PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

How Does Self-Pollination Evolve? Inferences from Floral Ecology and Molecular Genetic Variation

Daniel J. Schoen, Martin T. Morgan and Thomas Bataillon

Phil. Trans. R. Soc. Lond. B 1996 **351**, 1281-1290
doi: 10.1098/rstb.1996.0111

How does self-pollination evolve? Inferences from floral ecology and molecular genetic variation

DANIEL J. SCHOEN, MARTIN T. MORGAN AND THOMAS BATAILLON

Department of Biology, McGill University, 1205 Avenue Docteur Penfield, Montreal, Quebec, Canada H3A 1B1

SUMMARY

The automatic selection and reproductive assurance hypotheses provide the two most general explanations for the evolution of self-pollination. Under automatic selection, self-pollination is mediated by pollen vectors and the mating system modifier experiences a transmission bias through the pollen that leads to its selection. Under reproductive assurance, self-pollination is autonomous and the mating system modifier is selected as it allows seed production when pollinators are scarce. We present phenotypic selection models that examine the selection of floral traits influencing several modes of selfing simultaneously. Inferences from these models suggest that reproductive assurance may be more important than has been appreciated. Additional insight into the importance of automatic selection versus reproductive assurance may be gained by considering the distribution of neutral genetic diversity among populations within selfing species. A number of approaches are outlined for analysing patterns of neutral diversity as they pertain to the mechanism of the evolution of selfing.

1. INTRODUCTION

The majority of angiosperms bear perfect flowers, i.e. flowers containing both anthers and stigmas. Most are outcrossing, but a significant proportion are predominantly self-pollinating. Indeed, the adoption of self-pollination is one of the most common trends in the evolutionary history of the angiosperms (Stebbins 1974).

Jain (1976) summarized hypotheses for the evolution of self-pollination. Two of most general are the so called 'automatic selection' and 'reproductive assurance' hypotheses. The automatic selection advantage of selfing arises because a gene promoting selfing in a population of outcrossers is, on average, transmitted to the next generation in two doses through the progeny arising from self-fertilization as well as an additional dose through the male gametes that cross-fertilize ovules in the population, whilst the alternative gene for outcrossing is transmitted in only two doses (Fisher 1941). This transmission bias gives mutations that increase the rate of selfing a strong advantage that is negated only by inbreeding depression or by other correlates of selfing such as reduced male fertility (Holsinger *et al*. 1984). The reproductive assurance hypothesis, on the other hand, states that the selective advantage of self-pollination lies in assured seed production when pollinators are insufficient for full pollination of the ovules, e.g. due to poor climatic conditions or following long-distance dispersal to areas where pollinators or mates are absent (Baker 1955). Darwin (1876) believed that reproductive assurance is the chief reason for the evolution of selfing.

Our principal focus is on how theoretical and experimental approaches can be applied to determine the relative importance of automatic selection and reproductive assurance. We do this through examination of: (1) phenotypic selection models for the evolution of floral traits promoting selfing; and (2) theory of selectively neutral genetic variation, concentrating specifically on how patterns of neutral diversity are influenced on historical features accompanying the evolution of selfing.

2. PHENOTYPIC SELECTION OF FLORAL TRAITS PROMOTING SELF-POLLINATION

(*a*) *The selection of selfing*

The advantages of selfing proposed by the automatic selection and reproductive assurance hypotheses occur under opposing ecological conditions. Automatic selection is dependent on vector-mediated pollen transfer (for realization of the transmission bias), whereas reproductive assurance is manifested when conditions for vector-mediated pollen transfer are inadequate for full seed set. Under automatic selection, increased selfing evolves whenever the relative fitness of progeny from selfing is one half or greater than that of progeny from outcrossing (Fisher 1941). In contrast, autonomous selfing can evolve (through reproductive assurance) whenever lack of pollinators limits seed set, provided that progeny from selfing have at least some fitness and that selfing does not preempt ovules that would otherwise be cross-pollinated (Lloyd 1979). The automatic selection advantage has dominated much of the theoretical and experimental work on the evolution of self-pollination, e.g. as illustrated by the many studies of inbreeding depression in selfers and outcrossers. Moreover, the majority of models assume that selfing does not contribute to an increase in seed set (Fisher 1941; Nagylaki 1976; Wells 1979; Holsinger 1991), thereby minimizing any role for reproductive

Phil. Trans. R. Soc. Lond. B (1996) **351**, 1281–1290
Printed in Great Britain

assurance. The models developed below are meant to examine the ecological conditions under which selfing is expected to evolve, with particular emphasis on how these conditions may influence opportunities for realization of the fitness gain.

(*b*) *Simultaneous changes in several modes of selfing*

Lloyd (1979, 1992) demonstrated that how and when self-pollination occurs during the lifetime of the flower (the 'mode' of selfing) strongly influences the conditions for its selection. For instance, when self-pollination depends upon the activities of pollinators ('facilitated self-pollination'), conditions for its selection differ from when it occurs independently of pollinators ('autonomous self-pollination'). This is true as well when selfing precedes or comes after opportunities for outcrossing (Lloyd 1979, 1992).

Our approach departs from that of Lloyd and others, who model the evolution of self-pollination *per se*, and thus consider only one mode of self-pollination at a time. Instead, we ask how floral changes that simultaneously influence the degree of *several* modes of selfing are selected. We assume that mutations causing changes in the timing of maturation or relative positions of anthers and stigmas within the flower are unlikely to influence only a single mode of selfing. Thus, in a self-compatible plant, a mutation that reduces the degree of dichogamy (temporal separation of anther and stigma maturation) or herkogamy (spatial separation of anther and stigma maturation) may lead to increased facilitated selfing. Such a mutation, however, may also lead to increased autonomous selfing. Likewise, a mutation that alters maturation times of the anthers and stigmas may lead to increased within-flower selfing, but such mutations may also increase the amount of between-flower (geitonogamous) selfing.

