Evolutionary Dynamics of Sporophytic Self-Incompatibility Alleles in Plants

Mikkel H. Schierup,* Xavier Vekemans† and Freddy B. Christiansen*

**Department of Ecology and Genetics, University of Aarhus, DK-8000 Aarhus C., Denmark and †Université Libre de Bruxelles, B-1160 Brussels, Belgium*

Manuscript received February 11, 1997
Accepted for publication July 3, 1997

ABSTRACT

The stationary frequency distribution and allelic dynamics in finite populations are analyzed through stochastic simulations in three models of single-locus, multi-allelic sporophytic self-incompatibility. The models differ in the dominance relationships among alleles. In one model, alleles act codominantly in both pollen and style (SSIcod), in the second, alleles form a dominance hierarchy in pollen and style (SSIdom). In the third model, alleles interact codominantly in the style and form a dominance hierarchy in the pollen (SSIdomcod). The SSIcod model behaves similarly to the model of gametophytic self-incompatibility, but the selection intensity is stronger. With dominance, dominant alleles invade the population more easily than recessive alleles and have a lower frequency at equilibrium. In the SSIdom model, recessive alleles have both a higher allele frequency and higher expected life span. In the SSIdomcod model, however, loss due to drift occurs more easily for pollen-recessive than for pollen-dominant alleles, and therefore, dominant alleles have a higher expected life span than the more recessive alleles. The process of allelic turnover in the SSIdomcod and SSIdom models is closely approximated by a random walk on a dominance ladder. Implications of the results for experimental studies of sporophytic self-incompatibility in natural populations are discussed.

SINGLE-LOCUS, multi-allelic self-incompatibility (SI) systems in plants are examples of well defined allelic forms subject to frequency-dependent selection. Recently, the loci believed to be responsible for self-incompatibility (*S* loci) have been identified and cloned in both gametophytic systems (Solanaceae, Rosaceae, Papaveraceae) and in sporophytic systems (Brassicaceae) (see FRANKLIN *et al.* 1995 for a review). RICHMAN *et al.* (1995, 1996a) made the first surveys of nucleotide sequence variation in natural populations of *Solanum carolinense* and *Physalis crassifolia* (Solanaceae), and they attempted to infer the history and long-term effective population sizes of the species (RICHMAN *et al.* 1996b). Only 13 and 22 individuals were surveyed in the two species, but with further development of the methods, the analysis of larger samples from natural populations will soon be possible. With more accurate data on the number of alleles and on the distribution of allelic frequencies within and among populations, further information on the selection pertinent to a *S* locus can be obtained.

The single-locus, multi-allelic gametophytic self-incompatibility (GSI) system is found in a large number of species scattered throughout the phylogenetic tree of dicots (DE NETTANCOURT 1977; RICHARDS 1986). The incompatibility is determined by matching of the allele in the haploid pollen to the alleles in the diploid style, where the gene action is codominant. The stationary distribution of allelic frequencies and the allelic dynamics of GSI are well described in a large body of theory based on diffusion approximations, founded by WRIGHT (1939) and extended recently by YOKOYAMA and NEI (1979), CLARK (1993), and VEKEMANS and SLATKIN (1994). Data on the number of alleles and their relative frequency in natural populations are scarce, however. Number of alleles ranging from 13 to 45 have been reported (EMERSON 1939; ATWOOD 1942, 1944; LAWRENCE and O'DONNELL 1981; LEVIN 1993; RICHMAN *et al.* 1995) with concomitant maximum likelihood estimates of the total number of alleles in the species between 30 and 66 (LAWRENCE 1996). An exception to these figures is the estimated number of alleles in *Trifolium repens* and *T. pratense* that exceeds 100 (ATWOOD 1942; LAWRENCE 1996). The available frequency data seem to confirm the theoretical prediction that all alleles have equal frequency (EMERSON 1939; LEVIN 1993; RICHMAN *et al.* 1995, but see also CAMPBELL and LAWRENCE 1981).

Single-locus, multi-allelic sporophytic self-incompatibility (SSI) systems have been described in a number of species belonging primarily to three families, Asteraceae (GERSTEL 1950), Brassicaceae (BATEMAN 1954), and Convolvulaceae (MARTIN 1968), but are also found in species of Sterculiaceae (COPE 1962) and Caryophyllaceae (LUNDQUIST 1979, 1995). This scattered phylogenetic distribution of SSI may imply a polyphyletic origin. The SSI system is distinguished from the GSI system in that the pollen phenotype is determined by the diploid paternal plant, and the recognition process

Corresponding author: Mikkel Heide Schierup, Department of Ecology and Genetics, Building 540, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C., Denmark. E-mail: mheide@pop.bio.aau.dk

therefore involves interaction among two diploid genotypes, *i.e.*, four alleles. For two species of Caryophyllaceae, alleles appear to be codominant in both the pollen and the style (LUNDQUIST 1994, 1995). However, the vast majority of studies show dominance relationships between the alleles, and these may be complex and involve mutual weakening of allele specificities (RICHARDS 1986). However, codominant allele action seems to be most common in the stigma, whereas dominance relationships are more common in the pollen (RICHARDS 1986). In some species the alleles can be approximately sorted along a linear dominance hierarchy in the pollen with codominance in the style (OCKENDON 1974; STEVENS and KAY 1989). In other species a dominance hierarchy determines the phenotype of both pollen and style (COPE 1962; SAMAHA and BOYLE 1989; KOWYAMA *et al.* 1994).

Selection on the male side is gametic selection in GSI and sexual selection in SSI. A codominant SSI and a typical GSI are similar, in that the alleles in both systems are functionally equivalent, and therefore expected to occur in equal frequencies at equilibrium in the population. The typical SSI system, however, shows dominance resulting in asymmetry among alleles. This probably explains why population genetics theory for the SSI model is considerably less developed than the theory for GSI. A wide class of models of sex-symmetric dominance, however, shares an equilibrium property with the codominance models, in that the various self-incompatibility phenotypes occur in equal frequencies, an isoplethic equilibrium (FINNEY 1952; HEUCH 1979). BATEMAN (1952) derived the deterministic equilibrium frequencies for three alleles for the case of hierarchical dominance in pollen and codominance in style and for the case of an identical dominance hierarchy in both pollen and style. For the latter case, COPE (1962) determined the equilibrium allele frequencies for any number of alleles. IMRIE *et al.* (1972) presented the first numerical results on SSI in finite populations in a model without mutation. Recessive alleles reach higher frequencies than dominant alleles, the so-called "recessive effect," because they display their genotype less often in the pollen, and therefore selection against any allele becomes balanced when recessive alleles have higher frequencies (BATEMAN 1952; SAMPSON 1974). CHARLESWORTH (1988) investigated the stability of a self-incompatibility system by studying the introduction of a self-compatibility allele in populations with SSI. These theoretical investigations were done in deterministic models, and there has been virtually no investigation on the equilibrium frequencies and turnover of alleles at a SSI locus in finite populations under mutation pressure.

Estimates of the number of alleles in species with SSI are fewer and less precise than with GSI because of the complication of dominance. The reported number of alleles range from 10 to 49 in local populations, while estimates of the species-wide number of alleles range from 15 to 60 (SAMPSON 1967; STEVENS and KAY 1989; KARRON *et al.* 1990; DEVALL and THIEN 1992; KOWYAMA *et al.* 1994; FRANKLIN *et al.* 1995; LAWRENCE 1996). Little data on allelic frequencies exist, but the more recessive alleles appear to have the highest frequencies (OCKENDON 1974; STEVENS and KAY 1989; KOWYAMA *et al.* 1994).

We perform a numerical investigation of the stationary frequency distribution and allelic dynamics in finite populations in three models of SSI: (1) a model of codominance in both style and pollen (SSIcod), (2) a model with dominance hierarchy in both style and pollen (SSIdom), and (3) a model with codominance in the style and dominance hierarchy in the pollen (SSIdomcod). The last two models are studied in more detail as they are examples of asymmetrical frequency-dependent selection. We show that the allelic dynamics in the three models can be approximated by a random walk on a dominance ladder, and we compare the results to the well-known results from the GSI model.

