Allelic Genealogies in Sporophytic Self-Incompatibility Systems 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 [†]Laboratoire de Génétique et d'Ecologie Végétales, Université Libre de Bruxelles, B-1160 Brussels, Belgium

ABSTRACT

Expectations for the time scale and structure of allelic genealogies in finite populations are formed under three models of sporophytic self-incompatibility. The models differ in the dominance interactions among the alleles that determine the self-incompatibility phenotype: In the SSIcod model, alleles act codominantly in both pollen and style, in the SSIdom model, alleles form a dominance hierarchy, and in SSIdomcod, alleles are codominant in the style and show a dominance hierarchy in the pollen. Coalescence times of alleles rarely differ more than threefold from those under gametophytic self-incompatibility, and transspecific polymorphism is therefore expected to be equally common. The previously reported directional turnover process of alleles in the SSIdomcod model results in coalescence times lower and substitution rates higher than those in the other models. The SSIdom model assumes strong asymmetries in allelic action, and the most recessive extant allele is likely to be the most recent common ancestor. Despite these asymmetries, the expected shape of the allele genealogies does not deviate markedly from the shape of a neutral gene genealogy. The application of the results to sequence surveys of alleles, including interspecific comparisons, is discussed.

HOMOMORPHIC, single-locus self-incompatibility (SI) systems in plants have long been known to contain a very large number of alleles (Emerson 1939). With the recent molecular characterization of SI systems in both the gametophytic SI system (GSI) of Solanaceae and the sporophytic SI system (SSI) of Brassicaceae, a number of nucleotide sequences of functionally distinct alleles have been obtained. These sequences display extensive "transspecific polymorphism" (GSI, Ioerger *et al.* 1990; SSI, Dwyer *et al.* 1991). This may imply that some of the alleles diverged before divergence of species and, therefore, that the polymorphisms are extremely old. A somewhat similar pattern of transspecific polymorphism has been described in the MHC systems of mammals (*e.g.*, Klein 1980; Lawlor *et al.* 1988; McConnell *et al.* 1988). Takahata and Nei (1990) investigated by numerical simulations the properties of allelic genealogies under balancing selection and concluded that models like symmetrical overdominance or symmetrical frequency-dependent selection can produce the transspecific polymorphism observed in the MHC system. Relaxing the assumption of symmetry in selection coefficients they found even higher coalescence times but fewer alleles maintained. Takahata (1990) developed a theory of allelic genealogies under strong symmetrical balancing selection. He found the expected genealogical structure to be well described by the coalescence process for a random sample of neutral genes, but with a different time scale. His theory reduces the genealogical process to a Moran process in alleles, where the time unit is determined by allelic turnovers rather than generations, and he showed that it is approximated closely by just considering the probability of entry/loss of an allele at each turnover event and the average time between turnovers.

The MHC and GSI loci share an excess of nonsynonymous over synonymous substitutions in the putative specificity-determining regions [the antigen-binding region of MHC alleles (Takahata *et al.* 1992) and the hypervariable region of GSI alleles (Clark and Kao 1991)]. The pattern for sequences of SSI alleles in Brassicaceae is still ambiguous (Uyenoyama 1995). In theory, an increased nonsynonymous substitution rate is expected at sites conferring new specificities, because selection favors the invasion of new functional alleles (Maruyama and Nei 1981; Takahata 1990; Takahata and Nei 1990).

Vekemans and Slatkin (1994) used Wright's (1939) theory for the number of alleles maintained at a GSI locus to extend Takahata's theory of allelic genealogies to the case of GSI. They showed that the transspecific polymorphism observed in the Solanaceae is consistent with theoretical expectations. Richman *et al.* (1995, 1996a) surveyed sequence variation in natural populations of two species of Solanaceae with GSI, *Solanum carolinense* and *Physalis crassifolia.* They used the theory of Takahata to place broad bounds on the long-term effective population size of the two species, and from the shape of the allelic genealogies they suggested that

Corresponding author: Mikkel H. Schierup, Institute of Cell, Animal and Population Biology (ICAPB), University of Edinburgh, Ashworth Laboratory, King's Bldgs., W. Mains Rd., Edinburgh EH9 3JT, United Kingdom. E-mail: mikkel.schierup@biology.aau.dk

P. crassifolia had undergone a population bottleneck, perhaps during speciation (Richman and Kohn 1996; Richman *et al.* 1996b). Uyenoyama (1997) derived four ratios describing deviations from the genealogical structure expected under Takahata's theory and applied these to analyze the results of the above studies. She concluded that the genealogical tree of the alleles in *S. carolinense* is the more deviant from theoretical expectations, in that the terminal branches are longer than expected.

Sporophytic self-incompatibility in Brassicaceae also shows transspecific polymorphism (Dwyer *et al.* 1991; Kusaba *et al.* 1997). A number of alleles in mainly commercial varieties of *Brassica oleracea* and *B. campestris* have been sequenced and allelic genealogies constructed (Kusaba *et al.* 1997). However, theoretical expectations for the genealogy of alleles in SSI systems have not been developed. The application of Takahata's theory of allelic genealogies to SSI is not straightforward in these systems because, unlike in GSI, alleles are not identical in action. Dominance influences the determination of the incompatibility phenotype of the pollen and/or the style. In SSI, the phenotype of pollen is determined by the diploid genotype of the paternal plant and the ovule phenotype by the genotype of the maternal plant.

Schierup *et al.* (1997) and Vekemans *et al.* (1998) analyzed the allelic dynamics in finite populations using the three models of SSI shown in Table 1. The models were considered with and without fecundity selection (FS). With FS female fecundity is modified in that the probability of fertilization is proportional to the availability of compatible mates, and without FS maternal success is independent of the genotypic frequencies in the population. The three models are simplistic as they do not include phenomena like the mutual weakening of alleles in heterozygotes or the nonlinear dominance observed in some species (Bateman 1955; Sampson 1964), and they do not allow multiple alleles at any dominance level. Analyses of these models showed that dominance substantially decreases the number of alleles maintained and that recessive alleles attain higher equilibrium frequencies than dominant ones (Sampson 1974; Schierup *et al.* 1997). FS acts to increase the number of alleles, but it changes only the qualitative properties of the SSIdomcod model (Vekemans *et al.* 1998). The turnover process of alleles differs among the SSI models, but a good approximation was achieved by a model describing the "random-walk" of alleles on the dominance states (Schierup *et al.* 1997). This model is determined entirely by the probabilities of entry and loss of alleles depending on their position in the dominance hierarchy, and it is similar to the process applied by Takahata (1990).

We here extend and apply our previous results to reach a description of the genealogical structure of extant *S*-alleles in a finite population for the same three models of SSI. The differences in dominance, the varying degrees of asymmetry in selection coefficients, and the different turnover process of alleles are expected to affect both the coalescence times of alleles and the shape of their genealogical tree. We investigate both properties in our models by simulations. The substitution rates for mutations conferring new allelic specificities are investigated to formulate expectations for the rate of evolution in the specificity-determining region of an SSI system. We show that the random-walk approximation also predicts coalescence times accurately, and we use the approximation to predict the time since origin of extant alleles as a function of their current dominance levels. We investigate the effect of the specific mutation model in the models with dominance. In our earlier studies we assumed that the characteristics of a new mutant are independent of its origin, in genealogical terms, hence that the descendant of a given allele would have a random dominance level. Here, we also include a model of the opposite extreme, where the descendant allele is given a dominance level next to that of its ancestor. We show that the conclusions on the size and shape of the genealogy are virtually independent of the mutation model and discuss the detailed influence of the mutation model on the genealogy of alleles. This is relevant to inferences about the mutation model from the genealogy of alleles sampled from natural populations. Finally, we investigate the power of the statistics introduced by Uyenoyama (1997) for detecting deviations from the genealogical structure described by the neutral coalescent process.