(*c*) *Selection of floral traits influencing autonomous and facilitated self-pollination*

Assume a population is comprised of two plant phenotypes that differ in the expression of floral trait z. The trait value z_i influences the amounts of both facilitated and autonomous self-pollination, $c(z_i)$ and $d(z_i)$, respectively, where the subscript i refers to the phenotype in question. The notation reminds us that the amount of facilitated and autonomous selfing are functions of the trait value z. Assume that autonomous selfing occurs after opportunities for outcrossing, i.e. autonomous selfing occurs via the 'delayed' autonomous mode of selfing described by Lloyd (1979), though our approach can be extended to other autonomous selfing modes. The proportion of ovules fertilized with the aid of an external vector (either through outcrossing or facilitated selfing) is symbolized by e, and the relative reduction in fitness of progeny from selfing versus outcrossing (inbreeding depression) is represented as δ. For simplicity, consider a population composed of a common phenotype 1, with floral trait expression z_1, and a rare phenotype 2 with floral trait expression, z_2(phenotype 2 has the higher selfing rate). Facilitated selfing and outcrossing are each dependent on pollinators, and hence, the number of seeds produced through these processes is influenced by e. Autonomous (delayed) selfing, on the other hand, involves only that proportion of ovules remaining unfertilized after pollinator visitation, i.e. $1-e$. With inbreeding depression, progeny from all modes of selfing have fitness $1-\delta$ relative to progeny from outcrossing. Taking these factors together, we obtain the following expressions for female reproductive fitnesses:

$$W_{f1} = c(z_1)e(1-\delta)+d(z_1)(1-e)(1-\delta)+[1-c(z_1)]e$$
$$W_{f2} = c(z_2)e(1-\delta)+d(z_2)(1-e)(1-\delta)+[1-c(z_2)]e. \qquad (1)$$

Male reproductive fitness is gained both through outcrossing and selfing. For the outcrossing component of male fitness, assume that phenotype 2 is rare, and hence, pollen from either phenotype encounters only ovules from phenotype 1. Male fitnesses gained through non-autonomous and autonomous self-pollination, and outcrossing may be written as:

$$W_{m1} = c(z_1)e(1-\delta)+d(z_1)(1-e)(1-\delta)+[1-c(z_1)]e$$
$$W_{m2} = c(z_2)e(1-\delta)+d(z_2)(1-e)(1-\delta)+[1-c(z_1)]e. \qquad (2)$$

The relative fitness of phenotype 2 is:

$$\tilde{W}_2 = 1/2\ [W_{f2}/\bar{W}_f + W_{m2}/\bar{W}_m], \qquad (3)$$

where $\bar{W}_f$ and $\bar{W}_m$ denote average fitness through female and male reproductive functions. To find the evolutionary stable floral trait value, z^*, solve for the conditions under which $d\tilde{W}_2/dz_2 = 0$ (when $z_1 = z_2 = z^*$), yielding the relationship:

$$2[c(z_2)'e(1-\delta)+d(z_2)'(1-e)(1-\delta)]-c(z_2)'e = 0, \qquad (4)$$

where the apostrophes denote the derivatives of the functions $c(z_i)$ and $d(z_i)$.

Assume a linear relationship between z_i and the selfing rate under each mode. Note that the change in selfing caused by modifications to the floral trait may differ for each mode, so accordingly, $c(z_i) = m_c z_i + y_c$ and $d(z_i) = m_d z_i + y_d$, where m_c and m_d denote the respective slopes of the linear relationship between facilitated and autonomous selfing modes and the trait value, and y_c and y_d denote the respective y-intercepts. The mutant phenotype will invade the population whenever:

$$\delta < 1-m_c e/\{2[m_c e+m_d(1-e)]\}. \qquad (5)$$

Figure 1 illustrates how pollinator activity and inbreeding depression levels specified in equation (5) combine to select for the floral trait. Conditions for the evolution of the trait are intermediate between those for strict facilitated selfing ($\delta < 1/2$) and those for autonomous (delayed) selfing ($\delta < 1$). The more important result, however, is that in pollinator-limited environments, when a change in floral trait expression leads to increases in *both* facilitated and autonomous selfing, even if the increase in autonomous selfing is small (i.e. $m_d/m_c < 1$), the selection of selfing is subject to more relaxed conditions compared with facilitated selfing alone (i.e. inbreeding depression is less of an obstacle to the evolution of selfing). Reproductive

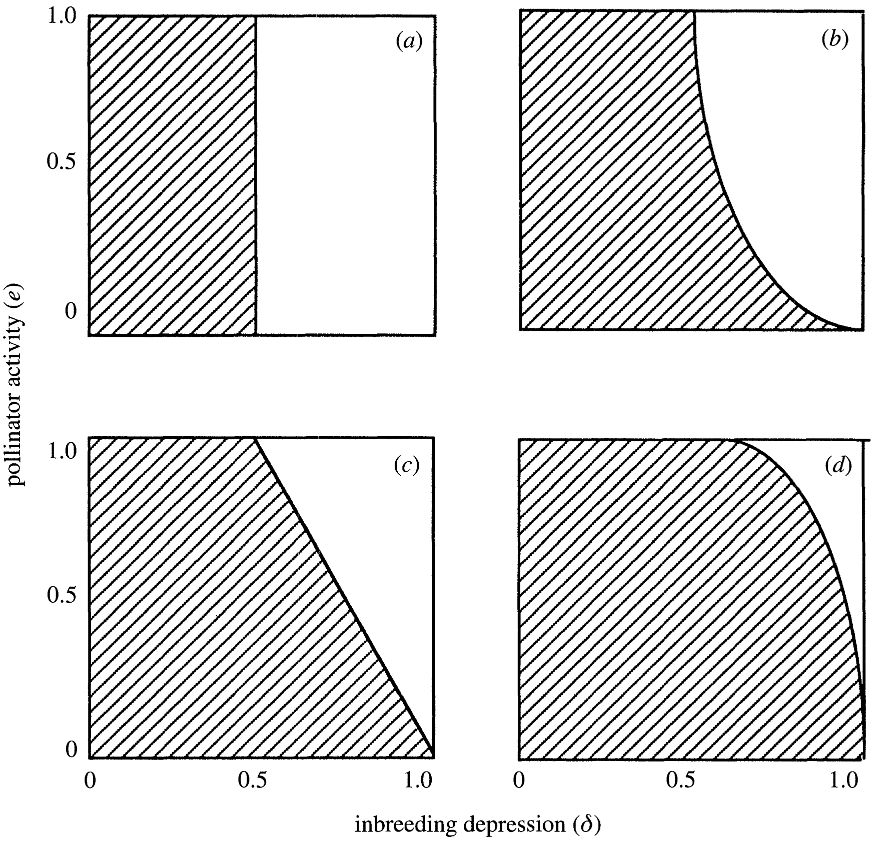

Figure 1. Conditions for the selection of a floral trait that causes a linear increase in both non-autonomous (slope m_c) and autonomous (slope m_a) self-pollination. Four cases: (*a*) $m_a/m_c = 0$; (*b*) $m_a/m_c = 0.1$; (*c*) $m_a/m_c = 1.0$; (*d*) $m_a/m_c = 3.0$. Hatched areas indicate parameter combinations where the trait is selected.

assurance may, therefore, play an important role, not only in the evolution of autonomous selfing, but facilitated selfing as well.