MATERIALS AND METHODS

Models and deterministic derivation: The compatibility of pollen and style is determined by alleles $S_1, S_2, \ldots, S_n$ at one locus in a diploid plant species.

Codominant models: Two codominant models are considered, the sporophytic (SSIcod) and the gametophytic (GSI), for comparison. In the gametophytic model, alleles are codominant in the style and a pollen expresses only its own allele. In the sporophytic model, alleles again have independent action in the style and the pollen expresses both alleles of the paternal plants. In both models, a mating is compatible only if no alleles are shared. The SSIcod model corresponds to the case F described by BATEMAN (1952).

Hierarchical dominance models (SSIdom and SSIdomcod): In these models, the extant alleles in the population (S_1, $S_2, \ldots, S_n$) are sorted along a linear dominance hierarchy ($S_1 < S_2 < \cdots < S_n$) for determination of the pollen phenotype *i.e.*, allele S_1 is recessive to all other alleles, and allele S_n is dominant to all other alleles. When a new allele arises by mutation, it is placed at random in one of the $n + 1$ possible states (including boundaries) within the dominance hierarchy, and the relative dominance level of the extant alleles shifted accordingly. In SSIdomcod, the alleles are codominant in the style, and a mating is compatible if the pollen phenotype is different from each of the two alleles in the style. The style phenotype in the SSIdom model is determined by the same dominance hierarchy as in the anthers. The SSIdom and SSIdomcod models correspond, respectively, to cases J and H in BATEMAN's (1952) classification and to models DOM and IND in CHARLESWORTH's (1988) paper.

Deterministic equilibrium frequencies: In the gametophytic system, the only stable equilibrium frequency of an allele was recently proved to be the intuitive value of $1/n$, where n is the number of alleles (BOUCHER 1993; STEINER and GREGORIUS 1994). We conjecture the same to be true in the SSIcod model.

For the SSIdom and SSIdomcod models, deterministic recursion equations are given for any number of alleles in APPENDIX A. HEUCH (1979) showed that in the SSIdom model, a fully polymorphic equilibrium has equal proportions of phenotypes. The recursion equations were iterated until convergence to equilibrium starting from equal allele frequencies for the cases of three, four, five, six, and 11 alleles, and to

check the uniqueness and stability of the equilibrium point; the iterations were also performed starting from 10,000 random allele frequencies. The results were also checked by stochastic simulations of a large population (5000 individuals) with a fixed number of alleles, as described below.

Computer simulations: We simulated reproduction in a diploid plant population of size N with nonoverlapping generations using the Wright-Fisher model. In each generation progeny are produced by randomly choosing one of the $2N$ genes as the female gamete, one as the pollen and checking for compatibility according to the self-incompatibility model. If the pollen is compatible, then a new zygote is formed. If not, a new pollen is chosen randomly for the same female gamete until a compatible pollen is found. The process is repeated until N new zygotes are produced. A number of mutations, drawn from a Poisson distribution with mean $2Nu$, are then applied at random to genes in the zygotes. Each mutation is assumed to produce a new functional allelic type (the infinite alleles model). For the SSIdom and SSIdomcod models the dominance level of a new allele was given a random position in the hierarchy as described above.

Each run started with $2N$ different alleles in the population. A mutation-selection-drift equilibrium was approximated by continuing until the average number of alleles in the population over subsequent intervals of 20 generations had oscillated five times (usually within 2000 generations). Then, either a number of alleles in the population were followed throughout their lifetime to study the allelic dynamics, or the population was left to evolve for a number of generations equal to the observed time to equilibrium, and then the number of alleles, the expected heterozygosity and frequency distribution of all extant alleles were recorded.

Number of alleles, expected heterozygosity and allele counts: For each of three parameter sets ($N = 50$, $u = 2 \times 10^{-5}$; $N = 100$, $u = 10^{-4}$; $N = 200$, $u = 5 \times 10^{-4}$), the number of alleles, the expected heterozygosity or gene diversity, H (calculated as $\Sigma p_i p_j$, $i \neq j$ where p_i is the frequency of allele S_i), and the mean allele counts (mean number of gene copies of an allele) were recorded and averaged over 2000 replicate runs for each model.

Stationary frequency distribution in finite populations: For one parameter set ($N = 100$, $u = 10^{-4}$), the frequency distribution was computed as described above in 50,000 replicate simulations for each of the four models. In the SSIdom and SSIdomcod models the frequency distribution was recorded for all alleles pooled, and as a function of the dominance level.

Allelic dynamics: For each of three parameter sets ($N = 50$, $u = 2 \times 10^{-5}$; $N = 100$, $u = 10^{-4}$; $N = 200$, $u = 5 \times 10^{-4}$), 30,000 alleles were followed from the time they arose by mutation to their loss by genetic drift. For the SSIdom and SSIdomcod models, the following statistics were recorded for each allele: (1) its initial relative dominance level at the time of appearance, (2) its relative dominance level at the time it was lost from the population, (3) the total life span of the allele in generations, (4) its maximum frequency throughout its life span. For the codominant models, only (3) and (4) were recorded.

Criteria for invasion of new alleles: Only a fraction of new mutants successfully invade the population. Most are lost in the first few generations through genetic drift. Therefore, to aid the subsequent analysis of allelic dynamics, we chose a criterion for successful invasion. For the SSIdom and SSIdomcod models, an allele was considered successful and included in the analysis if the maximal frequency of the allele during its life span was larger than the expected deterministic equilibrium frequency for an allele at the corresponding initial dominance level (obtained from APPENDIX A). The same approach was used for the codominant models, but here the maximum frequency was compared to the expected equilibrium frequency, given the number of alleles at equilibrium. When recording distributions of life spans, only the alleles that met these criteria were used.

Analytical approximation of the hierarchical dominance models: The dynamics of the SSIdom and SSIdomcod models are modeled as a random walk on a dominance ladder in APPENDIX B. From the simulations we obtain the relative probabilities of invasion and loss of alleles at each dominance level, conditional on the allele being successful. From these probabilities, APPENDIX B provides an approximation of the expected life span for an allele at each dominance level.

From the simulations of allelic dynamics, the relative probabilities of invasion and loss for each dominance class were estimated as follows. The observed rounded average number of alleles for a given parameter set (n_{obs}) was used as the constant number of alleles. The 30,000 alleles ordered by initial dominance level were then equally divided into n_{obs} dominance classes, with the $30{,}000/n_{obs}$ alleles that had the lowest initial dominance level in the most recessive class, etc. For each of the n_{obs} classes, the relative probability of invasion was calculated according to the above criteria. The relative probability of loss of an allele was calculated similarly for each class by ordering alleles by their dominance level when leaving the population and dividing the alleles into corresponding dominance classes. The probabilities of invasion and loss were then used for construction of the transition matrix from which the expected life span (in number of allelic turnovers) could be estimated for alleles invading into each dominance class (APPENDIX B). These values can be transformed to the number of generations by multiplying by the average time span between allelic turnovers, which is $1/(2NuP_{intro})$, where P_{intro} is the probability of invasion for an allele with a random dominance level.

RESULTS

Deterministic equilibrium frequencies in the hierarchical dominance models: Equilibrium allele frequencies in the SSIdom and SSIdomcod models calculated from the deterministic recursion equations (APPENDIX A) for each dominance level are shown in Table 1 for three, four, five, six and 11 alleles. In each case, allele frequencies were found to converge to these values for 10,000 different sets of iterations where the initial allele frequencies were randomly chosen. Therefore, we conjecture that the frequencies in Table 1 are the only stable equilibria. Furthermore, allele frequencies estimated from stochastic simulations in a large population agreed very closely with the values of Table 1 (data not shown). The allele frequencies obtained for the SSIdom model satisfy the condition of equal phenotypic frequencies (HEUCH 1979) and are in agreement with COPE's (1962) formulas.