MODELS

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. Four basic models of SI were investigated: the GSI model, where alleles act codominantly in the style and pollen expressing only its own allele, and the three different models of SSI (Table 1). In the two models involving dominance, the *dominance models*, the extant alleles in the population are sorted along a linear dominance hierarchy ($S_1 < S_2 < \ldots < S_n$) for determination of the phenotype; *i.e.*, allele S_i is recessive to S_j when $i < j$, allele S_1 is recessive, and allele S_n is dominant to all other alleles. Each of the models were considered with and without FS, and models with FS will explicitly be referred to as GSIFS, SSIcodFS, SSIdomFS, and SSIdomcodFS.

For the dominance models, two distinct mutation schemes were investigated. Under the *random* mutation scheme, a new allele arising by mutation is placed at random in one of the $n + 1$ possible states (including boundaries) within the dominance hierarchy, and the relative dominance levels of the extant alleles are shifted accordingly. Under the *one-step* mutation scheme, a new allele is placed on the dominance level above or below

TABLE 1

Determination of *S*-phenotypes of pollen and style in three models of sporophytic self-incombatibility

Model	Pollen phenotype (determined by diploid pollen parent)	Style phenotype (determined by diploid style)	Biological system resembling model
SSIcod	Codominance of alleles	Codominance of alleles	*Stellaria holostea* (Caryophyllaceae, Lundquist 1994)
SSIdom	Dominance hierarchy of alleles	Dominance hierarchy of alleles	Convolvulaceae (Kowyama *et al.* 1994)
SSIdomcod	Dominance hierarchy of alleles	Codominance of alleles	*Corylus avellana* (Mehlenbacher 1997); Brassicaceae (Thompson and Taylor 1966; Stevens and Kay 1989)

the dominance level of its parental allele. The random mutation scheme is used below, unless stated otherwise.

POPULATION SIMULATIONS

We simulated reproduction in a diploid plant population of size N with nonoverlapping generations using the Wright-Fisher model. In each generation, progeny were produced by the following procedure. One of the $2N$ genes is chosen at random as the maternal gamete and one as the pollen gamete. Compatibility is checked according to the self-incompatibility model, using the genotypes of the maternal and paternal plants. If the pollen grain is compatible, a new zygote is formed. If it is incompatible, then the other pollen phenotype of the same paternal parent is tried in the case of the GSI model, in which pollen of a given plant has two incompatibility types. If the pollen donor is incompatible, the ovule is discarded if FS is assumed; otherwise a new pollen genotype is drawn at random from the population. The process of choosing additional random pollen to test for compatibility with a given maternal plant is repeated 500 times in models without FS, and if no compatible pollen is found, the ovule is discarded, a new female gamete is chosen, and the process starts again. The mating process is repeated until N new zygotes are produced. A number of mutations, drawn from a Poisson distribution with mean $2Nu$, is then applied at random to genes in the offspring zygotes. Each mutation is assumed to produce a new functional allelic type according to the mutation scheme assumed.

Each run was started with $2N$ different alleles in the population and allowed to evolve for a number of generations until a mutation-selection-drift equilibrium was reached. Equilibrium was assumed when the average number of alleles in the population over subsequent intervals of 20 generations had oscillated 10 times. Beginning at that time genealogical information was recorded.

The genealogy was tracked in a forward manner by the method of Takahata and Nei (1990). This method assigns to each extant allele a vector that keeps record of the genealogical relationships among alleles (according to Maruyama and Nei 1981) and a vector that records the times when the mutations occurred. For the dominance models we added a vector recording the dominance level of parental alleles at the time of mutation. Simulations were continued for a number of generations equal to 5 times the expected coalescence time (T_c) of all alleles in the GSI model (Vekemans and Slatkin 1994, Equation 10). At this stage, a steady-state allelic genealogy was assumed and appropriate statistics were computed. The whole process was repeated 1000 times in most cases.

Number of alleles and coalescence times: Simulations were performed for each of the following six parameter sets, for the four models of SI with and without FS: $N = 50$, $u = 2 \times 10^{-5}$; $N = 100$, $u = 10^{-5}$; $N = 200$, $u = 5 \times 10^{-6}$; $N = 50$, $u = 2 \times 10^{-4}$; $N = 100$, $u = 10^{-4}$; $N = 200$, $u = 5 \times 10^{-5}$. We computed the mean number of extant alleles in the population (n), the mean coalescence time of all alleles [T_c, time back to the most recent common ancestor (MRCA)], and the average pairwise divergence times between alleles (T_d). To investigate the pattern of variation of T_c as a function of u, we also simulated cases with $N = 50$ and u from 2×10^{-3} to 2×10^{-5}.

Measures of genealogical structure: Simulations were performed for the four SI models without FS and for SSIdomcodFS with $N = 500$ and $u = 10^{-5}$. Two approaches were used to characterize the topological structure of allelic genealogies under the different models. The first was to draw, for each replicate run, a random sample of seven alleles from the population and compute the time intervals between successive coalescent events in the genealogy, $T(i)$, where i is the number of distinct lineages present during that interval ($2 \leq i \leq 7$). These time intervals were averaged over replicates, in units of the product $T(i)i(i - 1)$, and plotted against i. For a gene genealogy described by the coalescence process for a random sample of neutral genes, these standardized time intervals are independent and exponentially distributed with the same expectation $4N$ (Felsenstein 1992), so that the graph should be a hori-

zontal line. The second approach consists of computing ratios of divergence times, scaled by functions of allele number, which have expectations equal to one in the case of a random sample of neutral genes (Uyenoyama 1997). In addition to T_c and T_d, three statistics were computed when considering genealogies of all extant alleles in each replicate: the total time in the genealogy, T_t, the expected sum of the terminal branch lengths, S, and the average length of the basal branches emanating from the root, B. On the basis of these statistics, the following ratios were computed: $R_{PT} = 2T_d a_n / T_t$; $R_{ST} = Sa_n / T_t$; $R_{SD} = S(1 - 1/n)/T_c$; $R_{BD} = B(1-1/n)/T_c b_n$; with $a_n = \sum_{i=1}^{n-1} 1/i$ and $b_n = 1/n + \sum_{i=1}^{n-1} 1/i^2$. Averages and variances of these ratios over replicates were computed. The four ratios were also estimated from the branch lengths of genealogical trees obtained from published sequences of S-alleles in *B. oleracea* (23 alleles) and *B. campestris* (19 alleles; see Kusaba *et al.* 1997, and references therein). Amino acid sequences from each species were aligned using CLUSTAL X version 1.64 (Windows-based implementation of CLUSTAL W; Thompson *et al.* 1994), and phylogenetic trees were reconstructed using programs PROTDIST and FITCH from the PHYLIP package version 3.572 (Felsenstein 1995).

Maintenance of ancestral alleles in the population: As previous results suggested that some alleles in the dominance models have very long life spans (Schierup *et al.* 1997), we investigated the percentage of runs where one of the extant alleles was the MRCA of all extant alleles in the population. Simulations were performed for all four models without FS and for SSIdomcodFS with 500 replicates for $N = 50$, 100, and 500, and $u = 10^{-5}$. In addition to the proportion of replicates in which the MRCA of all extant alleles is present, the proportion of such cases where the extant MRCA is the most recessive allele in the population was determined in the dominance models.

Substitution rate: To investigate the pattern of variation of the substitution rate as a function of N, we performed simulations for all four models without FS and for SSIdomcodFS with $u = 10^{-5}$ and N varying between 30 and 5000. The substitution rate was computed by counting the number of mutations per gene that accumulated during the time T between the first generation of genealogical recording and the appearance of the MRCA of all extant alleles. This number divided by Tu estimated the substitution rate for nucleotide changes resulting in new alleles, or rate of codon substitution, expressed in units of $1/u$ generations.