(*d*) *Selection of floral traits influencing within- and between-flower self-pollination*

Self-pollination may occur within as well as between flowers on the same plant, especially when it is facilitated by pollinators. Changes in flowers may influence the rates of both of these modes of selfing. For example, in a normally protandrous species, a mutation that causes the stigmatic lobes to expand and become receptive earlier on in floral development, thereby lengthening the period of time when self pollen can be deposited by insects, may lead to higher rates of selfing within the flower. But this same change also lengthens the period of stigma receptivity on other flowers of the plant, thereby increasing the opportunity for between-flower selfing. Consider a trait z, with value z_i in phenotype i, that influences within- and between-flower selfing, symbolized $c(z_i)$ and $g(z_i)$, respectively. Assume that both within- and between-flower selfing are facilitated by external pollen vectors. The proportion of seeds arising from self-fertilization is thus given by $[c(z_1)+g(z_1)]e$. In a population with phenotypes 1 and 2, the female reproductive fitnesses are:

$$W_{f1} = \{[c(z_1)+g(z_1)](1-\delta)+[1-c(z_1)-g(z_1)]\}e$$
$$W_{f2} = \{[c(z_2)+g(z_2)](1-\delta)+[1-c(z_2)-g(z_2)]\}e. \quad (6)$$

Next, consider male reproductive fitnesses. When the floral trait influences the amount of between-flower selfing, it is incorrect to assume that it will have no influence on the amount of pollen available for outcrossing. This is because the same mechanism of between-flower transfer of (self) pollen by vectors is that used in the process of outcrossing (Lloyd 1992). In the terminology of Holsinger *et al.* (1984), pollen discounting may be associated with increased between-flower selfing. Let $p(z_2)$ denote the proportion of a plant's pollen available for fertilizing outcrossed ovules in the population. Assume that phenotype 2 is rare, so that outcrossed pollen from either phenotype encounters only ovules of phenotype 1. This leads to the following male fitness expressions:

$$W_{m1} = \{[c(z_1)+g(z_1)](1-\delta)+[1-c(z_1)-g(z_1)]\}e$$
$$W_{m2} = \{[c(z_2)+g(z_2)](1-\delta)+[1-c(z_1)-g(z_1)][p(z_2)/p(z_1)]\}e. \quad (7)$$

If one assumes that the trade-off between the rate of between-flower selfing and the amount of pollen available for outcrossing is complete (Lloyd 1992), then $p(z_i) = 1-g(z_i)$. Assuming, as before that the rate of self-pollination is a linear function of the trait value z_i, we have $c(z_i) = m_c z_i + y_c$ and $g(z_i) = m_g z_i + y_g$. The evolutionary stable floral trait value satisfies $d\tilde{W}_2/dz_2 = 0$ (when $z_1 = z_2 = z^*$), and it can be shown that the mutant phenotype will invade the population when:

$$\delta < 0.5 - km_g/[2(m_c+m_g)], \quad (8)$$

where k ($k \leqslant 1$) is a constant that reflects the proportion of total selfing in the wild type that is within-flower as opposed to between-flower. Values of

k near 1 indicate that within-flower selfing predominates. Figure 2 illustrates the conditions for selection of a floral trait that influences both within- and between-flower selfing rates. When $m_g/m_c = 0$ (i.e. no effect of the trait on between-flower selfing), conditions for selection of increased selfing are identical to those for facilitated selfing ($\delta < 1/2$). But as the influence of the trait on between-flower selfing increases ($m_g/m_c > 0$), conditions for the evolution of the trait become more restrictive. The threshold values of δ shown in figure 2 would be larger (less of an obstacle to the evolution of selfing) if pollen discounting accompanying between-flower selfing less severe, but the qualitative result shown there would still hold.

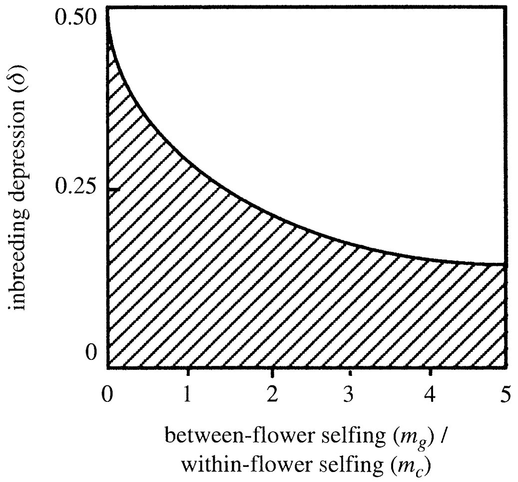

Figure 2. Conditions for the selection of a floral trait that causes a linear increase in both between- and within-flower self-pollination rates (slopes m_g and m_c, respectively) ($k = 0.75$). Hatched areas indicate parameter combinations where the trait is selected.

The results above indicate that increases in between-flower selfing occurring in conjunction with increases in facilitated selfing oppose the automatic selection advantage. Consequently, automatic selection as a driving force for the evolution of selfing may be less important than simpler (one mode) models suggest.

3. NEUTRAL GENE DIVERSITY AND THE EVOLUTION OF SELF-POLLINATION

When self-pollination evolves by automatic selection, the way in which selfing spreads to other populations may differ from when it evolves by reproductive assurance (figure 3). Specifically, as autonomous selfing allows for uniparental reproduction, one might expect that there will be more opportunities for the founding of populations by one or a few individuals (Baker