The equilibrium frequency of an allele decreases when its dominance level increases, but the unevenness in frequency between dominant and recessive alleles is substantially larger under SSIdom than under SSIdomcod. This is more pronounced with larger number of alleles, *e.g.*, with 11 alleles the most recessive allele has more than five times the frequency of the most dominant allele under SSIdom whereas the difference is <20% under SSIdomcod. The unevenness in frequency between dominant and recessive alleles increases with

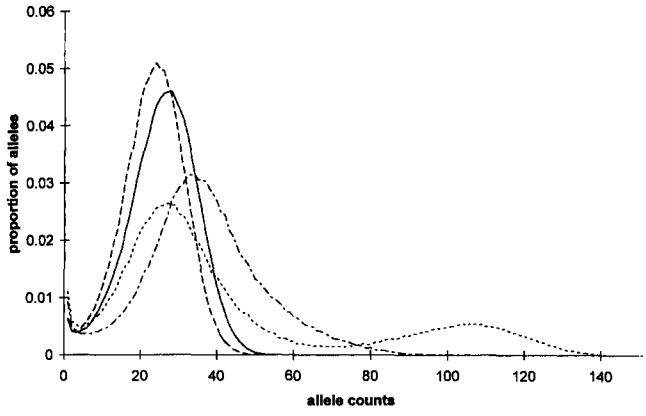

FIGURE 1.—Stationary frequency distribution for all four models and $N = 100$, $u = 10^{-4}$: GSI, —; SSIcod, ---; SSIdom, · · ·; SSIdomcod, -·-.

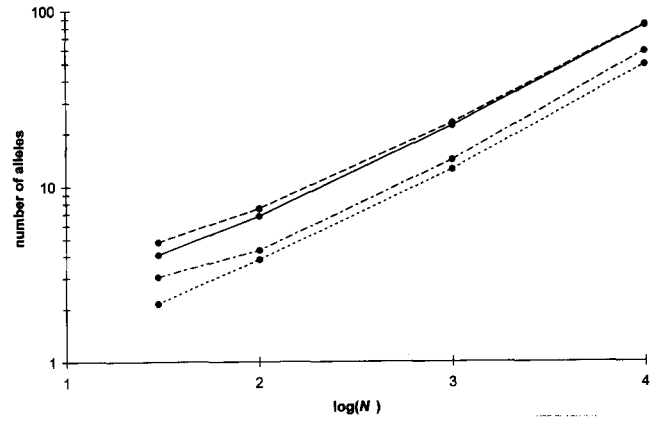

FIGURE 2.—Average number of alleles for all four models as a function of population size (N) with $u = 10^{-5}$: GSI, —; SSIcod, ---; SSIdom, · · ·; SSIdomcod, -·-.

increasing number of alleles in the SSIdom model, whereas it decreases under the SSIdomcod model. This unevenness in allele frequency is quantified as the frequency of the most recessive allele divided by the frequency of the most dominant allele in Table 1, bottom row.

Number of alleles and stationary frequency distributions in finite populations: In Figure 1, the stationary frequency distributions for the parameter set $N = 100$ and $u = 10^{-4}$ are shown for the three models of SSI and for GSI, and the corresponding mean number of alleles, mean expected heterozygosities, and mean allele counts are given in Table 2. The frequencies show a bell-shaped pattern of distribution with the exception of SSIdom that shows a bimodal distribution. The distribution for the SSIcod model is narrower and shifted toward lower frequencies as compared to the GSI model (Figure 1). The shift is caused by the higher mean number of alleles in the SSIcod model (Table 2). The stationary distribution of all alleles irrespective of their dominance level in the SSIdomcod model shows a much wider and flatter curve than GSI, and in the SSIdom model the distribution appears bimodal with a main peak in a range of frequencies similar to the distributions of the codominant models and a second peak at much higher frequencies. The number of alleles and expected heterozygosities observed in the SSIdom and SSIdomcod models are considerably smaller than in codominant models. The SSIdom model maintains only about half the number of alleles compared to the SSIcod model, and the average allele frequencies therefore are larger. Alleles at different dominance levels have different expected frequencies, and this causes the observed increase in the variance of allele counts (Table 2). Figure 2 gives the number of alleles maintained at equilibrium under a range of population sizes and a mutation rate of 10^{-5} per gene per generation. The codominant models consistently produce a higher number of alleles than the dominant models, the SSIcod model is higher than the GSI model, and the num-

TABLE 1

Deterministic equilibrium frequencies of the alleles in the two dominant models

Dominance level	Three alleles		Four alleles		Five alleles		Six alleles		11 alleles	
	SSIdom	SSIdomcod	SSIdom	SSIdomcod	SSIdom	SSIdomcod	SSIdom	SSIdomcod	SSIdom	SSIdomcod
1	0.610	0.487	0.525	0.336	0.467	0.254	0.425	0.204	0.310	0.102
2	0.223	0.286	0.200	0.261	0.181	0.220	0.167	0.186	0.125	0.099
3	0.167	0.227	0.150	0.215	0.137	0.194	0.126	0.170	0.095	0.097
4			0.125	0.187	0.114	0.174	0.106	0.157	0.080	0.095
5					0.100	0.158	0.093	0.146	0.070	0.093
6							0.083	0.137	0.063	0.090
7									0.058	0.089
8									0.054	0.087
9									0.051	0.085
10									0.048	0.083
11									0.045	0.081
FU	3.65	2.15	4.20	1.80	4.67	1.61	5.06	1.49	6.89	1.26

Allele frequency values are calculated from the recursion formulas in APPENDIX A for the cases of three, four, five, six and 11 alleles. Dominance level 1 is the most recessive. FU is the frequency unevenness calculated as the ratio of the frequencies of the most recessive allele to the most dominant allele.

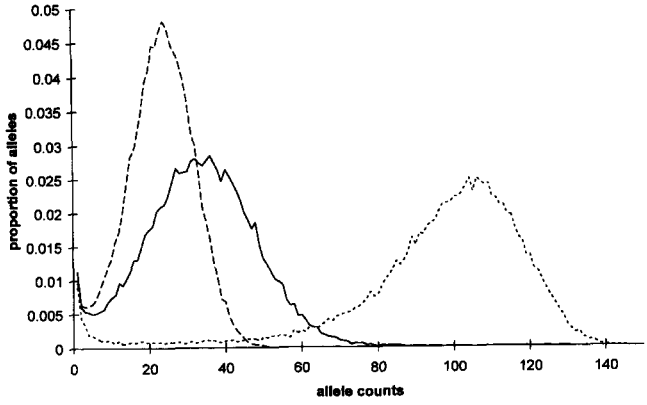

FIGURE 3.—Stationary frequency distribution of alleles belonging to different classes of dominance under SSIdom and $N = 100$, $u = 10^{-4}$: most dominant, ---; intermediate, —; most recessive, · · ·.

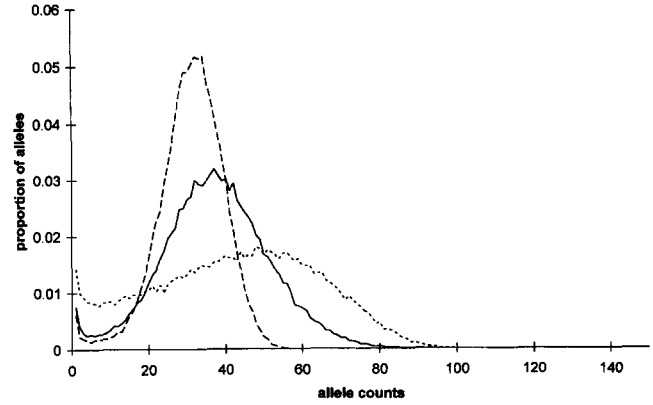

FIGURE 4.—Stationary frequency distribution of alleles belonging to different classes of dominance under SSIdomcod and $N = 100$, $u = 10^{-4}$: most dominant, ---; intermediate, —; most recessive, · · ·.

ber of alleles in the SSIdomcod model is higher than in the SSIdom model.

Figures 3 and 4 show more detail of the stationary frequency distribution in the dominant models. Figure 3 shows the stationary frequency distribution for the most recessive, an intermediate and the most dominant alleles in the SSIdom model and Figure 4 shows similar distributions for the SSIdomcod model. In both models, the dominant alleles show a bell-shaped distribution similar in mean and variance to alleles in the GSI model. The recessive alleles, however, show a distribution with a higher mean frequency and much larger variance. The unevenness in average frequency between the most dominant and the most recessive alleles is higher in SSIdom than in SSIdomcod, whereas the reverse relation holds for the variance of allele frequencies. The higher mean frequency for recessive alleles in finite populations is in agreement with the deterministic equilibrium frequencies (Table 1). An intermediate distribution is observed for alleles with an intermediate dominance level.