Dominance levels of internal and terminal nodes in allelic genealogies: Simulations were performed to assess the effect of the mutation scheme on dominance levels of different nodes in the allelic genealogies. Three models (SSIdom, SSIdomcod, SSIdomcodFS) were compared for the random and one-step mutation schemes, using 500 replicate runs with $N = 200$ and $u =$

5×10^{-5}. For each replicate, six alleles were randomly sampled (five alleles for the SSIdom model), and the dominance levels of the five (four) internal nodes were extracted from the vector of dominance levels of parental alleles at the time of mutation. Dominance levels of alleles at any generation t were computed from the number n_t of alleles at that time in the population as the position from 1 to n_t in the dominance hierarchy, divided by n_t, and thus vary between $1/n_t$ and 1. For each replicate, we also computed the difference in dominance levels between the two alleles that first coalesced in the genealogies of all extant alleles. This quantity was compared to the average difference in dominance levels between random pairs of alleles, which is given by $\sum_{i=2}^{n_t} i(i - 1)/n_t^2 (n_t - 1)$.

RANDOM-WALK APPROXIMATION

Time since origin of alleles: The random-walk model is given in Schierup *et al.* (1997, Appendix B). The dynamics of the population of genes are approximated by a process in alleles where time is measured in units of allelic turnovers. A fixed number of alleles (the rounded average of the number of alleles observed for a given set of parameters) is assumed, and at each turnover, an allele goes extinct and is immediately replaced by a new invading allele. Given probabilities of entry and loss of alleles as a function of their dominance levels, a transition matrix of the allelic turnover process can be constructed, and then the expected life span of alleles as a function of their dominance level at mutation can be calculated. These were found to be in close agreement with population simulations (Schierup *et al.* 1997). The transposed matrix may be used to find the time elapsed since the origin of extant alleles as a function of their current dominance level, because the process back in time is very similar to the forward process with invasion and extinction interchanged. The time since origin of alleles is therefore calculated from the transposed transition matrix by the same procedure used for allelic life spans in Schierup *et al.* (1997).

Time to coalescence: The random-walk model also provides an estimate of T_c if the dynamics of incompatibility alleles in finite populations are simulated as a Moran process (Moran 1958; see Ewens 1979, p. 84). Consider a population of n_{RW} alleles and count time in number of allelic turnovers. At time points $t = 1, 2, 3, \ldots$, an allele was chosen to produce a new descendant allele by mutation and an allele was chosen to be lost from the population. In the GSI and the SSIcod models, all alleles in the population have equal probabilities of being chosen either to produce a descendant or to be lost. In the dominance models, the probability of choosing a given allele to produce a descendant is proportional to the deterministic equilibrium frequency of alleles at the corresponding dominance level, as com-

puted from recursion equations given by Schierup *et al.* (1997) and Vekemans *et al.* (1998). The dominance level of the new allele was chosen between 1 and n_{RW} by weighting each level by the relative probability of entry of alleles at the corresponding dominance level. The probability of choosing a given allele to be lost from the population is proportional to the relative probability of loss of alleles at the corresponding dominance level. The mean number of alleles n, the allelic turnover time, and relative probabilities of entry and loss of alleles as a function of their dominance were estimated as in Vekemans *et al.* (1998) by using a population simulation at equilibrium to track the fate of 50,000 alleles. The rounded n was used as n_{RW}. A new allele was placed in the dominance hierarchy after deleting the exiting allele, and the dominance levels of the extant alleles were shifted accordingly. The process was repeated until $t = 20n_{RW}^2$, and then the genealogical information was tracked in the same way as in the population simulations. The T_c-value obtained was then corrected for error in the rounding of n by multiplication by $n^2/(n_{RW})^2$. This correction is inspired by the dependence of T_c on n^2 in the codominant selection model (Takahata 1990), and it improved the correspondence to the population simulations.

RESULTS AND DISCUSSION

Coalescence times: Table 2 shows results for three population sizes ($N = 50$, 100, and 200) and two values of Nu, using the population simulation program. For each of the four models, the number of alleles, mean coalescence time of all alleles, T_c, and average pairwise divergence times between alleles, T_d (shown in units of N generations), are given both with and without FS. Values of T_c and T_d covary closely over the different models, so only the pattern of T_c needs to be discussed.

For the cases without FS, GSI, SSIcod, and SSIdom can be ordered with respect to T_c as SSIcod > GSI > SSIdom. The difference between SSIcod and GSI is about 10%, whereas SSIdom has markedly lower T_c-values ($\sim\frac{1}{3}$ of SSIcod). The differences are most pronounced for the smaller value of Nu. With constant Nu, these three models also share the pattern that T_c, when expressed in units of N generations, increases with N as observed by Vekemans and Slatkin (1994) for GSI, whereas in the SSIdomcod model it decreases. Thus, T_c is larger in SSIdomcod than in the SSIdom model for $N = 50$ and smaller for $N = 200$. Apart from this, values of T_c in the two dominance models are of comparable magnitude.

For the case with FS, the ranking orders of SSIcodFS, GSIFS, and SSIdomFS with respect to T_c and T_d are the same. Quantitatively, FS has a larger effect on T_c for the SSIdom and SSIcod models (T_c increases by 10–20%) than for GSI (T_c increases by $\sim$5%). However, the major effects of FS on T_c are in the SSIdomcod model. SSIdom-

codFS shows the same increase of T_c with N for Nu constant as the other models, and the increase in T_c by FS is more than threefold for $N = 200$. The T_c-values in the SSIdomcodFS model are distinctly larger than in the SSIdom and SSIdomFS models, making it intermediate between the SSIdom and SSIcod models for all parameters. The difference between SSIdomcod and SSIdomcodFS results from the qualitative differences in the dynamics of these models (Schierup *et al.* 1997; Vekemans *et al.* 1998). In the SSIdomcod model, the dominant alleles are most likely to invade while recessive alleles are most likely to be lost. This results in a directional turnover process of alleles, where most alleles enter the population as dominant, become gradually more recessive as more dominant alleles invade, and are finally lost as recessives. In the remaining dominance models (SSIdomcodFS, SSIdom, and SSIdomFS), however, dominant alleles are lost more easily than recessive alleles, and the differences in probabilities of entry and loss as a function of dominance level tend to balance such that an allele is likely to enter and exit at about the same dominance level. Because the SSIdomcodFS model shows qualitatively different behavior from SSIdomcod, it is included in all other figures and tables, whereas the GSIFS, SSIcodFS, and SSIdomFS models are mostly omitted in the subsequent discussions.

The pattern of T_c when u is varied and N kept constant ($N = 50$) is shown in Figure 1. For comparison, the expected curve for a symmetric overdominant locus with selection coefficient $s = 1$ is also shown. All models, except SSIdomcod, show the same pattern of increasing T_c with decreasing u, with the lines almost exactly parallel, and the same ordering of models as in Table 2. The SSIdomcod model, however, shows a smaller relative increase in T_c with decreasing u.

In Figure 2, T_c-values from simulations based on the random-walk model are compared with the population simulations of Table 2 for $4Nu = 0.004$ and $N = 50$, 100, and 200. The random-walk simulations underestimate T_c slightly in all models except SSIdomcod, but the discrepancy is smallest for $N = 200$. The fit of the random-walk model is sufficiently accurate that we have attempted extrapolation to larger population sizes, where the population simulations are too time consuming to perform (Figure 2; $N = 500$, 1000, 5000). The results for larger population sizes show that the ranking of the models is unchanged, but the absolute differences in T_c-values between the SSIdomcod model and the other models are further increased such that T_c for the SSIcod model is 15 times greater than T_c for the SSIdomcod model for $N = 5000$ ($u = 2 \times 10^{-7}$).