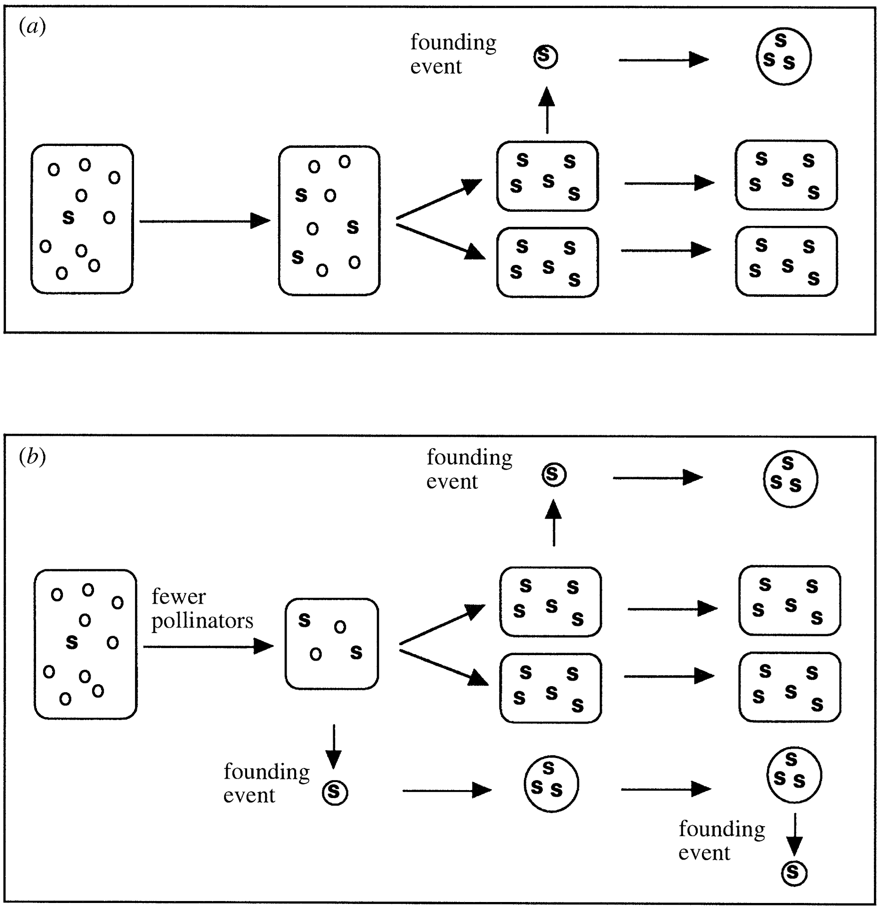

Figure 3. Hypothetical population histories associated with the selection of self-pollination. Under automatic selection (*a*), the mating system modifier spreads due to the transmission bias, but the probability of founding events is not enhanced. Under reproductive assurance (*b*), the mating system modifier spreads because it increases seed set in an environment where vector-mediated pollination is uncertain. Founding events are more likely due to uniparental reproduction. The symbols '**s**' and '**o**' indicate selfing and outcrossing morphs.

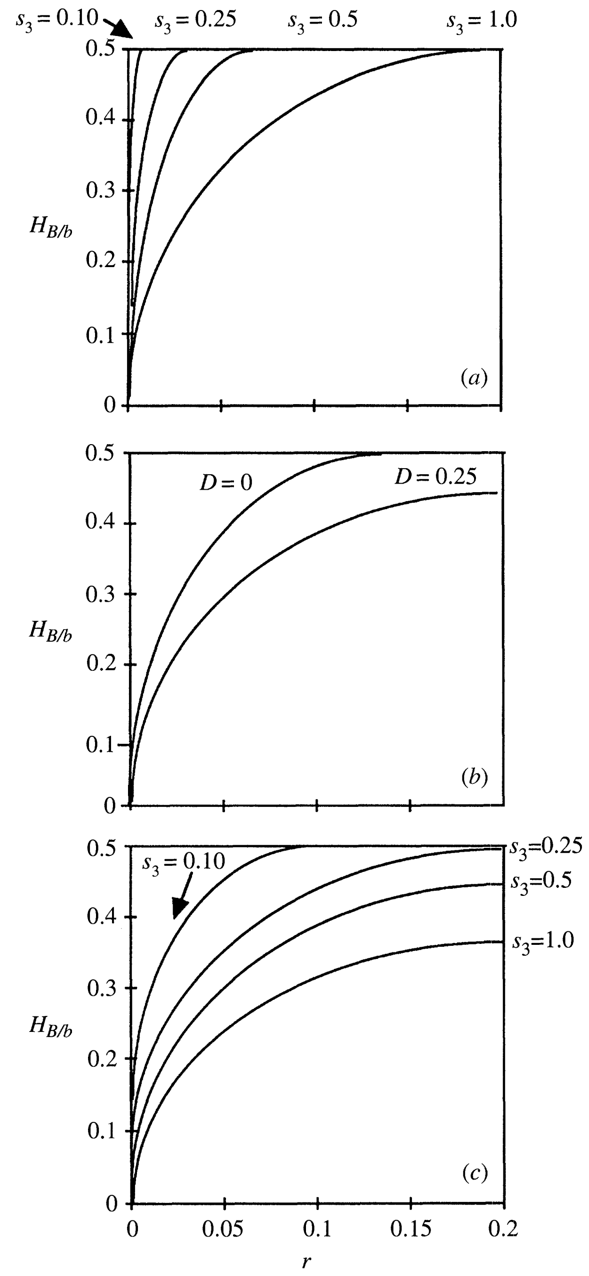

Figure 4. The effect of selection of a mating system modifier (*a* allele) on Nei's gene diversity at neutral, diallelic locus *B* ($H_{B/b}$) as a function of selfing rate (in the *aa* homozygote, s_3), and recombination fraction *r*. The mating system modifier is introduced into: (*a*) an outcrossing population in low frequency ($q_1 = 0.001$) and in near linkage equilibrium with locus *B*; (*b*) a population of partial selfers ($s_1 = 0.75$) in low frequency ($q_1 = 0.001$) and near zero linkage equilibrium with locus *B* ($D = 0$ curve), or into a founder population at high frequency ($q_1 = 0.5$) and in complete linkage disequilibrium with locus *B* ($D = 0.25$ curve) (the selfing rate of the *aa* homozygotes is $s_3 = 1.0$, both curves); and (*c*) a founder population at high frequency ($q_1 = 0.5$) and in complete linkage disequilibrium with locus *B*.

1955). The expectation of increased founding events does not accompany the evolution of selfing by automatic selection, as seed set remains dependent on the presence of pollinators and mates. These differences in population history may have a pronounced effect on the distribution of neutral diversity among populations within selfing species, as discussed below.

(*a*) *Hitchhiking and the selection of mating system modifiers under automatic selection and reproductive assurance*

The selection of a favourable mutation and its influence on variation at a linked neutral locus has been investigated by several researchers (Maynard Smith & Haigh 1974; Hedrick 1980). The issue of how the selection of a mating system modifier directly influences diversity at a neutral locus has not, however, been investigated. The question we pose here is whether the effect of hitchhiking differs under automatic selection versus reproductive assurance.

Consider a population with two diallelic loci *A* and *B*. Locus *A* determines the selfing rate, whilst locus *B* is selectively neutral. Let the frequencies of the ten two-locus genotypes be denoted as v_i, the frequencies of the gametes *AB*, *Ab*, *aB* and *ab* as x_{11}, x_{10}, x_{01}, x_{00}, and the frequencies of alleles at loci *A* and *B* as p_1, q_1 and p_2, q_2. The selfing rates of genotypes *AA*, *Aa* and *aa* are s_1, s_2, and s_3, respectively, with $s_1 \leqslant s_2 < s_3$. For brevity, consider only the case where s_2 is exactly intermediate between s_1 and s_3 (codominance). Recursion equations for the v_i's can be obtained by modifying those in Strobeck (1979), i.e. specifying the selfing rate for each genotype at the *A* locus.