Allelic dynamics: For the case with $N = 100$ and $u = 10^{-4}$, Figure 5 shows the distribution of life span of alleles for the four models. Only alleles that have successfully invaded the population, according to our criteria, are included. The distributions for the codominant models are close to exponential. The frequency distributions for the dominant models are different, since the alleles have different dynamics depending on their current dominance level. Thus, separate distributions for the most recessive, intermediate and the most dominant alleles for the same parameter set are shown for SSIdom (Figure 6) and SSIdomcod (Figure 7). For the SSIdom model the life span distribution is shifted

TABLE 2

Results of numerical simulations in finite populations

Model	No. of alleles	H	Allele counts	Probability of invasion (%)	Average life span
$N = 50$, $u = 0.00002$					
GSI	5.10 ± 0.52	0.790 ± 0.019	19.61 ± 5.16	42.2	5940
SSIcod	5.80 ± 0.58	0.815 ± 0.017	17.23 ± 4.74	48.4	5753
SSIdom	2.96 ± 0.38	0.526 ± 0.076	33.80 ± 21.47	24.3	5148
SSIdomcod	3.44 ± 0.57	0.663 ± 0.055	29.06 ± 11.63	41.2	3773
$N = 100$, $u = 0.0001$					
GSI	7.80 ± 0.94	0.856 ± 0.014	25.65 ± 8.91	27.3	1482
SSIcod	8.62 ± 0.94	0.870 ± 0.012	23.20 ± 8.06	28.9	1453
SSIdom	4.52 ± 0.79	0.654 ± 0.064	44.21 ± 33.21	15.6	1346
SSIdomcod	5.46 ± 0.96	0.784 ± 0.035	36.66 ± 15.50	24.8	1089
$N = 200$, $u = 0.0005$					
GSI	14.78 ± 1.77	0.914 ± 0.008	27.07 ± 14.18	14.0	498
SSIcod	15.62 ± 1.85	0.919 ± 0.008	25.61 ± 13.22	16.0	459
SSIdom	9.38 ± 1.66	0.790 ± 0.048	42.63 ± 42.04	11.0	381
SSIdomcod	11.37 ± 1.79	0.879 ± 0.017	35.17 ± 21.55	11.9	462

The results are shown for three parameter sets in each of the four models. The columns show the average number of alleles, the expected heterozygosity H, the average allele counts, the probability of successful invasion by a mutant, and the average observed life span of a successful allele. Means and standard deviations are calculated for 2000 replicates.

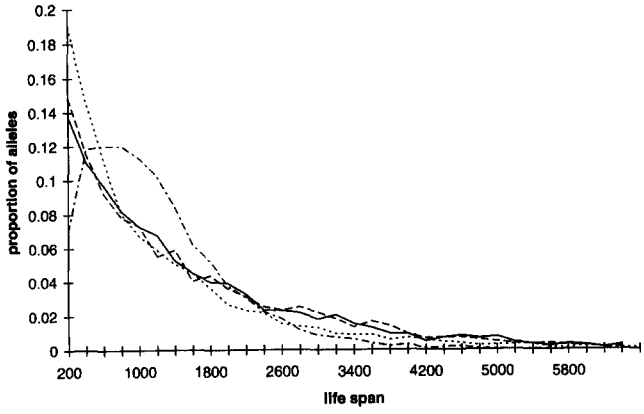

FIGURE 5.—Distribution of life span of alleles for all four models and $N = 100$, $u = 10^{-4}$: GSI, —; SSIcod, ---; SSIdom, · · ·; SSIdomcod, -·-.

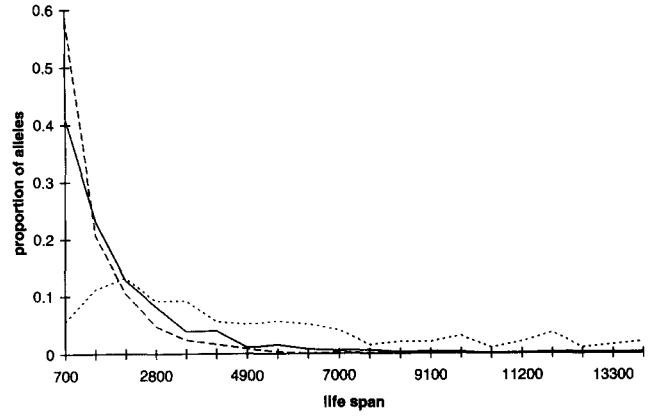

FIGURE 6.—Distribution of life span of alleles under SSIdom and $N = 100$, $u = 10^{-4}$: most dominant, ---; intermediate, —; most recessive, · · ·.

to lower values with increasing initial dominance level and is close to exponential except for the most recessive class that shows an intermediate peak. For the SSIdomcod model, the pattern is reversed with distributions shifted to the right with increased initial dominance level. In addition, the distributions for high dominance levels have a peak at intermediate values. Note, however, that the distribution of recessive alleles is based on fewer observations than the distributions of the dominant alleles, since the likelihood that a dominant allele invades the population is far greater. Our criterion for inclusion of an allele may seem restrictive, and therefore an alternative criterion using half the expected deterministic equilibrium frequency as the border value was tried. This criterion was found to include alleles with a very short life span. However, the increase in number of considered alleles was <5%, and although relatively more recessive alleles were included (because of their higher expected equilibrium frequency), the difference in observed average life spans of the various classes was very slight.

For each model, the probability of successful invasion by a new allele and the average observed life span of a successful allele are also given in Table 2 for the three parameter sets investigated ($N = 50$, $u = 0.00002$; $N = 100$, $u = 0.0001$; $N = 200$, $u = 0.0005$). The probabilities of successful invasion show the same pattern as the number of alleles, with the highest values in the SSIcod model and the lowest in the SSIdom model. The average life spans are smaller for the SSIcod than for the GSI model, and they are generally smaller in the hierarchical dominant than in the codominant models. The two hierarchical dominance models are compared to the random walk model of APPENDIX B in Table 3. The expected life span for each dominance class is generally within 10% of the observed life span for both dominant models. Table 3 shows, separately for alleles at each dominance level, the average observed and expected life spans of alleles, and the probability of invasion and loss conditional on a turnover event. The number of dominance classes for the analytical approximation was chosen as the rounded average number of alleles taken from Table 2. Because alleles at different dominance levels show different probabilities of invasion, the life span for all alleles in the dominant models in Table 2 is not a simple average of the life spans from each of the dominance classes in Table 3. In both hierarchical dominance models the probability of successful invasion is much higher for a new dominant allele than for a new recessive allele, and this effect is about equally strong in the two models. The probability of loss of alleles shows a very different pattern. In the SSIdom model, the probability of loss of a recessive allele is much lower than for more dominant alleles while the reverse is true in the SSIdomcod model. Moreover, the difference in probability of loss between recessive and dominant alleles is much higher in the SSIdomcod model than in the SSIdom model.

The observed life span of an allele decreases monotonically with its initial dominance level in the SSIdom model, and the most recessive alleles have an observed life span up to five times as long as the most dominant alleles (Table 3). The situation is reversed in the SSIdomcod model. The alleles introduced as dominant in the SSIdomcod model (and as recessive in the SSIdom

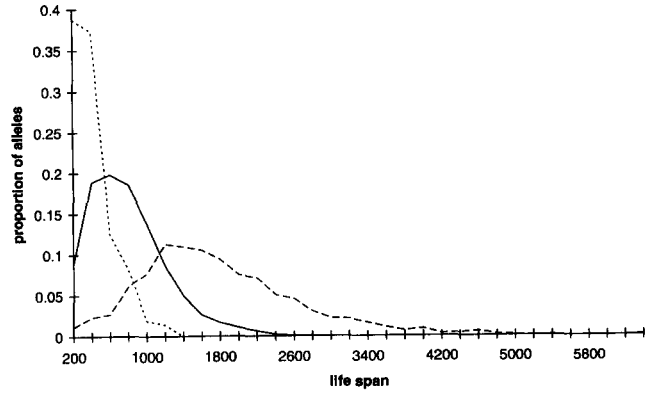

FIGURE 7.—Distribution of life span of alleles under SSIdomcod and $N = 100$, $u = 10^{-4}$: most dominant, ---; intermediate, —; most recessive, · · ·.