The random-walk simulations are based solely on the probabilities of entry and loss as functions of the dominance levels and the average probability that a new mutant invades. The close fit of T_c-values obtained from the random-walk simulations to the population simulations therefore strengthens the conclusion that the qual-

TABLE 2

Coalescence times in models of SI systems

			No fecundity selection			Fecundity selection		
$4Nu$	N	u	n	T_c/N	T_d/N	n	T_c/N	T_d/N
					GSI			
0.004	50	2.0×10^{-5}	4.93 ± 0.55	519.9 ± 341.4	327.4 ± 198.4	5.13 ± 0.53	537.4 ± 353.6	333.3 ± 199.5
	100	1.0×10^{-5}	6.38 ± 0.52	593.1 ± 394.6	347.6 ± 199.0	6.63 ± 0.67	613.8 ± 378.9	354.7 ± 188.4
	200	5.0×10^{-6}	8.49 ± 0.77	698.8 ± 435.9	397.0 ± 221.7	8.77 ± 0.76	729.5 ± 448.4	410.8 ± 222.1
0.04	50	2.0×10^{-4}	6.01 ± 0.85	88.6 ± 54.9	53.6 ± 30.9	6.26 ± 0.87	92.6 ± 59.5	54.5 ± 30.6
	100	1.0×10^{-4}	7.82 ± 0.92	102.5 ± 63.0	59.3 ± 32.5	8.09 ± 0.98	107.1 ± 68.6	59.9 ± 31.6
	200	5.0×10^{-5}	10.29 ± 1.07	119.5 ± 68.4	66.4 ± 33.7	10.64 ± 1.13	124.4 ± 74.1	68.3 ± 35.6
					SSIcod			
0.004	50	2.0×10^{-5}	5.66 ± 0.61	599.1 ± 375.4	362.0 ± 210.7	6.46 ± 0.59	648.1 ± 385.5	380.0 ± 203.2
	100	1.0×10^{-5}	7.20 ± 0.65	662.1 ± 403.5	382.6 ± 203.2	8.63 ± 0.77	772.2 ± 455.8	433.8 ± 225.8
	200	5.0×10^{-6}	9.28 ± 0.64	761.5 ± 434.6	425.4 ± 210.5	11.55 ± 0.81	883.5 ± 471.9	474.6 ± 216.5
0.04	50	2.0×10^{-4}	6.72 ± 0.82	97.1 ± 58.5	56.8 ± 30.2	7.85 ± 0.89	105.8 ± 60.4	60.8 ± 30.7
	100	1.0×10^{-4}	8.66 ± 0.97	110.9 ± 67.2	63.0 ± 33.1	10.35 ± 1.04	119.8 ± 66.2	67.3 ± 33.3
	200	5.0×10^{-5}	11.11 ± 1.02	134.6 ± 77.6	73.2 ± 37.4	13.84 ± 1.15	152.4 ± 84.2	82.8 ± 40.5
					SSIdom			
0.004	50	2.0×10^{-5}	2.82 ± 0.52	190.2 ± 156.2	149.0 ± 119.0	3.24 ± 0.49	206.1 ± 151.1	150.3 ± 104.5
	100	1.0×10^{-5}	3.53 ± 0.57	197.9 ± 134.2	139.6 ± 90.6	4.33 ± 0.55	249.2 ± 173.8	163.0 ± 103.8
	200	5.0×10^{-6}	4.57 ± 0.56	238.9 ± 147.9	153.6 ± 85.6	5.90 ± 0.67	299.5 ± 179.6	182.0 ± 97.1
0.04	50	2.0×10^{-4}	3.55 ± 0.71	33.9 ± 24.1	23.8 ± 15.7	4.16 ± 0.80	36.1 ± 24.5	24.0 ± 14.7
	100	1.0×10^{-4}	4.51 ± 0.79	40.1 ± 27.4	25.7 ± 15.9	5.50 ± 0.91	47.2 ± 30.9	28.7 ± 17.0
	200	5.0×10^{-5}	5.87 ± 0.88	47.3 ± 30.3	28.6 ± 16.3	7.33 ± 0.94	55.3 ± 33.6	32.2 ± 17.7
					SSIdomcod			
0.004	50	2.0×10^{-5}	3.37 ± 0.53	192.8 ± 139.5	138.0 ± 94.2	4.18 ± 0.53	310.7 ± 215.2	209.0 ± 133.6
	100	1.0×10^{-5}	4.05 ± 0.62	168.8 ± 126.9	112.0 ± 75.3	5.65 ± 0.66	401.4 ± 252.2	248.8 ± 136.9
	200	5.0×10^{-6}	4.70 ± 0.70	144.7 ± 96.6	92.2 ± 56.8	7.61 ± 0.72	488.2 ± 297.2	286.6 ± 143.0
0.04	50	2.0×10^{-4}	4.40 ± 0.83	39.6 ± 25.5	25.7 ± 14.9	5.21 ± 0.81	58.9 ± 38.7	36.9 ± 22.3
	100	1.0×10^{-4}	5.47 ± 0.95	37.8 ± 23.5	23.5 ± 13.2	7.00 ± 0.93	70.2 ± 42.0	40.9 ± 22.0
	200	5.0×10^{-5}	6.75 ± 1.10	36.8 ± 24.1	21.7 ± 12.1	9.46 ± 1.00	85.7 ± 49.7	49.6 ± 25.7

Number of alleles (n), total coalescence time (T_c), and pairwise divergence time (T_d) for the four models with and without fecundity selection. The random mutations scheme is used. Means are given $\pm$ standard deviations.

itatively different pattern observed in the SSIdomcod model must be caused primarily by the directional turnover process of alleles in this model (Schierup *et al.* 1997). Takahata (1990) showed for symmetrical overdominance that the expected number of allelic turnovers until coalescence is proportional to the square of the number of extant alleles. This is because all alleles are equally likely to be lost in each generation. Therefore, T_c is also expected to be proportional to n^2 in GSI and SSIcod, where all alleles are equivalent. In the SSIdomcod model, however, the dependence of T_c on n may be reduced because of the directionality of the turnover process. This deterministic influence becomes stronger in larger populations, and this explains why T_c/N decreases with N in the SSIdomcod model whereas it increases in any other model (Figure 2).

Overall measures of genealogical structure: Figure 3 shows the standardized time intervals between coalescence events for seven alleles sampled from 1000 replicate simulations of $N = 500$, $u = 10^{-5}$, for each of the models. None of the models departs very much from
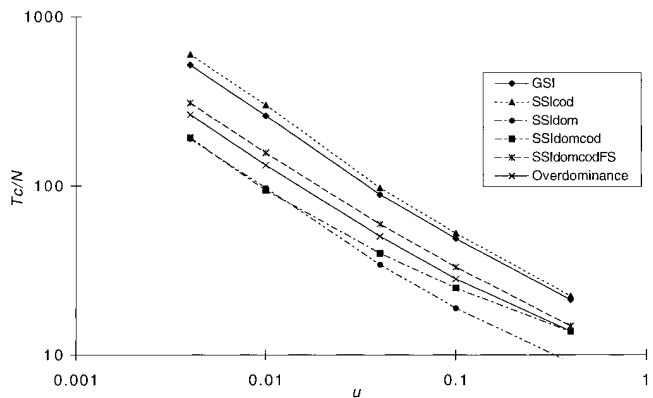

