with the mutation rate (Figure 2), which means that Wright's conjecture is more inaccurate for small and biologically realistic mutation rates than for the four cases he analyzed (Wright 1939).

The effect of two different migration models chosen to represent extreme scenarios is shown in Table 1. The "pollen packet" migration model assumes that m_p is equal to the proportion of pollen migrants that are *successful* in fertilization, whereas the "single pollen" migration model assumes that m_p is equal to the proportion of migrant pollen in the total pollen pool. Although the latter model seems more biologically plausible under most circumstances, the former was assumed by Wright since it is easier to model. Table 1 shows that when u , m_p and deme size are small, the number of alleles maintained in the total population for a given m_p is lower under the "single pollen" migration model than under the "pollen packet" migration model. This is because the average proportion of compatible pollen

within the deme is small when m_p is small, and migrating pollen have a larger probability of compatibility than nonmigrating pollen because demes are differentiated, leading to an increase in the “effective” migration rate under “single pollen” migration. Therefore, “single pollen” migration decreases the threshold migration rate below which subdivision has the effect of increasing the number of alleles above the value in an undivided population.

The results using a symmetrical overdominant model show that the pattern of an intermediate migration rate that maintains a minimum of alleles is not an effect of the specific kind of frequency-dependent selection under GSI, but rather an effect of balancing selection. This broadens the perspectives of the study to other systems than GSI, *e.g.*, sporophytic SI and MHC. Takahata (1993) used simulations similar to the ones described here, but he gives results only for $Nm \leq 1$, and did not observe an intermediate minimum number of alleles with intermediate migration rates. However, his finding that a substantial increase in the number of alleles happens only when $Nm \ll 0.1$ is in accordance with the present study.

To help understand why Wright’s theory fails, the stationary frequency distributions of an allele in a deme and in the total population were investigated (Figures 1, 4 and 5). While Wright’s theory accurately predicts the stationary distribution in an undivided population (Figure 1), it underestimates the variance of the stationary distribution for a subdivided population, and the discrepancy increases with decreasing migration. The stationary distribution within demes shows the striking pattern that at an intermediate migration rate ($m_p = 0.05$), it has a significant skew with a distribution dominated by low frequency alleles, whereas at the highest and lowest migration rates it is bell-shaped (Figure 5). Combining these observations, the conclusion is that at low migration rates (Figure 5, $m_p = 0.0001$ and 0.001), the variation in allele frequencies within demes is small, but the variation in the number of demes where the allele occurs is very large, violating Wright’s assumption 3. At intermediate migration rates (Figure 5, $m_p = 0.05$), on the other hand, the variance in the frequency of an allele in the total population is dominated by the variance within demes. This violates Wright’s assumption 2, which substitutes an allele’s frequency in a deme by its mean value, which is clearly not appropriate when the true distribution is highly asymmetrical with a large variance. At the high migration limit ($m_p = 1.0$), Figure 6 shows that Wright’s assumption 1 that an allele occupies a number of demes proportional to its frequency is violated. This is easily understood because when the migration rate is high a new allele is likely to spread to other demes before it has attained its equilibrium frequency in the deme, and therefore low frequency alleles are present in more demes than expected from

their population frequency under Wright’s assumption 1.

It appears that Wright’s theory fails for different reasons in different ranges of the migration spectrum. Unfortunately, it has not been possible to combine these observations into an explanation of why the number of alleles maintained in GSI (and symmetrical overdominance) has a minimum at intermediate migration rates, in contrast to neutrality. One important difference between balancing selection and neutrality is that with balancing selection the probability of invasion of a new mutant allele in a specific deme is dependent on the number of alleles already present in the deme, *i.e.*, the higher the number of alleles, the smaller the probability of invasion against genetic drift because selection for a rare allele is weaker. At the high migration limit, the population behaves (almost) as an undivided one and the probability of invasion is determined by the number of alleles in the total population. At the low migration limit ($m_p = 0.0001$), the fate of a new allele is almost completely determined by the probability of invasion into a deme, and although the demes contain few alleles, the total number of alleles is large because demes are highly differentiated. At intermediate migration rates, the probability of invasion of a new mutant allele is still primarily determined by its probability of invading the deme, but because $m_p \gg u$, migration maintains many more alleles in the deme than at the low migration limit, and the probability of invasion in the deme is reduced, while at the same time migration is high enough that demes are not sufficiently differentiated to contain many unique alleles.

The pattern for the effective number of alleles maintained was included mainly for comparison with neutral theory. The qualitative pattern as a function of the migration rate is similar to the pattern for the actual number of alleles. The quantitative differences arise because the variance in the stationary distribution of an allele in the total population increases with decreasing migration rate, and this has the effect of decreasing n_E more than n_a . Therefore, the difference between these two measures is largest for symmetrical overdominance, where selection is weaker than for GSI, resulting in a larger variance in the stationary distribution (Ewens 1964).