Under automatic selection, the mating system modifier (*a* allele) will be introduced into the population at low frequency by mutation, and hence, linkage disequilibrium between the modifier and neutral locus will be near zero. Consider a population that is initially outcrossing. Figure 4*a* summarizes the effect of hitchhiking on diversity at neutral locus *B* as a function of the recombination fraction between the two loci when initial frequencies at the neutral locus *B* are $p_2 = q_2 = 0.5$ and the mating system modifier allele is introduced at $q_1 = 0.001$. Only when the effect of *a* on the selfing rate is large and the recombination fraction between the two loci is moderate to small (0.1 or less) is there any pronounced decrease in diversity at the neutral locus. If the influence of the *a* allele is reduced (e.g. from causing 100% selfing in the homozygote to only 10% selfing), the hitchhiking effect diminishes quite rapidly and is confined to a much smaller genomic region (figure 4*a*). Hitchhiking has a more pronounced effect on neutral diversity when the mating system modifier is introduced into a population that is already partially selfing (figure 4*b*), presumably because the selection of the modifier occurs more rapidly in the homozygous state.

To determine the effect of hitchhiking, assumptions must be made about the magnitude of selfing rate modification arising from floral change. Since most modifications to floral traits cause relatively small changes in the rate of selfing, it would appear from the analytical results that automatic selection will not lead to any broad (i.e. throughout the genome) reduction in initially outcrossing populations. A stronger (more genome-wide) effect on neutral diversity levels can, however, occur if such modifiers arise in partially selfing populations.

What about the effect of hitchhiking under reproductive assurance? As noted above, there is an

Table 1. *The effects of hitchhiking, background selection, and genetic drift on neutral gene diversity under two hypothetical mechanisms for the evolution of self-pollination*

mechanism of evolution of self-pollination	associated population history	hitchhiking	background selection	genetic drift
automatic selection	no association of the evolution of selfing with increased founding events	decline in diversity at neutral loci	decline in neutral diversity in all populations with high selfing rates.	decline in diversity due to inbreeding effects on N_e
reproductive assurance	the evolution of selfing coupled with population disturbance or founding events (bottlenecks)	as above, but additional bottlenecks increase the probability of linkage disequilibrium, between selected and neutral loci, thereby contributing further to loss of neutral diversity	as above	as above, but additional bottlenecks contribute to loss of neutral diversity through genetic drift

increased likelihood of founding events in the history of a selfing species under this hypothesis, and thus an increased probability of linkage disequilibrium between the mating system modifier locus and any neutral loci remaining in a polymorphic state. This linkage disequilibrium is expected to enhance the decline in neutral diversity following a founding event (Hedrick 1980). For example, compare the situation in which the mating system modifier is introduced into an outcrossing population, in linkage equilibrium with the locus B (figure 4*a*), to the case where a double heterozygote individual founds a new population – where linkage disequilibrium between the two loci is necessarily maximum (figure 4*c*). The reduction in neutral diversity due to hitchhiking is more pronounced and extends to genomic regions more distant from the modifier (figure 4*c*). The same effect can be seen when the ancestral (source) population is partially selfing (figure 4*b*). The enhanced hitchhiking effect under reproductive assurance will contribute to increases in among-population variation in levels of neutral diversity (see below).

(*b*) *Indirect effects of selection of mating system modifiers on diversity at neutral loci*

Will automatic selection and reproductive assurance have different indirect influences on diversity levels at neutral loci? First, note that inbreeding leads to a reduction in effective population size (N_e). The expected reduction is given by $N_e(selfer) = N_e(outcrosser)/(1+F)$, where F is the inbreeding coefficient of the population arising from selfing (Pollak 1987). According to the sampling properties of the neutral model (Ewens 1972), a reduction in effective population size due to inbreeding will lead to a reduction in neutral allele diversity. Second, selfing leads to slower decay of linkage disequilibrium, and as noted in the case of hitchhiking between mating system loci and neutral alleles, this should contribute to reductions in neutral allelic diversity whenever any selected and neutral loci start out in linkage disequilibrium (Hedrick 1980). Third, with high levels of selfing, a reduction of neutral diversity may occur because of ‘background selection’, in which recurrent deleterious mutations located throughout the genome are selectively eliminated (Charlesworth *et al.* 1993). Fourth, when autonomous selfing increases the likelihood of bottlenecks, there is the expectation of reduced neutral diversity due to genetic drift (Nei *et al.* 1975; Tajima 1989).

Table 1 summarizes some expectations pertaining to patterns of neutral diversity under the two hypotheses for the evolution of selfing. Note that neutral diversity levels within single selfing populations are expected to be reduced by inbreeding effects on N_e and background selection, regardless of how selfing has evolved and, therefore, examination of average population levels of neutral diversity alone may not help to distinguish between automatic selection and reproductive assurance.

On the other hand, examination of variation in levels of neutral diversity *among* a number of populations of a selfing species has the potential to suggest the mechanism of evolution of selfing. In particular, species in which selfing has evolved by reproductive assurance are expected to consist of populations that have recently recovered from a bottleneck, together with older and more stable populations (figure 3). Those populations that have recently passed through bottlenecks will have reduced diversity levels due to both drift and the enhancement of hitchhiking, whilst those populations that have been demographically stable for some time will store higher levels of neutral diversity. Thus, in species where selfing has evolved through reproductive assurance, one might expect significant among-population variation in neutral diversity. This expectation is not a necessary corollary of automatic selection.