Figure 1.—Coalescence times of all alleles, T_c, expressed in units of N generations, as a function of the mutation rate, u, ranging from 2×10^{-3} to 2×10^{-5}. Values are shown for all models without fecundity selection as well as for SSIdomcodFS. Expected values for symmetrical overdominance with selection coefficient $s = 1$ are also shown.
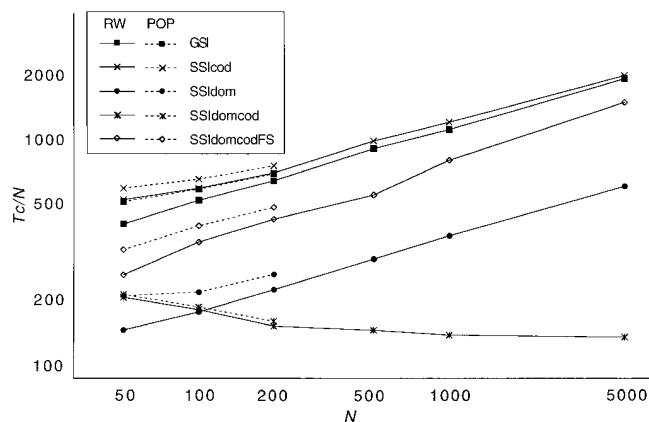
Figure 2.—Random-walk simulations. Comparison between T_c/N estimated from simulations using the random-walk (RW) model (solid lines) and the population (POP) simulations (from Table 1, dashed lines) for $N = 50$, 100, and 200. For the random-walk simulations, the rounded value of n observed from simulations was used as the number of allelic classes and the values of T_c subsequently corrected for the error induced by this rounding (see text). For $N > 200$, population simulations with genealogical recording are too time consuming to perform. The random mutation scheme is used. Nu is kept constant at 0.001.

the horizontal pattern expected for a gene genealogy described by the coalescence process for a random sample of neutral genes (Felsenstein 1992). There is a tendency in the SSIdom models for the time intervals at the top of the genealogy ($Ti = 2,3,4$) to be relatively shorter than expected for a neutral gene genealogy, but the effect is very small compared to the large variance of the time intervals.

Uyenoyama (1997) suggested four ratios that describe different aspects of the shape of the genealogy. Mean values of these ratios ±SD are shown, for the same parameter set, in Figure 4. Again, the differences among the models are small compared to standard deviations. In most cases, the mean values of the ratios are close to one, the expected value for neutral gene genealogies. For GSI, the results agree with the previously reported results by Uyenoyama (1997). The main departure from unity is again observed with the SSIdom and SSIdomFS models, where R_{ST} and R_{SD} are increased relative to the other models. These ratios measure the length of the terminal branches relative to the total length of the genealogy and the size of the genealogy, respectively. The increase in R_{ST} and R_{SD} for the two SSIdom models therefore agrees with Figure 3 in that the upper branches in the genealogy are relatively shorter in the SSIdom models than in the other models.

Time since origin of extant alleles: Figure 5 shows the expected times since origin of extant alleles as a function of their dominance levels at the present time for one parameter set ($N = 200$, $u = 5 \times 10^{-6}$), in the random-walk model. These times differ less than 10% from the results obtained from population simulations.
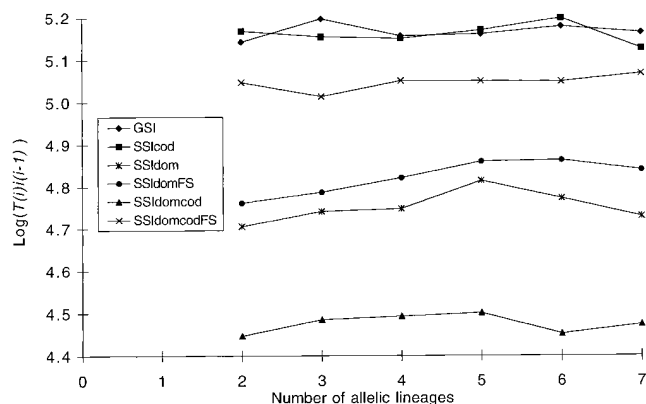

Figure 3.—Standardized time intervals, $T(i)i(i - 1)$, between coalescent events in genealogies from random samples of seven alleles in simulations with $N = 500$ and $u = 10^{-5}$, as a function of the number of allelic lineages during the corresponding interval. Values are shown for all models without fecundity selection as well as for SSIdomcodFS.

Figure 5 shows that, in all models, recessive alleles are expected to have an older origin than dominant alleles. For the SSIdomcod model, this appears at first to contradict that the expected life span of an allele is largest if the allele is dominant when it arises (Schierup et al. 1997). However, extant recessive alleles observed in the population in the SSIdomcod model are unlikely to be new mutants because very few alleles enter as recessive. Instead, recessives consist primarily of alleles that entered the population as dominant and have, through the directional turnover process, become recessive. Thus the recessive alleles have an origin more distant than that of dominant alleles that are likely to have invaded recently.

The effect of asymmetrical interaction among alleles: Table 3 shows the proportion of runs where the MRCA allele is still present in the population, for $u = 10^{-5}$ and various population sizes. This proportion clearly shows dependence on the number of alleles and the asymmetry in their interaction, and we therefore also included the number of alleles present. For the dominance models, extant recessive alleles are older and present at higher frequencies than are dominant alleles, and we include (in parentheses) the frequency with which an extant MRCA belongs to the most recessive class.

The results show that the probability that the MRCA is present drops rapidly with the number of alleles in the models with codominance. The number of allelic turnovers that have occurred in these models since the first occurrence of the MRCA is proportional to n^2, as explained above. The probability that a given allele goes extinct is $1/n$ at each turnover, because all alleles are equivalent in the codominant models, and when n is large, the number of turnovers since the MRCA is so large that each initial allele is likely to have gone extinct. In the SSIdom model, however, the probability that the MRCA is present is very high, and this allele is usually
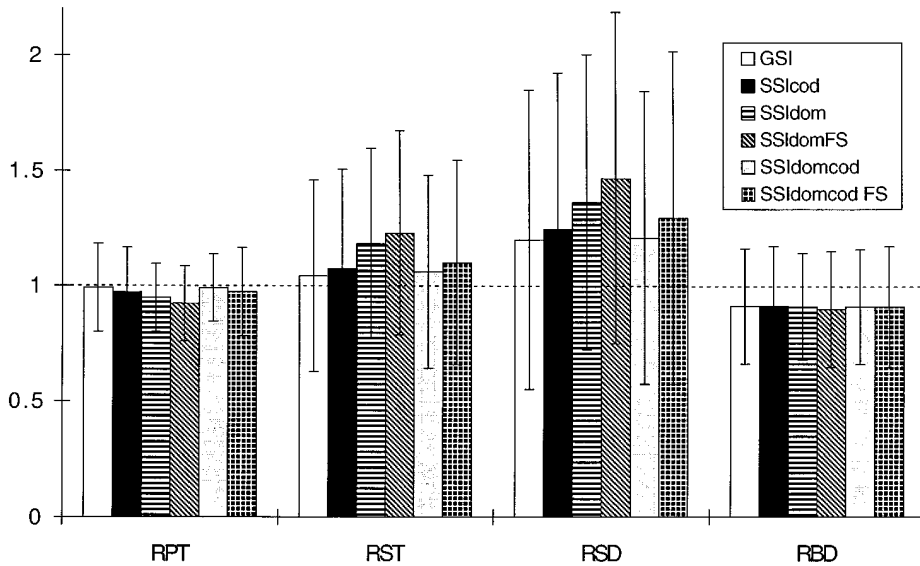
Figure 4.—Ratios of divergence time ($\pm$SD), scaled by functions of allele numbers according to Uyenoyama (1997), for allelic genealogies in simulations with $N = 500$ and $u = 10^{-5}$.

the most recessive. This happens because recessive alleles in the SSIdom model have a much higher life span and higher allele frequency than dominant alleles (Schierup *et al.* 1997). Because of its high frequency, the most recessive allele is the most likely ancestor of each new invading allele, and its long life span assures that it can stay sufficiently long in the population to become the MRCA of all extant alleles. In the SSIdomcodFS model, a pattern similar to the SSIdom model is observed, but the probability that the MRCA is present is smaller, because the frequency of the most recessive allele is lower and it has a shorter life span. This model has the same dynamics as the SSIdom model (Vekemans *et al.* 1998). The SSIdomcod model, in contrast, shows a different pattern in that the probability that the MRCA is present is much lower, and the MRCA is not particularly likely to be recessive if it is present. This is again caused by the directional turnover process, where any given allele is doomed to become more and more recessive and eventually to be replaced by new dominant alleles.