In *Papaver rhoeas*, the only species in which *S*-allele frequencies have been estimated from multiple natural populations, the variance in *S*-allele frequencies in local populations is indeed significantly larger than expected from Wright’s theory for an undivided population (Lawrence and O’Donnell 1981). Furthermore, a large overlap in allele specificities is present among four populations (Lawrence *et al.* 1993; Lane and Lawrence 1993). These results are consistent with the results of this study for a subdivided population with an intermediate migration rate ($m_p = 0.05$). In this case, the expected variance in allele frequencies within demes is very large (Figure 5), while at the same time the

numbers of alleles per deme, and in the total population, are 15 and 24, respectively; *i.e.*, a significant overlap in allele specificities among the demes is expected. Brooks *et al.* (1996) report only a small increase in the variance in the stationary frequency distribution of alleles in a model with restricted pollen and seed dispersal, but they only considered restricted dispersal within a local population and not migration between populations.

The numbers of alleles and their sequence divergence in species with GSI and for MHC-loci have been used to put broad bounds on the current (short-term) and long-term effective population sizes and the likelihood of population bottlenecks during speciation (Richman *et al.* 1996b and Richman and Kohn 1996 for GSI in Solanaceae; Takahata 1993 for MHC in humans). While this article does not address the estimation of the long-term effective population size, the results of Table 2 show that short-term effective sizes calculated from the total number of *S*-alleles in the population may be much smaller than the effective size for a neutral locus if the species in question has a subdivided population structure, as indeed, seems to be the case in the above-mentioned studies. Therefore, calculations of the short-term effective size from the number of *S*-alleles do not yield a precise estimate if the population is subdivided, in particular because the bias depends, qualitatively, on the migration rate. It should be stressed that natural populations are rarely expected to conform to the simple island model analyzed in this study, and that the conclusions cannot be directly extended to include the effect of geographic subdivision, *e.g.*, a stepping-stone model, on the number of *S*-alleles maintained. Still, the results show that the number of *S*-alleles (or MHC-alleles) maintained seems to be a good estimator of the number of individuals in the total population over a large range of migration rates ($Nm > 0.005$). For a subdivided population, the number of *S*-alleles (or MHC-alleles) may therefore provide a different and supplementary kind of information than can be obtained from the analysis of neutral variation.

This study was initially developed through discussions with M. Turilli. I thank F. B. Christiansen and X. Vekemans for numerous discussions and comments to previous versions of the manuscript, and D. Charlesworth, C. Damgaard and two anonymous reviewers for many valuable comments to the manuscript. X. Vekemans kindly provided a simulation program to test the results of my simulation program. The Department of Computer Sciences, University of Aarhus provided computing facilities. This study was supported by grants nos. 9400065 and 9401631 from the Danish Natural Science Research Council.