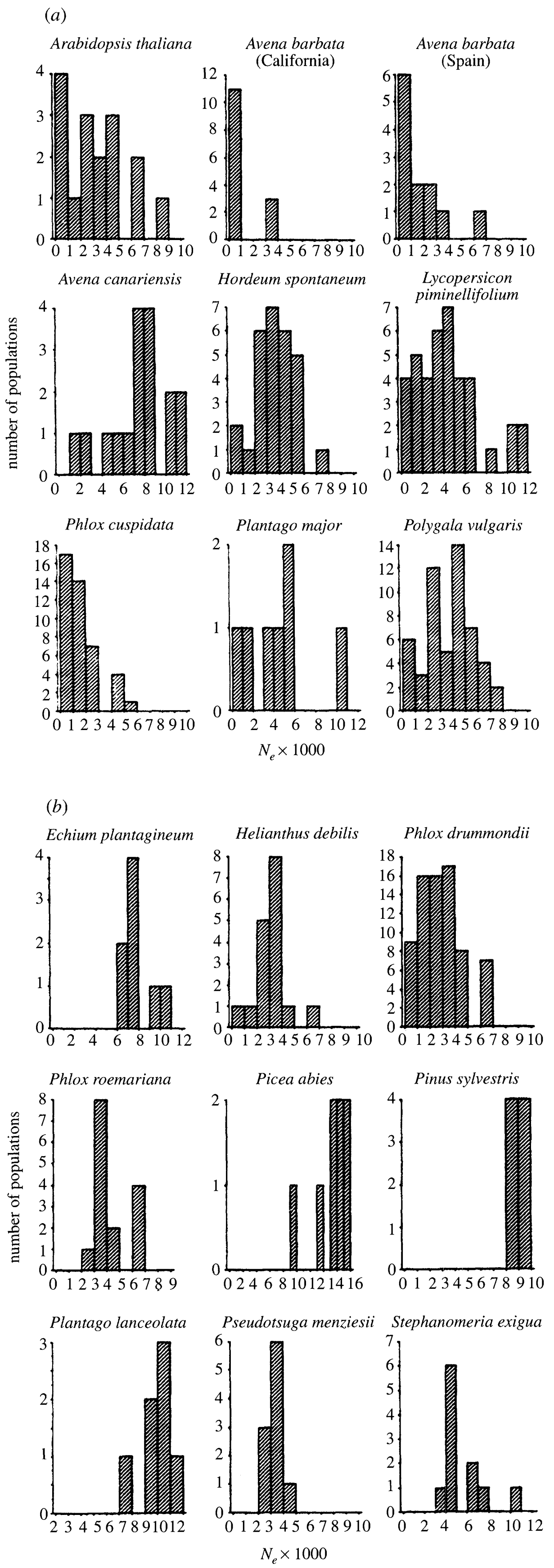

Figure 5. Variation among populations in allozyme diversity levels in (*a*) selfing and (*b*) outcrossing species, after Schoen & Brown (1991). Diversity is expressed as an estimate of effective population size (N_e) and is calculated from allelic richness at the sampled allozyme loci (Ewens 1979).

(*c*) *Variation in allozyme diversity in the populations of selfing species*

There have been many studies of allozyme variation in selfing plant species (Hamrick & Godt 1990; Schoen & Brown 1991) (figure 5). Substantially more variation in allozyme diversity is found among populations of selfers compared with outcrossers. For example, most selfers surveyed consist of populations with no detectable allozyme diversity together with other populations having relatively high levels of diversity. Such a tendency towards L-shaped distributions of allozyme diversity within selfing species suggests a possible history of recent bottlenecks in some but not all populations, as envisioned under the reproductive assurance model. The inference is an indirect one, however, and is based on assumptions about neutrality of allozyme loci (Gillespie 1991). Moreover, L-shaped distributions of neutral variation could be produced if selfing populations within the species studied have evolved repeatedly from outcrossing ancestors, but at different times in the past, coupled with background selection (Charlesworth *et al.* 1993), i.e. selective sweeps that have removed neutral genetic variation in the more ancient selfing populations, but not in those in which selfing has recently evolved. Clearly, there is a need for methods that more directly reveal the effect of population history on neutral variation. One possible approach is examined next.

(*d*) *DNA sequence variation and the history of selfing populations*

Here we focus on how coalescent theory might be applied to investigate historical events accompanying the evolutionary transition from outcrossing to self-fertilization. Of particular relevance is whether population bottlenecks occur in conjunction with this mating system transition, and how these bottlenecks influence the evolutionary history of a sample of genes. Slatkin & Hudson (1991) have used coalescent theory in an analogous manner to examine historical changes in population size.

Coalescent analyses characterize the statistical properties of a sample of genes by tracing the ancestors of these sampled genes backward over successive generations (Hudson 1990). A coalescent event occurs when two genes derive from a single ancestral gene in the previous generation. After many generations, the ancestry of all genes traces back to a single ancestral gene, the most recent common ancestor of the sample. The result of tracing the ancestry of sampled genes is a genealogy (figure 6). Ecological factors such as population size and rate of self-fertilization determine the topology and length of branches in the genealogy. Mutations that accumulate along the branches of the genealogy provide information for statistical characterization of the sampled genes. Statistics include the number of segregating sites in the sample (Watterson 1975) and the average pairwise divergence between sampled genes (Tajima 1983). As all sampled genes trace back to the most recent common ancestor, the time of the most recent common ancestor of the

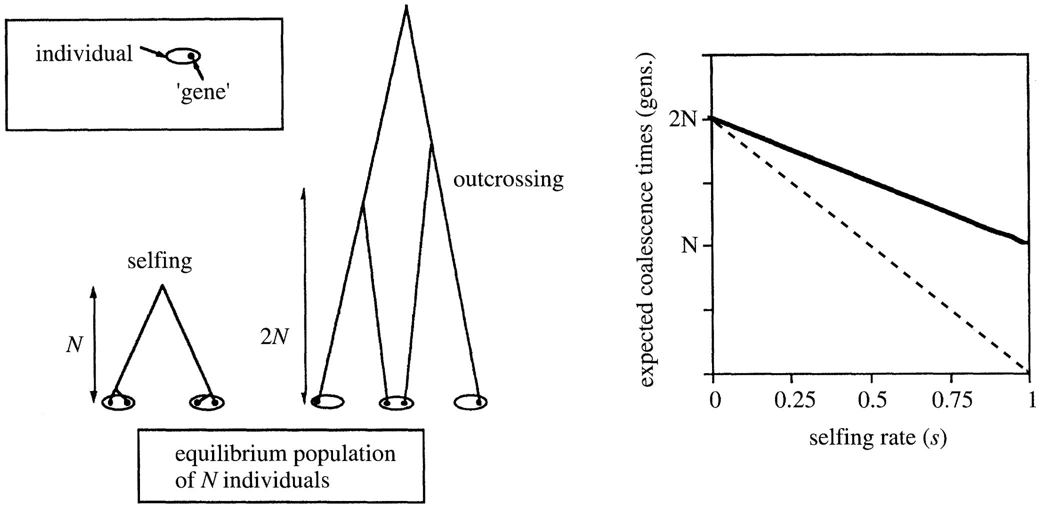

Figure 6. Hypothetical genealogies involving four sampled genes. In a completely selfing population, genes sampled within the same individual tend to coalesce early, whereas in an outcrossing population, genes sampled initially within the same individual coalesce (on average) at roughly the same time as genes sampled in separate individuals. The accompanying graph shows average coalescence times for pairs of genes sampled within (broken line) and between (solid line) individuals as a function of the selfing rate in a population containing N individuals.

sampled genes sets an upper limit for the period about which evolutionary inferences can be drawn.