TABLE 3

The process of invasion and loss in the two dominant models

Dominance level	SSIdomcod model			SSIdom model		
	Probability of invasion	Probability of loss	Life span obs. (exp.)	Probability of invasion	Probability of loss	Life span obs. (exp.)
$N = 50$, $u = 0.00002$						
All			3773			5148
1	0.075	0.951	1146 (1279)	0.061	0.090	18034 (16703)
2	0.400	0.046	2411 (2530)	0.413	0.366	5121 (6581)
3	0.525	0.003	5188 (4822)	0.526	0.544	3675 (4643)
$N = 100$, $u = 0.0001$						
All			1089			1346
1	0.033	0.691	307 (294)	0.039	0.067	5393 (3629)
2	0.123	0.218	421 (453)	0.247	0.221	1599 (1600)
3	0.215	0.053	652 (674)	0.331	0.339	1170 (1121)
4	0.287	0.024	1063 (976)	0.382	0.373	919 (960)
5	0.342	0.012	1723 (1518)			
$N = 200$, $u = 0.0005$						
All			462			381
1	0.020	0.261	195 (171)	0.022	0.018	1544 (1557)
2	0.028	0.181	191 (204)	0.070	0.055	599 (747)
3	0.039	0.139	245 (234)	0.085	0.090	501 (520)
4	0.055	0.115	272 (264)	0.109	0.109	390 (426)
5	0.072	0.076	311 (302)	0.128	0.133	352 (371)
6	0.084	0.071	324 (339)	0.120	0.141	267 (340)
7	0.103	0.052	356 (384)	0.159	0.148	347 (323)
8	0.124	0.038	416 (440)	0.159	0.149	299 (312)
9	0.135	0.029	469 (511)	0.149	0.157	274 (299)
10	0.155	0.023	599 (610)			
11	0.186	0.015	725 (786)			

Results are shown for three parameter sets for the two dominant models. The first column refers to dominance level, with "1" being most recessive; "All" refers to the average over all alleles. Probabilities of successful invasion and loss for each dominance level are conditional on the occurrence of an allelic turnover. The observed life spans refer to the values from the simulations, and the expected life spans are values calculated from the random walk model in APPENDIX B using the probabilities of invasion and loss in the table.

model) generally have larger life spans than alleles in the codominant models.

DISCUSSION

Our models of sporophytic self-incompatibility are extremely simplified. The limited number of experimental investigations provides examples of both dominance and codominance of the SI alleles and of differences in dominance relations in the expression in style and anthers. Thus, we model the observed phenomena by simplified situations. Either alleles are codominant in the anther (SSIcod) or alleles show different levels of dominance in the anther, arranged in a dominance hierarchy (SSIdom, SSIdomcod). Alleles are either codominant in the style (SSIcod, SSIdomcod) or express a dominance hierarchy identical to the anther relationships (SSIdom). Differences between anther and style expression are only modeled as dominance in the anther and codominance in the style (SSIdomcod), because the opposite has not been observed. The situation for many species may be somewhat intermediate, in that several dominance levels each with a number of codominant alleles may exist (RICHARDS 1986; STEVENS and KAY 1989; KOWYAMA *et al.* 1994). Recent molecular studies of Brassicaceae have shown that more genes are involved in SI specificity, but the responsible genes were found to be closely linked and most likely to evolve as a single nonrecombining unit (BOYES *et al.* 1997).

The SSIcod model: The model with codominance in both anther and style was found to behave similarly to the well studied GSI model. Allele frequencies show narrow stationary distributions, and allelic life spans are approximately exponentially distributed. These properties are expected in a model where all alleles are functionally equivalent. However, the minimum number of alleles in a population with SSIcod is four, as opposed to three in the GSI model, and SSIcod leads to a much lower average cross-compatibility than GSI in populations with few alleles (BATEMAN 1952). This argues for stronger selection in SSIcod than in GSI, and we found it expressed in finite populations in the SSIcod model as a larger number of alleles maintained, a higher probability of successful invasion of alleles, and a more narrow stationary frequency distribution. Selection in the GSI model is stronger than selection in an overdominant model with lethal homozygotes (VEKEMANS and SLATKIN 1994), and so the SSIcod model appears as an

even more extreme example of balancing selection. The average life span of alleles was found to be lower in SSIcod than in GSI. We see this as a result of a higher probability of invasion by a new allele in the SSIcod model caused by stronger selection.

The SSIdom model: Iteration of the recursion equations (APPENDIX A) confirmed the existence and stability of the unique isoplethic equilibrium (COPE 1962; HEUCH 1979). Similar results for equilibrium frequencies were obtained by BYERS and MEAGHER (1992) in stochastic simulations of SSIdom in a large population for three to nine alleles. The deterministic results confirm SAMPSON's (1974) "recessive effect" that recessive alleles are found at higher frequencies. The unevenness in frequency between recessive and dominant alleles is increasing with the number of alleles.

In finite populations the average number of alleles at equilibrium is much lower than under the GSI and SSIcod models. This supports the argument by BYERS and MEAGHER (1992) that small populations are unable to maintain a high diversity of alleles under SSIdom. As the equilibrium number of alleles shows weak dependence on the mutation rate (VEKEMANS and SLATKIN 1994), the expected number of alleles seems mainly to depend on population size and on the genetic determination of self-incompatibility.

In the SSIdom model, rare recessive alleles are affected by average selection coefficients that are lower than for rare dominant alleles, because a rare recessive allele rarely displays its genotype. A rare recessive allele is therefore primarily affected by random genetic drift (see Figure 3). A sufficiently common recessive allele, however, forms homozygotes frequently, so that selection becomes effective and tends to keep the allele near the deterministic equilibrium. The probability of invasion increases with increasing level of dominance because a rare dominant allele is subject to stronger selection. The probability of loss of an allele from the population also increases with dominance level because of an increased sensitivity to drift due to a smaller equilibrium frequency (BYERS and MEAGHER 1992). The combined result is that the expected life span of alleles decreases with increasing level of dominance. During an allelic turnover there is therefore a relatively large probability that a dominant allele enters and a dominant allele exits the population. Any allele will spend most of its life in the population at or near its initial dominance level and therefore its probability of extinction at an allelic turnover is fairly constant. This resembles the situation for an allele in the codominant models, and it explains why the life span distributions for each initial dominance level (Figure 6) approximated exponential distributions.

COPE (1962) argued that recessive alleles always have a selective advantage because they reach higher equilibrium frequencies. IWASA and SASAKI (1987) challenged this view based on their deterministic analysis showing that recessive alleles do not invade the population as easily as dominant ones. Our analysis shows that both effects are important for the dynamics and equally strong in the sense that the absolute dominance level of the extant alleles does not evolve in a specific direction.

The SSIdomcod model: Deterministic equilibrium frequencies in the SSIdomcod model were investigated by BATEMAN (1952) for two and three alleles, and IMRIE *et al.* (1972) for three and six alleles. Their results for three alleles differed. Our results, based on iterations of the recursion equations (APPENDIX A), agree with Bateman's results, and we show in a separate paper that the model of IMRIE *et al.* (1972) corresponds to a modification of the SSIdomcod model that introduces differential maternal fecundities (X. VEKEMANS, M. H. SCHIERUP and F. B. CHRISTIANSEN, unpublished results).

The recessive alleles reach higher frequencies than dominant alleles in the deterministic SSIdomcod model, but the unevenness in frequency between most recessive and most dominant alleles is smaller than in the SSIdom model. Moreover, this unevenness in frequency decreases with increasing number of alleles in the SSIdomcod model whereas it increases in the SSIdom model.

The number of alleles maintained in finite populations under SSIdomcod is intermediate between the number of alleles maintained in the SSIcod and SSIdom models. However, in the SSIdomcod model the stationary frequency distributions and the patterns of allelic turnover have unique properties that are not intermediate to those of the SSIcod and SSIdom models.