Substitution rates: The substitution rate in units of $1/u$ generations as a function of the population size ($u = 10^{-5}$) is shown in Figure 6. In all cases, the rate of substitution of alleles exceeds the relative rate of one expected for neutral alleles (Kimura 1968). The SSIdomcod model is peculiar in having very high substitution rates. The pattern of substitution rates in the other models agrees with selection intensities for new alleles, *i.e.*, SSIcod $\geq$ GSI > SSIdomcodFS > SSIdom. The high substitution rate in the SSIdomcod model is again caused by the directional substitution process, because a substitution occurs whenever an allele is replaced by its dominant descendants. For all models the substitution rate initially increases with population size because selection for new alleles increasingly outweighs genetic drift. For large populations, the substitution rate decreases again because the intensity of selection fa

voring new alleles decreases with increasing number of extant alleles. The population size yielding the maximum substitution rate is therefore expected to depend on u. For this mutation rate ($u = 10^{-5}$), an intermediate population size of the order of 500–1000 yields a maximum substitution rate, but the substitution rate decreases only slowly for $N > 1000$.

The effect of the mutation scheme: Table 4 shows the effect of the mutation scheme on various aspects of the genealogical structure and allelic dynamics. Under the one-step mutation scheme, a descendant allele is given a dominance level next to that of its ancestor, and this causes a relative increase of recessive mutations because recessive alleles are more common. However, this does not affect the number of alleles maintained in any model (Table 4). Values of T_c are insensitive to the mutation scheme in the SSIdom model and the SSIdomcodFS model; *i.e.*, the differences are less than 15%. The SSIdomcod model, in contrast, shows a more
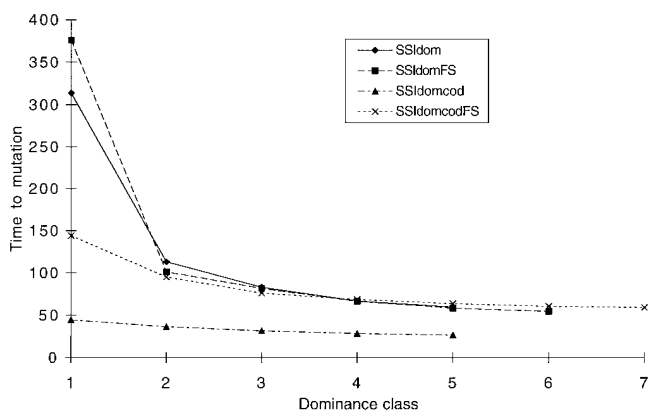

Figure 5.—Time since the origin of alleles by mutation (in N generations) as a function of their dominance level in the dominance models. Results are shown for $N = 200$, $u = 5 \times 10^{-6}$ and are calculated from probabilities of entry and loss using the random-walk model.

TABLE 3

The proportion of runs (percentage) where one of the extant alleles is the common ancestor of all alleles present in the population as a function of the population size ($u = 10^{-5}$) for the different models

	N = 50		N = 100		N = 500	
Model	n	MRCA present (%)	n	MRCA present (%)	n	MRCA present (%)
GSI	4.74	25	6.38	11	14.16	0
SSIcod	5.37	18	7.26	8	15.16	0
SSIdom	2.65	84 (86)	3.51	78 (83)	7.83	56 (88)
SSIdomcod	3.17	24 (42)	4.07	15 (37)	7.82	6 (33)
SSIdomcodFS	4.02	51 (62)	5.62	35 (56)	12.98	9 (44)

For the dominance models, the proportion of such cases where the allele is also the most recessive allele is given in parentheses. *n*, number of alleles.

than threefold decrease in T_c with the one-step mutation scheme, indicating that the turnover process in this model becomes even more deterministic. This is because successful dominant alleles are generated exclusively from extant long-lived dominant alleles under the one-step mutation scheme, and they force other alleles to become more recessive and hence more prone to loss. Substitution rates are always lower under the one-step scheme because the mutation rate to dominant alleles that are more likely to be successful is reduced as compared to the random scheme. These results were also observed with other sets of population sizes and mutation rates (results not shown).

The lower part of Table 4 shows for each model the average dominance level at internal nodes (at the time of mutation) in the genealogical tree of a sample of alleles (five in SSIdom, and six in SSIdomcod and SSIdomcodFS). The table starts from the tip in the genealogy (two lineages present at the time of coalescence, *i.e.*, MRCA) to the bottom of the tree (five or six lineages in the genealogy present at the time of coalescence, *i.e.*, the first coalescence). In the SSIdom model for both mutation schemes, the dominance level decreases steadily when moving up the tree (back in time), and this
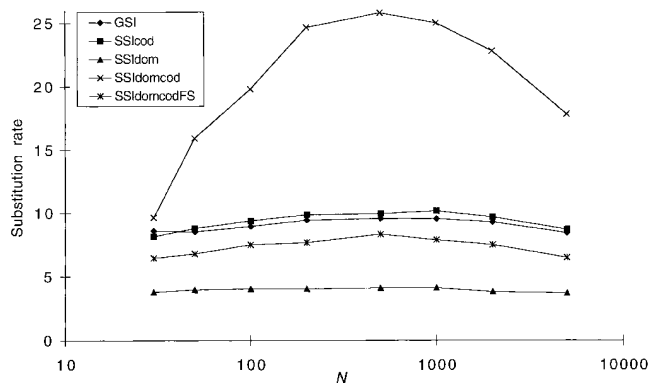
agrees with the observation that the most recessive allele is often the MRCA (Table 4, rows 3–5). The results for the SSIdomcodFS model are qualitatively similar to those for the SSIdom model, but again, the SSIdomcod model shows a deviant pattern, in that the dominance level increases when moving up the tree. This pattern becomes more pronounced with the one-step mutation scheme, such that the MRCA was almost certainly the most dominant allele at the time of coalescence of all alleles. Furthermore, the probability that this MRCA is still present is very large. However, it is not particularly likely that it is still dominant (18.1%), because alleles change dominance levels through their life span in the SSIdomcod model.

The difference in dominance levels between the two alleles involved in the first (most recent) coalescence is shown in Table 4, row 6. This measure is included because it can be estimated from genealogies reconstructed from sequence data of extant alleles. For comparison, the average difference in dominance levels between two randomly chosen alleles is 0.39 for six alleles present and 0.37 for nine alleles present. The observed value for the SSIdom model with random mutation is slightly higher than 0.39, showing that there is a tendency for the first coalescence to be between a dominant and a recessive allele. This is expected because recessive alleles attain high frequencies and generate many descendants and, among these, dominant descendants are more likely to invade the population. However, this effect is barely detectable. With the one-step mutation scheme, the difference in dominance levels is much smaller. This reflects the fact that the first coalescence is (almost always) between alleles at two adjacent dominance levels, not surprisingly given the mutation process assumed.