For instance, consider a sample of two genes in a diploid population of size N individuals with selfing rate s. For such a sample, there are two possible configurations for the genes – either both genes may be in the same individual, or each gene may be in separate individuals. Ecological parameters (e.g. population size, selfing rate) determine the probabilities of transition between different configurations (Milligan 1996). Consider two genes that in the present generation are in the same individual. Suppose that such an individual is derived from self-fertilization, which occurs with probability s. In this case, the ancestors of the sampled genes in the previous generation (the 'parental' genes) must be in the same individual. The laws of Mendelian segregation tell us that half of the time both sampled genes are derived from the same parental gene (i.e. the sampled genes coalesce), and half of the time the sampled genes are derived from different genes which are nonetheless in the same individual. On the other hand, suppose that the sampled genes are in an individual that arose from outcrossing, occurring with probability $1-s$. The parental genes are in separate individuals with probability approximately equal to $1-1/N$, and in the same individual (because the same parental individual was randomly sampled twice) with probability $1/N$. If the genes are in the same individual, Mendelian rules of segregation tell us that there is probability $1/2$ that a coalescent event occurs, and probability $1/2$ that the genes are distinct but in the same individual. Combining the probabilities of self-fertilization and outcrossing, two genes in the same individual coalesce with probability $s(1/2)+(1-s)(1/N)(1/2)$, are in the same individual *but do not coalesce* with probability $s(1/2)+(1-s)(1/N)/(1/2)$, or are in different individuals with probability $(1-s)(1-1/N)$. Changes in population size or selfing rate alter these transition probabilities. Iteration of the transition probabilities allowsstraightforward determination of the genealogy of sampled genes, and hence the statistical properties of the sample.

(*e*) *The evolution of selfing and expected coalescent patterns*

We describe results from an investigation of the historical change in rates of self-fertilization, reduction in population size, and the simultaneous effects of change in *both* selfing rate and population size. The models with population bottlenecks involve a reduction of populations to one tenth of their original size, persistence of the population at small size for a number of generations, and then expansion of the population to its original size. The time scale of all scenarios is in units of $\tau = N$ generations, so that the results presented do not depend on the exact population size of the simulations (only the relative size). The symbol T represents the period during which the bottlenecked populations remain small. For each scenario of ecological change, we calculated the expected time, t, to coalescence of two genes sampled in different individuals.

Several conclusions can be seen from the portion of this analysis that deals with historical changes in selfing First, the expected coalescent time of genes sampled within and between individuals decreases linearly with rate of self-fertilization (figure 6). The decrease of expected coalescent times of genes sampled between individuals from $2N$ in a random mating population to N in a completely self-fertilizing population is comparable to the well known linear increase in coefficient of consanguinity, just as the decrease in within-individual coalescent times parallels the increasing inbreeding coefficients found using traditional population genetic techniques (Crow & Kimura 1970). Second, the coalescent times of genes sampled within individuals in self-fertilizing populations provide little insight into the evolutionary history of populations because they coalesce quickly. On the other hand, genes sampled *between* individuals can potentially provide information on historical population parameters occurring over time spans of between $2N$ generations (in historically selfing populations) and $4N$ generations (in populations with historical outcrossing). This can be seen in figure 7*a*, where the x-axis is the number of generations before the present, τ, when the population was last outcrossing and the y-axis is the expected between-individual coalescent time. The result illustrates the important point that demographic parameters such as the number of individuals in the population impose an upper limit on the historical resolution attainable through sampling of genetic information in contemporary populations. Third, as the time of the transition from outcrossing to self-fertilization recedes into the past, the expected coalescent time changes monotonically from that characterizing outcrossing populations ($2N$) to that characterizing completely self-fertilizing populations (N).

Next, consider the portion of the analysis that deals with historical reduction in population size (figure 7*b*). It can be seen that historical bottlenecks occurring without any alteration in the rate of self-fertilization reduce expected between-individual coalescent times. Focusing on the solid line in figure 7*b* (which corresponds to the coalescent in populations with intermediate self-fertilization and a bottleneck persisting for $T = 0.05N$ generations), it can be seen that when the bottleneck first occurred at a time τ less than the duration of the bottleneck, $\tau < T$, the contemporary population is at the reduced size and the coalescent times are those expected in a population suddenly reduced in size $T - \tau$ generations ago. Population size recovers from the bottleneck as τ becomes larger than T, and as the bottleneck recedes into history, the coalescent times gradually return to the value expected for the larger population. The main effect of population bottlenecks then, is to reduce the between-individual coalescent times. The size of the reduction can be considerable and is proportional to the magnitude of the bottleneck, contrasting with the more modest effects of increase in self-fertilization that alter coalescent times by at most a factor of 2. Like the effects of changes in self-fertilization, though, demography imposes an upper limit on historical resolution of about $3N$ generations.

Finally, consider the evolution of self-fertilization occurring in conjunction with population bottlenecks. This leads to coalescent patterns that are approximately the additive combination of the separate effects of each individual factor (i.e. changes in self-fertilization, bottlenecks) (figure 7*c*). When bottlenecks are not too severe (either of short duration or only small reduction in size), the consequences of past outcrossing express themselves in expected coalescent times greater than expected with historically constant rates of selfing.