In the stationary frequency distributions, the variance decreases strongly with the level of dominance, *i.e.*, the more recessive alleles show much more variation in frequency than the more dominant alleles. In particular, the probability of having a low allele frequency is large for recessive alleles, and they are therefore more sensitive to genetic drift. This explains why, despite their higher expected equilibrium frequency, recessive alleles have a much shorter expected life span than dominant alleles. This result is the opposite of the results from the SSIdom model. The reason for the higher sensitivity of recessive alleles to genetic drift under SSIdomcod can be intuitively described as follows. A rare recessive allele is very close to neutral in both dominant models, because there is no selection on alleles in the style and a rare recessive allele is never expressed in the pollen phenotype. Thus, when the allele becomes common, homozygotes are formed that express their phenotype in the pollen. However, in the SSIdomcod model this happens at an allele frequency where the allele is already commonly expressed in the style, where allele interaction is codominant, and positive selection for the recessive allele is therefore diminished. Hence, a recessive allele never experiences selection as strong as selection on a dominant allele. In the SSIdom model, the phenotype corresponding to a recessive allele is formed equally often in pollen and style, and so pollen

of the recessive phenotype does not encounter its phenotype in the style as frequently as in the SSIdomcod model. Thus, the difference between the SSIdom and SSIdomcod models for a recessive allele is caused primarily by different selection forces when the allele is common. For rare recessive alleles the probabilities of invasion are very similar in the two models (Table 3). As in the SSIdom model, the probability of invasion by an allele under SSIdomcod increases monotonically with its dominance level. The probability of loss of an allele, however, decreases monotonically with its dominance level. The net effect is that a new allele tends to have a high dominance level, then it experiences a decrease in its dominance relative to extant alleles in the population, and it tends to exit at a low dominance level. This process also contributes to the shapes of the distributions of life spans as a function of initial dominance level (Figure 7), since the life span of a given dominant allele is limited primarily by invasion by new and more dominant alleles.

Through time, the extant alleles will evolve toward ever increasing levels of absolute dominance. An upper limit may exist for the absolute dominance level. Our mutation model therefore may be unrealistic, because it assumes that mutations to any level in the dominance hierarchy are equally likely and independent of the dominance of the gene in which the mutation happens. An alternative, simplified mutation model assuming an absolute dominance level, however, is more complex to analyze because, after some time, most alleles are expected to be close to the maximal dominance level and almost all new mutations would generate recessive alleles quickly lost from the population. In our model, this would amount to a change in the effective mutation rate through time and to a development of heterogeneity in mutation rate among allelic classes. If a limit to dominance exists, our result that dominant alleles are more easily maintained in the population suggests that the most dominant alleles should have an even longer life span than in our simulations, because they are less likely to be replaced by incoming dominant mutations.

Approximation of the allelic dynamics in the dominant models: In the random walk model of APPENDIX B, we have presented an analytical approximation of the allelic dynamics in the dominance models. The approximation is based on the assumption that invasion and loss of an allele is independent of the number and frequencies of other alleles in the population and only dependent on the dominance level of the allele in focus. The model therefore views evolution of an allele as a random walk on a dominance ladder. In small populations as studied here, the stochastic fluctuations in number and frequencies of alleles are dramatic (Table 2 and Figure 4). Nevertheless, expected life spans for each dominance level as determined from the random walk model agreed closely with the observed values (Table 3). Therefore, as a reasonable first approximation, the dynamics may be viewed as a process in the population of alleles rather than a process in the population of genes. A similar approximation was obtained by TAKAHATA (1990) in his theory of allelic genealogies in the simpler situation of selection by symmetric overdominant viabilities.

Experimental implications: Our study shows that the equilibrium frequencies and allelic dynamics in finite populations with SSI are highly dependent on the dominance relationships among alleles. Therefore, a thorough understanding of the genetics of SSI in a given species is a prerequisite for any evolutionary inference about the system. The only way to investigate the genetics of a system is by scoring for compatibility reaction in controlled crosses. However, the use of molecularly defined alleles (determined *e.g.*, through nucleotide sequences or RFLP-typing) may greatly reduce the number of crosses that are necessary to establish the exact dominance relationships. Hence, only individuals with a new molecularly defined allele need to be characterized phenotypically by crosses to a set of tester plants of already known phenotypes. This design would also test the hypothesis that a molecularly defined allele corresponds to an allele with a phenotypically distinct effect. In the only studies of natural populations using molecular determination of alleles in the GSI system, RICHMAN *et al.* (1995, 1996a) did not confirm that all of their molecularly defined alleles were functionally distinct. However, the sequences of their alleles were very divergent, and it is therefore extremely unlikely that any two of their molecularly defined alleles would belong to the same functional allele (VEKEMANS and SLATKIN 1994). With simpler molecular markers (*e.g.*, RFLP-typing) of alleles, only limited information on the genetic divergence between molecularly defined alleles is obtained, and two molecularly defined alleles may well belong to the same functional allele. In the SSIdom model, for instance, recessive alleles have a much longer life span and a higher population frequency than dominant alleles, so we expect much deeper genealogies of gene copies sampled within individual recessive alleles than among genes of a dominant allele.

If detailed information about the dominance relationships among alleles is at hand, the analysis of the models presented here provides testable predictions of *e.g.*, the expected variation in the average and variance of allele frequency as a function of the dominance level.

We thank D. CHARLESWORTH for comments and two anonymous reviewers for suggestions to improve the manuscript. The study was supported by grants no. 9400065 (M.H.S) and 94-0163-1 (F.B.C.) from the Danish Natural Science Research Council, and a travel grant from the European Science Foundation Scientific Programme in Population Biology (M.H.S.)

LITERATURE CITED

ATWOOD, S. S., 1942 Oppositional alleles causing cross-incompatibility in *Trifolium repens*. Genetics **27:** 333–338.

ATWOOD, S. S., 1944 Oppositional alleles in natural populations of *Trifolium repens*. Genetics **29:** 428–435.

BATEMAN, A. J., 1952 Self-incompatibility systems in angiosperms. 1. Theory. Heredity **6:** 285–310.

BATEMAN, A. J., 1955 Self-incompatibility systems in angiosperms. III. Cruciferae. Heredity **9:** 53–68.

BOUCHER, W., 1993 A deterministic analysis of self-incompatibility alleles. J. Math. Biol. **31:** 149–155.

BOYES, D. C., M. E. NASRALLAH, J. VREBALOV and J. B. NASRALLAH, 1997 The self-incompatibility (*S*) haplotypes of Brassica contain highly divergent and rearranged sequences of ancient origin. Plant Cell **9:** 237–247.

BYERS, D. L., and T. R. MEAGHER, 1992 Mate availability in small populations of plant species with homomorphic sporophytic self-incompatibility. Heredity **68:** 353–359.

CAMPBELL, J. M., and M. J. LAWRENCE, 1981 The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. II. The number and frequency of S-alleles in a natural population (R106). Heredity **46:** 81–90.

CHARLESWORTH, D., 1988 Evolution of homomorphic sporophytic self-incompatibility. Heredity **60:** 445–453.

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, Sunderland, MA.

COPE, F. W., 1962 The effects of incompatibility and compatibility on genotype proportions in populations of *Theobroma cacao* L. Heredity **17:** 183–195.

DE NETTANCOURT, D., 1977 *Incompatibility in Angiosperms.* Springer Verlag, Berlin.

DEVALL, M. S., and L. B. THIEN, 1992 Self-incompatibility in *Ipomoea pes-caprea* (Convolvulaceae). Am. Midl. Nat. **128:** 22–29.

EMERSON, S., 1939 A preliminary survey of the *Oenothera organensis* population. Genetics **24:** 524–537.

FINNEY, D. J., 1952 The equilibrium of a self-incompatible polymorphic species. Genetica **26:** 33–64.

FRANKLIN, F. C. H., M. J. LAWRENCE and V. E. FRANKLIN-TONG, 1995 Cell and molecular biology of self-incompatibility in flowering plants. Int. Rev. Cytol. **158:** 1–64.

GERSTEL, D. U., 1950 Self-incompatibility studies in Guayule. II. Inheritance. Genetics **35:** 482–506.

HEUCH, I., 1979 Equilibrium populations of heterostylous plants. Theoret. Popul. Biol. **15:** 43–57.