Figure 6.—Substitution rate, expressed in units of $1/u$ generations, as a function of population size, *N*, varying between 30 and 5000 for $u = 10^{-5}$. Values are shown for all models without fecundity selection as well as for SSIdomcodFS.

CONCLUSIONS

The expected T_c for alleles in models of SSI are generally long and comparable to those for GSI. Transspecific polymorphism due to a deep root of the genealogical

TABLE 4

Properties of allelic genealogies under two different mutation schemes for SSI models with dominance

	SSIdom random	SSIdom one-step	SSIdomcod random	SSIdomcod one-step	SSIdomcodFS random	SSIdomcodFS one-step
n	5.83	5.77	6.69	6.71	9.35	9.28
T_c/N[a]	42.1	47.9	38.5	13.7	86.1	75.7
Substitution rate[a]	3.3	2.7	14.9	9.9	6.2	5.0
Allele older than T_c (%)[b]	54.2	49.8	3.8	75.0	16.6	14.2
Alleles recessive (%)[c]	81.9	69.5	31.6	18.1	57.8	53.5
Difference in dominance level[d]	0.44 ± 0.22	0.18 ± 0.03	0.36 ± 0.20	0.15 ± 0.03	0.38 ± 0.23	0.11 ± 0.01
Dominance levels of internal nodes[e]						
2	0.22	0.22	0.76	1.00	0.23	0.16
3	0.29	0.34	0.78	0.92	0.27	0.27
4	0.34	0.44	0.73	0.84	0.34	0.35
5	0.39	0.49	0.66	0.77	0.39	0.43
6	—	—	0.61	0.69	0.43	0.48

Results are shown for $N = 200$, $u = 5 \times 10^{-5}$. n, number of alleles.

[a] The average observed in the simulations.

[b] The percentage of simulations where an extant allele is the common ancestor of all alleles present.

[c] The percentage of such alleles that are recessive.

[d] The average difference in dominance levels between the two most closely related, extant alleles.

[e] Average dominance levels of internal nodes (at the time of mutation) are shown for the reconstructed trees from random samples of five or six alleles; for explanation see text.

tree of alleles is therefore equally likely in GSI and SSI. The only exception may be for species with an incompatibility system that may be described by the SSIdomcod model. Values of T_c are not likely to distinguish any of the other models, because the differences are small, usually less than threefold. The current and historical mutation rates probably differ among different SSI systems, and because the mutation rate has a large effect on T_c, such differences may well override the differences in T_c seen in comparisons among models. One striking observation, though, is that differences between sexes in the expression of dominance (SSIdomcod) can lead to a directional turnover process of alleles, which results in reduced values of T_c, especially when the population size is large and the mutation rate is small.

In Brassica species, sex-specific differences in dominance have been observed (Hatekayama *et al.* 1998). This might suggest that the allelic genealogy for species in this family should be shorter than that for species with GSI for similar rates of mutation and long-term effective population sizes. The extensive transspecific polymorphism observed among sequences of alleles of *B. campestris* and *B. oleracea* does not support this suggestion. However, the SSI system in Brassicaceae deviates markedly from our simplistic sex-asymmetric SSIdomcod model (Thompson and Taylor 1966; Stevens and Kay 1989). Furthermore, the addition of FS in the SSIdomcod model completely changes the turnover process and substantially increases T_c. The level of FS would

need to be estimated for each species by studying the pollination biology (Vekemans *et al.* 1998). Therefore, it is premature from our models to draw conclusions about the expected size of the genealogy in Brassicaceae before the genetics of the system and the level of FS are known sufficiently well to allow it to be modeled more precisely than by the SSIdomcod model. Such a description of the genetics requires extensive crossing experiments.

The occurrence of transspecific polymorphism is usually taken as evidence that a set of allelic lines inherited by both descendent species during the speciation process have survived until the present time. However, introgression of *S*-alleles may be an alternative hypothesis for the transspecific polymorphism in Brassicaceae. Hybridization is common in the Brassicaceae. *B. oleracea* and *B. campestris* can be hybridized in the laboratory, and both species hybridize with the self-compatible *B. napus* [*B. oleracea* and *B. napus* with difficulty (McNaughton 1976), *B. campestris* and *B. napus* readily in the wild (Jorgensen and Andersen 1994; Landbo *et al.* 1996)]. Kusaba *et al.* (1997) found that two SLG alleles from each of these species differ by only 4.4% at the deduced amino acid sequence, with no differences found in the hypervariable and putative specificity-encoding domains. This low divergence may be difficult to reconcile with the split of these two species, which has tentatively been estimated as on the order of 10 mya (Uyenoyama 1995). If hybridization occurs at all, introgression of strongly selected alleles (*e.g.*, rare self-

incompatibility alleles) and genes closely linked to the S-alleles may take place even when hybrid viability or fertility is very low; even neutral markers can cross partial postzygotic barriers between populations or species (Bengtsson 1985; Barton and Bengtsson 1986). The quantitative effects of this hypothesis are currently under study.

The expected elevation in substitution rate at nonsynonymous sites involved in the determination of specificity depends on the model of incompatibility. The directional substitution process of alleles in the SSIdomcod model may lead to accelerated evolution at these sites that is not explained by the selection intensity of the balancing selection alone. The magnitude of this effect is surprisingly high and may be detectable if species with sex differences in dominance levels (perhaps in Brassicaceae or Corylaceae) are compared to species with similar dominance relations in both sexes (*e.g.*, Convolvulaceae).

Dominance in both pollen and style (SSIdom model) causes decreased T_c-values compared to the codominant models. This is mainly due to the decrease in the number of alleles in this model. The SSIdom model can be thought of as a case of asymmetrical balancing selection, where dominant alleles are more strongly favored when rare and more strongly disfavored when common. If selection coefficients of individual alleles are constant, asymmetrical overdominance decreases the number of alleles maintained in finite populations, in that a subset of strongly selected alleles survives longer (Robertson 1962). Takahata and Nei (1990) found in a study of asymmetrical overdominance that despite the decrease in the number of alleles, T_c was increased by increased asymmetry of alleles, in contrast to our results with the SSIdom model. In their model, the selection coefficients are constants assigned at the time of mutation, and because strongly selected alleles are preferentially kept in the population, the mean selection coefficients of extant alleles increase and may not be at equilibrium at any given time. Consequently, the effective mutation rate may decrease, accounting for the observations reported by Takahata and Nei (1990).

The overall shape of the allelic genealogies observed was quantified by two different approaches in the different models, namely the standardized time intervals between coalescence events and Uyenoyama's (1997) ratios. The results clearly show that none of these approaches can reliably distinguish between models, using experimental data, because the models behave similarly. However, the insensitivity of Uyenoyama's ratios to the presence or absence of dominance in SSI models broadens their application to detect significant deviations from expectation in genealogies of SSI systems. Uyenoyama (1997) showed, using these ratios, that the genealogical tree of *S. carolinense* has lower branches significantly longer than expected given the selection inherent in a GSI system. Table 5 shows the ratios esti-

TABLE 5

Ratios of divergence times in two Brassica species

	Number of sequences	R_{PT}	R_{ST}	R_{SD}	R_{BD}
Brassica oleracea	23	0.45	2.74	5.15	0.58
Brassica campestris	19	0.46	3.02	8.87	0.04

mated in *B. oleracea* and *B. campestris.* In both species R_{SD}, which measures the length of terminal branches relative to the size of the tree, is well above the expected values for any of the models of SSI (Figure 4), indicating that terminal branches are longer than expected under an infinite alleles model as in the case of *S. carolinense.*

The random-walk model captures the essential information with respect to allelic dynamics and coalescence times for all the models investigated (including the codominant models). This shows that the probabilities of invasion and extinction of alleles as a function of their dominance levels is the main determinant of S-allele evolution. In the models of SSI, these probabilities currently must be estimated from simulations, and a theory in line with Takahata's (1990) theory for MHC would ease their estimation considerably.