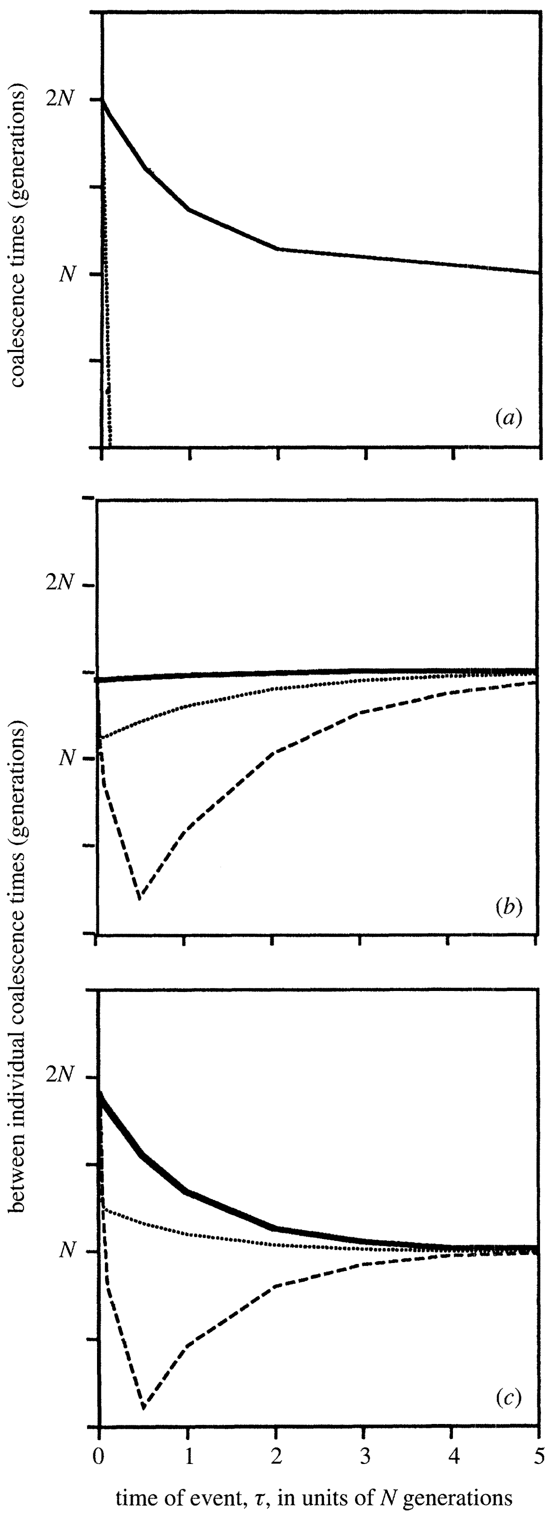

Figure 7. Coalescence times for pair of genes sampled in a population: (*a*) that was historically outcrossing but where a transition to complete selfing occurred τN generations ago (N denotes population size and is assumed to be constant in this case; dotted line denotes within individuals; solid line denotes between individuals); (*b*) with constant selfing rate ($s = 0.5$), but where a bottleneck occurred τN generations ago reducing population size by a factor of 10 for a duration of TN generations (after which the population regained its original size N) (solid line denotes $T = 0.005N$; dotted line denotes $T = 0.05N$; broken line denotes $T = 0.5N$); (*c*) where a transition from historical outcrossing to complete selfing occurred τN generation ago in conjunction with a bottleneck reducing the population size by a factor of 10 for a duration of TN generations (after which the population regained its original size N) (solid line denotes $T = 0.005N$; dotted line denotes $T = 0.05N$; broken line denotes $T = 0.5N$).

Increasing the size or duration of the bottleneck overwhelms the effects of the evolution of self fertilization, leading to coalescent times much smaller than expected based on current population size and mating system.

These results hold some promise for inferring the mechanism of evolution of self fertilization based on observation of contemporary population structure. If under the automatic selection hypothesis the transition from outcrossing to self-fertilization generally occurs in the absence of population bottleneck, then estimates of coalescent times and related statistics from contemporary populations may be characteristic of larger or more outcrossing populations than census and mating system estimates suggest. On the other hand, under reproductive assurance, coalescent times and derived statistics should reflect much smaller population sizes than census size would indicate.

There are important caveats. Population demographic parameters impose an upper time limit, of several N generations, during which mating system transitions are detectable. This restricts application to species where the mating system transition is thought to have occurred relatively recently. The effects of the selfing transition and of population bottleneck may cancel each other out so that, in the absence of additional information, the conclusion reached is simply that the population has experienced approximately constant selfing rates and population sizes for at least several N generations. Finally, all inferences drawn from contemporary samples rely on comparing observed with expected statistics (e.g. comparing coalescent times observed with those expected if the selfing rate and population size were constant). The inferential process presented above, therefore, relies on external information (e.g. additional information from sources other than the genetic data about contemporary population size and mating system). One way of circumventing this difficulty is through development of statistics that characterize the topology of the tree. For instance, Tajima (1989) shows that his D statistic, which contrasts the number of segregating sites with the average pairwise divergence, reflects recent changes in population size. Development of such statistics represent areas of active research.

We thank Spencer Barrett, Kent Holsinger, Russell Lande and Stewart Schultz for discussing the material presented here.

REFERENCES

Baker, H. G. 1955 Self-compatibility and establishment after 'long-distance' dispersal. *Evolution* **9**, 347–348.

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

Crow, J. F. & Kimura, M. 1970 *An introduction to population genetics theory*. New York: Harper and Row.

Darwin, C. 1876 *The effects of cross- and self-fertilization in the vegetable kingdom*. London: John Murray.

Ewens, W. J. 1972 The sampling theory for selectively neutral alleles. *Theor. Popul. Biol.* **3**, 87–112.

Fisher, R. A. 1941 Average excess and average effect of a gene substitution. *Ann. Eugen.* **11**, 53–63.

Gillespie, J. H. 1991 *The causes of molecular evolution*. Oxford University Press.

Hamrick, J. L. & Godt, M. J. 1990 Allozyme diversity in plant species. In *Plant population genetics, breeding, and genetic resources* (eds. A. H. D. Brown, M. T. Clegg, A. L. Kahler & B. S. Weir), pp. 43–63. Sunderland: Sinauer.

Hedrick, P. W. 1980 Hitch-hiking: a comparison of linkage and partial selfing. *Genetics* **94**, 791–808.

Holsinger, K. E. 1991 Mass-action models of plant mating systems: the evolutionary stability of mixed mating systems. *Am. Nat.* **138**, 606–622.

Holsinger, K. E., Feldman, M. W. & Christiansen, F. B. 1984 The evolution of self-fertilization in plants. *Am. Nat.* 124, 446–453.

Hudson, R. R. 1990 Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology* **9**, 1–44.

Jain, S. K. 1976 The evolution of inbreeding in plants. *Ann. Rev. Ecol. Syst.* **7**, 469–495.

Lloyd, D. G. 1979 Some reproductive factors affecting the selection of self-fertilization in plants. *Am. Nat.* **113**, 67–69.

Lloyd, D. G. 1992 Self- and cross-fertilization in plants. II. The selection of self-fertilization. *Int. J. Plant. Sci.* **153**, 370–380.

Maynard Smith, J. & Haigh, J. 1974 The hitch-hiking effect of a favorable gene. *Genet. Res.* 231, 1114–1116.

Milligan, B. G. 1996 Estimating long-term mating systems using DNA sequences. *Genetics* **142**, 619–627.

Nagylaki, T. 1976 A model for the evolution of self-fertilization and vegetative reproduction. *J. theor. Biol.* **58**, 55–58.

Nei, M., Maruyama, T. & Chakraborty, R. 1975 The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1–10.

Pollak, E. 1987 On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* **117**, 353–360.

Schoen, D. J. & Brown, A. H. D. 1991 Interspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4494–4497.

Slatkin, M. & Hudson, R. R. 1991 Pairwise comparison of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555–562.

Stebbins, G. L. 1974 *Flower plants: evolution above the species level*. Cambridge, MA: Harvard University Press.

Strobeck, C. 1979 Partial selfing and linkage: the effect of a heterotic locus on a neutral locus. *Genetics* **92**, 305–315.

Tajima, F. 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460.

Tajima, F. 1989 The effect of change in population size on DNA polymorphism. *Genetics* **123**, 597–601.

Watterson, G. 1975 On the number of segregating sites in genetic models without recombination. *Theor. Popul. Biol.* **7**, 256–276.

Wells, H. 1976 Self-fertilization: advantageous or disadvantageous? *Evolution* **33**, 252–255.