IMRIE, B. C., C. J. KIRKMAN and D. R. ROSS, 1972 Computer simulation of a sporophytic self-incompatibility breeding system. Australian J. Biol. Sci. **25:** 343–349.

IWASA, Y., and A. SASAKI, 1987 Evolution of the number of sexes. Evolution **41:** 49–65.

KARRON, J. D., D. L. MARSHALL and D. M. OLIVERAS, 1990 Numbers of sporophytic self-incompatibility alleles in populations of wild radish. Theoret. Appl. Genet. **79:** 457–460.

KOWYAMA, Y., H. TAKAHASI, K. MURAOKA, T. TANI, K. HARA *et al.,* 1994 Number, frequency and dominance relationships of S-alleles in diploid *Ipomoea trifida.* Heredity **73:** 275–283.

LAWRENCE, M. J., 1996 Number of incompatibility alleles in clover and other species. Heredity **76:** 610–615.

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.

LEVIN, D. A., 1993 S-gene polymorphism in *Phlox drummondi.* Heredity **71:** 193–198.

LUNDQUIST, A., 1979 One-locus sporophytic self-incompatibility in the carnation family, Caryophyllaceae. Hereditas **91:** 307.

LUNDQUIST, A., 1994 "Slow" and "quick" S-alleles without dominance interaction in the sporophytic one-locus self-incompatibility system of *Stellaria holostea* (Caryophyllaceae). Hereditas **120:** 191–202.

LUNDQUIST, A., 1995 Concealed genes for self-incompatibility in the carnation family Caryophyllaceae? Hereditas **122:** 85–89.

MARTIN, F. W., 1968 The system of self-incompatibility in *Ipomoea.* J. Hered. **59:** 263–267.

OCKENDON, D. J., 1974 Distribution of self-incompatibility alleles and breeding structure of open-pollinated cultivars of Brussels sprouts. Heredity **33:** 159–171.

RICHARDS, A. J., 1986 *Plant Breeding Systems.* Unwin Hyman, London.

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., 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.

SAMAHA, R. R., and T. H. BOYLE, 1989 Self-incompatibility of *Zinnia angustifolia* HBK (Compositae): II. Genetics. J. Hered. **80:** 368–372.

SAMPSON, D. R., 1967 Frequency and distribution of self-incompatibility alleles in *Raphanus raphanistrum.* Genetics **56:** 241–251.

SAMPSON, D. R., 1974 Equilibrium frequencies of sporophytic self-incompatibility alleles. Can. J. Genet. Cytol. **16:** 611–618.

STEINER, W., and H.-R. GREGORIUS, 1994 Single-locus gametophytic incompatibility: the symmetric equilibrium is globally asymptotically stable. J. Math. Biol. **32:** 515–520.

STEVENS, J. P., and Q. O. N. KAY, 1989 The number, dominance relationships and frequencies of self-incompatibility alleles in a natural population of *Sinapis arvensis* L. in South Wales. Heredity **62:** 199–205.

TAKAHATA, N., 1990 A simple genealogical structure of strongly balanced allelic lines and *trans*-species evolution of polymorphism. Proc. Natl. Acad. Sci. USA **87:** 2419–2423.

VEKEMANS, X., and M. SLATKIN, 1994 Gene and allelic genealogies at a gametophytic self-incompatibility locus. Genetics **137:** 1157–1165.

WRIGHT, S., 1939 The distribution of self-sterility alleles in populations. Genetics **24:** 538–552.

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: M. K. UYENOYAMA

APPENDIX A: RECURRENCE EQUATIONS

Male dominance, female codominance: Consider a sporophytic self-incompatibility system of n alleles $S_1 \ldots S_n$ with the dominance relation $S_1 < S_2 < \cdots < S_n$ in anthers and codominance in the stigma. The anther phenotype of an S_kS_l plant is $\max(k, l)$, and so the cross S_iS_j (macrosporophyte parent) $\times$ S_kS_l (microsporophyte parent) is fertile when $i \neq \max(k, l)$ and $j \neq \max(k, l)$. The homozygote S_iS_i occurs in frequency z_{ii} and the heterozygote S_iS_j occurs in frequency $2z_{ij}$, $i, j = 1, 2, \ldots, n$. The frequency of allele S_i is p_i,

$$p_i = \sum_{j=1}^{n} z_{ij} \quad \text{and} \quad \sum_{i=1}^{n}\sum_{j=1}^{n} z_{ij} = 1.$$

The population frequency of pollen that are fertile on an S_iS_j stigma is

$$\pi_{ij} = 1 - 2\sum_{k=1}^{i-1} z_{ik} - 2\sum_{k=1}^{j-1} z_{jk} - z_{ii} - z_{jj}$$

for $i \neq j$ and

$$\pi_{ii} = 1 - 2\sum_{k=1}^{i-1} z_{ik} - z_{ii}$$

for a homozygote.

The probability of fertilization of a macrosporophyte is assumed to be independent of the frequencies of genotypes in the population. The reciprocal of the frequencies π_{ij}, $i, j = 1, 2, \ldots, n$, therefore is the set of male fitnesses on the various stigmas.

Let $z^*_{\alpha\beta}$ be the frequency of fusion of an S_α macrosporophyte and an S_β microsporophyte in the population, that is, the frequency of production of an $S_\alpha S_\beta$ zygote by way of an S_β pollen. Given these frequencies of fusion, the recurrence equations of the population frequencies are given by

$$z'_{\alpha\beta} = \tfrac{1}{2}(z^*_{\alpha\beta} + z^*_{\beta\alpha}).$$

The expression for the fusion frequencies depends on the relation between the alleles in that for $\alpha < \beta$ we get

$$z^*_{\alpha\beta} = \sum_{i=1}^{\beta-1} \frac{z_{i\alpha}}{\pi_{i\alpha}} p_\beta + \frac{z_{\alpha\beta}}{\pi_{\alpha\beta}} \left(p_\beta - \sum_{k=1}^{\beta} z_{k\beta} \right) + \sum_{i=\beta+1}^{n} \frac{z_{i\alpha}}{\pi_{i\alpha}} (p_\beta - z_{i\beta}).$$

For homozygotes we get

$$z^*_{\alpha\alpha} = \sum_{i=1}^{\alpha} \frac{z_{i\alpha}}{\pi_{i\alpha}} \left(p_\alpha - \sum_{k=1}^{\alpha} z_{k\alpha} \right) + \sum_{i=\alpha+1}^{n} \frac{z_{i\alpha}}{\pi_{i\alpha}} \left(p_\alpha - \sum_{k=1}^{\alpha} z_{k\alpha} - z_{i\alpha} \right)$$

and for $\alpha > \beta$ we get

$$z^*_{\alpha\beta} = \sum_{i=1}^{\beta-1} \frac{z_{i\alpha}}{\pi_{i\alpha}} (p_\beta - z_{\alpha\beta}) + \frac{z_{\alpha\beta}}{\pi_{\alpha\beta}} \left(p_\beta - \sum_{k=1}^{\beta} z_{k\beta} - z_{\alpha\beta} \right) + \sum_{i=\beta+1}^{\alpha-1} \frac{z_{i\alpha}}{\pi_{i\alpha}} (p_\beta - z_{\alpha\beta} - z_{i\beta}) + \frac{z_{\alpha\alpha}}{\pi_{\alpha\alpha}} (p_\beta - z_{\alpha\beta}) + \sum_{i=\alpha+1}^{n} \frac{z_{i\alpha}}{\pi_{i\alpha}} (p_\beta - z_{\alpha\beta} - z_{i\beta}).$$

Dominance in both sexes: Now assume that the dominance relation $S_1 < S_2 < \cdots < S_n$ holds in both anthers and stigma. The anther and stigma phenotye of an $S_k S_l$ plant is $\max(k, l)$, and so the cross $S_i S_j \times S_k S_l$ is fertile when $\max(i, j) \neq \max(k, l)$. The frequency of phenotype i in the population is

$$\zeta_i = z_{ii} + 2 \sum_{j=1}^{i-1} z_{ij}.$$

The population frequency of pollen that are fertile on an $S_i S_j$, $j = 1, \ldots, i$, stigma is $\pi_i = 1 - \zeta_i$. Again, the probability of fertilization of a macrosporophyte is assumed to be independent of the frequencies of genotypes in the population.