When the dominance levels of the alleles are considered, the superficial similarity in genealogical shape disappears. In the SSIdom and SSIdomcodFS models, a recessive allele is generally very old, and it is likely to be the MRCA of all extant alleles. In the SSIdomcod model, recessive alleles also tend to be oldest, but the age differences among alleles are small. The age of an allele cannot be inferred experimentally from the reconstruction of an intraspecific genealogy, but a joint genealogy of alleles from more species may indicate allelic age. This requires that the domain of allelic specificity of the alleles is identified, such that allele homology can be identified in comparisons among species. The identification of allelic specificity may, however, be obtained from transformation experiments where S-alleles are transferred between species (see Kao and McCubbin 1996). The prediction is that a recessive allele is more likely to be preserved in different species and to show transspecific polymorphism as a reflection of an ancient state of the population. On the other hand, if hybridization between the species occurs or has occurred after speciation, then introgression of S-alleles is likely because of the selective advantage of rare alleles with a variant specificity. The probability of entrance of a new dominant allele is higher, because rare dominant alleles are selected more strongly than rare recessive alleles. Thus, when transspecific polymorphism is a reflection of genetic introgression, then the shared allele is expected to be placed high in the dominance hierarchy at the time of introgression.

The details of the mutation model do not appear to

influence T_c very much, as judged from the two extreme mutation schemes modeled in this study. On the other hand, from a genealogical tree of the alleles where the dominance levels of the nodes are known, inferences on the mutation scheme may be made. If closely related alleles have similar dominance levels, a mutation scheme, where the dominance level of a descendant allele is close to that of its ancestor is likely.

LITERATURE CITED

Barton, N. H., and B. O. Bengtsson, 1986 The barrier to gene exchange between hybridizing populations. Heredity **56:** 357–376.

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

Bengtsson, B. O., 1985 The flow of genes through a genetic barrier, pp. 31–42 in *Evolution: Essays in Honour of John Maynard Smith*, edited by J. J. Greenwood and M. Slatkin. Cambridge University Press, Cambridge, UK.

Clark, A. G., and T.-H. Kao, 1991 Excess nonsynonomous substitution at shared polymorphic sites among self-incompatibility alleles of Solonaceae. Proc. Natl. Acad. Sci. USA **88:** 9823–9827.

Dwyer, K. G., M. A. Balent, J. B. Nasrallah and M. E. Nasrallah, 1991 DNA sequences of self-incompatibility genes from *Brassica campestris* and *B. oleracea*: polymorphism predating speciation. Plant Mol. Biol. **16:** 481–486.

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

Ewens, W. J., 1979 Mathematical population genetics. Springer-Verlag, Berlin.

Felsenstein, J., 1992 Estimating effective population size from samples of sequences: inefficiency of pairwise and segregating sites as compared to phylogenetic estimates. Genet. Res. **59:** 139–147.

Felsenstein, J., 1995 PHYLIP (Phylogeny Inference Package) version 3.572. Distributed over the World Wide Web, Seattle.

Hatakeyama, K., M. Watanabe, T. Takahasi, K. Ojima and K. Hinata, 1998 Dominance relationships between *S*-alleles in self-incompatible *Brassica campestris* L. Heredity **80:** 241–247.

Ioerger, T. R., A. G. Clark and T.-H. Kao, 1990 Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc. Natl. Acad. Sci. USA **87:** 9732–9735.

Jorgensen, R. B., and B. Andersen, 1994 Spontaneous hybridization between oilseed rape (*Brassica napus*) and weedy *B. campestris* (Brassicaceae): a risk of growing genetically modified oilseed rape. Am. J. Bot. **81:** 1620–1626.

Kao, T.-H., and A. G. McCubbin, 1996 How flowering plants discriminate between self and non-self pollen to prevent inbreeding. Proc. Natl. Acad. Sci. USA **93:** 12059–12065.

Kimura, M., 1968 Evolutionary rate at the molecular level. Nature **217:** 624–626.

Klein, J., 1980 Generation of diversity at MHC loci: implications for T cell receptor repertoires, pp. 239–253 in *Immunology 80*, edited by M. Fougerreau and J. Dausset. Academic Press, London.

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.

Kusaba, M., T. Nishio, Y. Satta, K. Hinata and D. Ockendon, 1997 Striking sequence similarity in inter- and intra-specific comparisons of class I *SLG* alleles from *Brassica oleracea* and *Brassica campestris*: implications for the evolution and recognition mechanism. Proc. Natl. Acad. Sci. USA **94:** 7673–7678.

Landbo, L., B. Andersen and R. B. Jorgensen, 1996 Natural hybridisation between oilseed rape and a wild relative: hybrids among seeds from weedy *B. campestris*. Hereditas **125:** 89–91.

Lawlor, D. A., F. E. Ward, P. D. Ennis, A. P. Jackson and P. Parham, 1988 *HLA-A* and *B* polymorphisms predate the divergence of humans and chimpanzees. Nature **355:** 268–271.

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.

Maruyama, T., and M. Nei, 1981 Genetic variability maintained by mutation and overdominant selection in finite populations. Genetics **98:** 441–459.

McConnell, T. J., W. S. Talbot, R. A. McIndoe and E. K. Wakeland, 1988 The origin of MHC class II polymorphisms within the genus *Mus*. Nature **332:** 651–654.

McNaughton, I. H., 1976 Swedes and rapes, pp. 53–56 in *Evolution of Crop Plants*, edited by N. W. Simmonds. Longman, London.

Mehlenbacher, S. H., 1997 Revised dominance hierarchy for *S*-alleles in *Corylus avellana* L. Theor. Appl. Genet. **94:** 360–366.

Moran, P. A. P., 1958 Random processes in genetics. Proc. Camb. Philos. Soc. **54:** 60–71.

Richman, A. D., and J. R. Kohn, 1996 Learning from rejection: the evolutionary biology of single-locus incompatibility. TREE **11:** 497–502.

Richman, A. D., 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 Solonaceae. Science **273:** 1212–1216.

Robertson, A., 1962 Selection for heterozygotes in small populations. Genetics **47:** 1291–1300.

Sampson, D. R., 1964 A one-locus self-incompatibility system in *Raphanus raphinistrum*. Can. J. Genet. Cytol. **6:** 435–445.

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

Schierup, M. H., X. Vekemans and F. B. Christiansen, 1997 Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. Genetics **147:** 835–846.

Stevens, J. P., and Q. O. N. Kay, 1989 The number, dominance relationships and frequencies of self-incompatibility alleles in a natural population of *Sinapsis 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.

Takahata, N., and M. Nei, 1990 Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics **124:** 967–978.

Takahata, N., Y. Satta and J. Klein, 1992 Polymorphism and balancing selection at major histocompatibility complex loci. Genetics **130:** 925–938.

Thompson, K. F., and J. P. Taylor, 1966 Non-linear dominance relationships between *S* alleles. Heredity **21:** 345–362.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Uyenoyama, M. K., 1995 A generalized least-squares estimate for the origin of sporophytic self-incompatibility. Genetics **139:** 975–992.

Uyenoyama, M. K., 1997 Genealogical structure among alleles regulating self-incompatibility in Angiosperms. Genetics **147:** 1389–1400.

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

Vekemans, X., M. H. Schierup and F. B. Christiansen, 1998 Mate availability and fecundity selection in multi-allelic self-incompatibility systems in plants. Evolution **52:** 19–29.

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

Communicating editor: M. K. Uyenoyama