LITERATURE CITED

- Anderson, M. A., E. C. Cornish, S.-L. Mau, E. G. Williams, R. Hoggart *et al.*, 1986 Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana glauca*. *Nature* **321**: 38–44.
- Ayala, F. J., 1995 The myth of Eve: molecular biology and human origins. *Science* **270**: 1930–1936.
- Ayala, F. J., and A. A. Escalante, 1996 The evolution of human populations: a molecular perspective. *Mol. Phylogenet. Evol.* **5**: 188–201.
- Brooks, R. J., A. M. Tobias and M. J. Lawrence, 1996 The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. XI. The effects of limited pollen and seed dispersal, overlapping generations and variation in plant size on the variance of *S*-allele frequencies in populations at equilibrium. *Heredity* **76**: 367–376.
- Clark, A. G., 1993 Evolutionary inferences from molecular characterisation of self-incompatibility alleles, pp. 79–108 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Sinauer Associates, Inc., Sunderland, MA.
- Crosby, J. L., 1966 Self-incompatibility alleles in the population of *Oenothera organensis*. *Evolution* **20**: 567–579.
- Emerson, S., 1939 A preliminary survey of the *Oenothera organensis* population. *Genetics* **24**: 524–537.
- Ewens, W. J., 1964 The maintenance of alleles by mutation. *Genetics* **50**: 891–898.
- Ewens, W. J., and P. M. Ewens, 1966 The maintenance of alleles by mutation - Monte-Carlo results for normal and self-sterility populations. *Heredity* **21**: 371–378.
- Finney, D. J., 1952 The equilibrium of a self-incompatible polymorphic species. *Genetica* **26**: 33–64.
- Fisher, R. A., 1958 *The Genetical Theory of Natural Selection*. Dover Publications Inc., New York.
- Footo, H. C. C., J. P. Ride, V. E. Franklin-Tong, E. A. Walker, M. J. Lawrence *et al.*, 1994 Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas* L. *Proc. Natl. Acad. Sci. USA* **91**: 2265–2269.
- Ioerger, T. R., A. G. Clark and T.-H. Kao, 1990 Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proc. Natl. Acad. Sci. USA* **87**: 9732–9735.
- Kimura, M., and J. F. Crow, 1964 The number of alleles that can be maintained in a finite population. *Genetics* **49**: 725–738.
- Klein, J., J. Gutknecht and N. Fischer, 1990 The major histocompatibility complex and human evolution. *Trends Genet.* **6**: 7–11.
- Lane, M. D., and M. J. Lawrence, 1993 The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. VII. The number of *S*-alleles in the species. *Heredity* **71**: 596–602.
- Lawrence, M. J., and S. O'Donnell, 1981 The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. III. The number and frequency of *S*-alleles in two further natural populations (R102 and R104). *Heredity* **47**: 53–61.
- Lawrence, M. J., M. D. Lane, S. O'Donnell and V. E. Franklin-Tong, 1993 The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. V. Cross-classification of the *S*-alleles of samples from three natural populations. *Heredity* **71**: 581–590.
- Maruyama, M., 1970 Effective number of alleles in subdivided populations. *Theor. Popul. Biol.* **1**: 273–306.
- Maruyama, T., and M. Nei, 1981 Genetic variability maintained by mutation and overdominant selection in finite populations. *Genetics* **98**: 441–459.
- Mayo, O., 1966 On the problem of self-incompatibility alleles. *Biometrics* **22**: 111–120.
- Moran, P. A. P., 1962 *The Statistical Processes of Evolutionary Theory*. Clarendon Press, Oxford.
- Nagylaki, T., 1985 Homozygosity, effective number of alleles, and interdem differentiation in subdivided populations. *Proc. Natl. Acad. Sci. USA* **82**: 8611–8613.
- Nei, M., and N. Takahata, 1993 Effective population size, genetic diversity, and coalescence time in subdivided populations. *J. Mol. Evol.* **37**: 240–244.
- Richman, A. D., T.-H. Kao, S. W. Schaeffer and M. K. Uyenoyama, 1995 *S*-allele sequence diversity in natural populations of *Solanum carolinense* (Horsenettle). *Heredity* **75**: 405–415.
- Richman, A. D., and J. R. Kohn, 1996 Learning from rejection: the evolutionary biology of single-locus incompatibility. *TREE* **11**: 497–502.
- Richman, A. D., M. K. Uyenoyama and J. R. Kohn, 1996a *S*-allele diversity in a natural population of *Physalis crassifolia* (Solanaceae) (ground cherry) assessed by RT-PCR. *Heredity* **76**: 497–505.
- Richman, A. D., M. K. Uyenoyama and J. R. Kohn, 1996b Allelic

- diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. *Science* **273**: 1212–1216.
- Satta, Y., C. O'Huigin, N. Takahata and J. Klein, 1994 Intensity of natural selection at the major histocompatibility loci. *Proc. Natl. Acad. Sci. USA* **91**: 7184–7188.
- Takahata, N., 1990 Allelic genealogy and MHC polymorphism, pp. 267–286 in *Population Biology of Genes and Molecules*, edited by N. Takahata and J. F. Crow. Baifukan Co. Ltd., Tokyo.
- Takahata, N., 1993 Evolutionary genetics of human paleo-populations, pp. 1–22 in *Mechanisms of Molecular Evolution*, edited by N. Takahata and A. G. Clark. Sinauer Associates, Inc., Sunderland, MA.
- Takahata, N., and M. Nei, 1990 Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**: 967–978.
- Takahata, N., Y. Satta and J. Klein, 1992 Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics* **130**: 925–938.
- Vekemans, X., and M. Slatkin, 1994 Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* **137**: 1–9.
- Walker, E. A., J. P. Ride, S. Kurup, V. E. Franklin-Tong, M. J. Lawrence *et al.*, 1996 Molecular analysis of two functional homologues of the *S3* allele of the *Papaver rhoeas* self-incompatibility gene isolated from different populations. *Plant Mol. Biol.* **30**: 983–994.
- Wright, S., 1939 The distribution of self-sterility alleles in populations. *Genetics* **24**: 538–552.
- Wright, S., 1960 On the number of self-incompatibility alleles maintained in equilibrium by a given mutation rate in a population of given size: A reexamination. *Biometrics* **16**: 61–85.
- Wright, S., 1964 The distribution of self-incompatibility alleles in populations. *Evolution* **18**: 609–619.
- Yokoyama, S., and L. E. Hetherington, 1982 The expected number of self-incompatibility alleles in finite plant populations. *Heredity* **48**: 299–303.
- Yokoyama, S., and M. Nei, 1979 Population dynamics of sex-determining alleles in honey bees and self-incompatibility alleles in plants. *Genetics* **91**: 609–626.

Communicating editor: A. G. Clark