The fusion frequencies for $\alpha < \beta$ are given by

$$z^*_{\alpha\beta} = \sum_{i=1}^{\alpha} \frac{z_{i\alpha}}{\pi_\alpha} p_\beta + \sum_{i=\alpha+1}^{\beta-1} \frac{z_{i\alpha}}{\pi_i} p_\beta + \frac{z_{\alpha\beta}}{\pi_\beta} \left(p_\beta - \sum_{k=1}^{\beta} z_{k\beta} \right) + \sum_{i=\beta+1}^{n} \frac{z_{i\alpha}}{\pi_i} (p_\beta - z_{i\beta}).$$

For homozygotes we get

$$z^*_{\alpha\alpha} = \sum_{i=1}^{\alpha} \frac{z_{i\alpha}}{\pi_\alpha} \left(p_\alpha - \sum_{k=1}^{\alpha} z_{k\alpha} \right) + \sum_{i=\alpha+1}^{n} \frac{z_{i\alpha}}{\pi_i} (p_\alpha - z_{i\alpha})$$

and for $\alpha > \beta$ we get

$$z^*_{\alpha\beta} = \sum_{i=1}^{\alpha} \frac{z_{i\alpha}}{\pi_\alpha} (p_\beta - z_{\alpha\beta}) + \sum_{i=\alpha+1}^{n} \frac{z_{i\alpha}}{\pi_i} (p_\beta - z_{i\beta}).$$

APPENDIX B: RANDOM WALK ON DOMINANCE LADDER

Consider a population containing n alleles $S_1 \ldots S_n$ with the dominance relation $S_1 < S_2 < \cdots < S_n$. The alleles exit the population from time to time and when an exit occurs, allele S_i exits with probability d_i. Immediately following the exit of one allele, another allele enters, and this allele is of type S_j with probability e_j. Thus, $\Sigma_{i=1}^n d_i = 1$ and $\Sigma_{j=1}^n e_j = 1$.

The dominance level of an allele in the population may change every time an exit event occurs. For a given allele, let X_t be its state after the tth exit event following its entrance into the population, and the state of an allele that exits as S_i is $-i$, that is, $\text{Prob}(X_{t+1} = -i \mid X_t = i) = d_i$ and $\text{Prob}(X_{t+1} = -i \mid X_t = -i) = 1$. The goal is to find the ultimate fate of the alleles, that is to find $\text{Prob}(X_\infty = -j \mid X_0 = i)$, and the expected time an allele will reside in the population.

Consider an allele in state i, and assume that an exit event occurs. Then S_i exits with probability d_i. If allele S_k, $k \neq i$, exits then

$$\begin{cases} i \mapsto i-1 & \text{when } k < i: \quad \text{probability } D_i, \\ i \mapsto i & \text{when } k > i: \quad \text{probability } 1 - D_i - d_i, \end{cases}$$

where

$$D_i = \sum_{j=1}^{i-1} d_j$$

is the probability that an allele below S_i in the dominance hierarchy exits. A new allele S_l enters at or below S_j in the dominance hierarchy with probability

$$E_j = \sum_{\alpha=1}^{j} e_a,$$

and the state of our allele of interest increases by one when the new allele enters at a state at or below its position after the exit. Therefore, the change in state after the exit and entrance become:

$$\begin{cases} i \mapsto i & \text{when } k < i \text{ and } l < i: \\ & \qquad \text{probability } D_i E_{i-1}, \\ i \mapsto i-1 & \text{when } k < i \text{ and } l \geq i: \\ & \qquad \text{probability } D_i (1 - E_{i-1}), \\ i \mapsto i+1 & \text{when } k > i \text{ and } l \leq i: \\ & \qquad \text{probability } (1 - D_i - d_i) E_i, \\ i \mapsto i & \text{when } k > i \text{ and } l > i: \\ & \qquad \text{probability } (1 - D_i - d_i)(1 - E_i). \end{cases}$$

Thus, the transition probabilities of the processes are given by

$$p_{ii} = D_i E_{i-1} + (1 - D_i - d_i)(1 - E_i),$$

$$p_{ii-1} = D_i(1 - E_{i-1}),$$

$$p_{ii+1} = (1 - D_i - d_i)E_i,$$

$$p_{i-i} = d_i,$$

for $i = 1, 2, \ldots, n$ and zero otherwise ($p_{ii} + p_{ii-1} + p_{ii+1} + p_{i-i} = 1$). The process is a Markov chain described by a random walk on the transient states 1, 2, . . . , n given by the matrix

$$\mathbf{A}^{\mathrm{T}} = \begin{Bmatrix} p_{11} & p_{12} & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\ p_{21} & p_{22} & p_{23} & \ldots & 0 & 0 & 0 & \ldots & 0 \\ 0 & p_{32} & p_{33} & \ldots & 0 & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & p_{i-1i-1} & p_{i-1i} & 0 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & p_{ii-1} & p_{ii} & p_{ii+1} & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 & p_{i+1i} & p_{i+1i+1} & \ldots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & p_{nn} \end{Bmatrix}$$

and by the absorbing states $-1, -2, \ldots, -n$, *i.e.*, $\mathbf{A}$ is the transposed matrix.

The probabilities of having made an exit from state i after one, two or three turnovers after the entrance are

$$\mathrm{Prob}(X_1 = -i) = d_i e_i,$$

$$\mathrm{Prob}(X_2 = -i) = d_i e_i + \sum_{j=1}^{n} d_i p_{ji} e_j,$$

$$\mathrm{Prob}(X_3 = -i) = d_i e_i + \sum_{j=1}^{n} d_i p_{ji} e_j + \sum_{j=1}^{n}\sum_{k=1}^{n} d_i p_{ji} p_{kj} e_k,$$

and in general

$$\begin{Bmatrix} \mathrm{Prob}(X_t = -1)/d_1 \\ \mathrm{Prob}(X_t = -2)/d_2 \\ \vdots \\ \mathrm{Prob}(X_t = -n)/d_n \end{Bmatrix} = \sum_{k=1}^{t} \mathbf{A}^{k-1}\mathbf{e} \rightarrow (\mathbf{I} - \mathbf{A})^{-1}\mathbf{e},$$

as $t \rightarrow \infty$, where

$$\mathbf{d} = \begin{Bmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{Bmatrix}, \quad \mathbf{e} = \begin{Bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{Bmatrix},$$

and the dot product of two vectors is $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i$. Therefore, $\mathrm{Prob}(X_\infty = -j \mid X_0 = i)$ equals the (j, i)th element of $(\mathbf{I} - \mathbf{A})^{-1}$ multiplied by d_j. The average time to exit of an allele is

$$\begin{aligned} ET &= \sum_{t=1}^{\infty} t\mathrm{Prob}(X_t < 0 \mid X_{t-1} > 0)\mathrm{Prob}(X_{t-1} > 0) \\ &= \sum_{t=1}^{\infty} t(1 - \mathrm{Prob}(X_t > 0 \mid X_{t-1} > 0)) \\ &\quad \mathrm{Prob}(X_{t-1} > 0) \\ &= \sum_{t=1}^{\infty} t\mathrm{Prob}(X_{t-1} > 0) - \sum_{t=1}^{\infty} t\mathrm{Prob}(X_t > 0) \\ &= \sum_{t=1}^{\infty} \mathrm{Prob}(X_{t-1} > 0). \end{aligned}$$

The probability of staying in the population after t allele-exit events is the probability of walking t steps on the states 1, 2, . . . , n, and this is

$$\mathrm{Prob}(X_t > 0) = \mathbf{1} \cdot \mathbf{A}^t \mathbf{e},$$

where $\mathbf{1}$ is the vector with all elements equal to 1. Inserting this in the above expression for ET and summing we get the average life time of an allele as

$$ET = \mathbf{1} \cdot (\mathbf{I} - \mathbf{A})^{-1}\mathbf{e}.$$

Therefore, the average life time of an allele that enters at the ith position on the dominance ladder, $E(T \mid X_0 = i)$, equals the sum of the ith column of the matrix $(\mathbf{I} - \mathbf{A})^{-1}$.