A Generalized Least-Squares Estimate for the Origin of Sporophytic Self-Incompatibility

Marcy K. Uyenoyama

Department of Zoology, Duke University, Durham, North Carolina 27708-0325

Manuscript received June 7, 1994
Accepted for publication October 7, 1994

ABSTRACT

Analysis of nucleotide sequences that regulate the expression of self-incompatibility in flowering plants affords a direct means of examining classical hypotheses for the origin and evolution of this major feature of mating systems. Departing from the classical view of monophyly of all forms of self-incompatibility, the current paradigm for the origin of self-incompatibility postulates multiple episodes of recruitment and modification of preexisting genes. In Brassica, the *S* locus, which regulates sporophytic self-incompatibility, shows homology to a multigene family present both in self-compatible congeners and in groups for which this form of self-incompatibility is atypical. A phylogenetic analysis of *S*-allele sequences together with homologous sequences that do not cosegregate with self-incompatibility permits dating the change of function that marked the origin of self-incompatibility. A generalized least-squares method is introduced that provides closed-form expressions for estimates and standard errors for function-specific divergence rates and times of divergence among sequences. This analysis suggests that the age of the sporophytic self-incompatibility system expressed in Brassica exceeds species divergence within the genus by four- to fivefold. The extraordinarily high levels of sequence diversity exhibited by *S* alleles appears to reflect their ancient derivation, with the alternative hypothesis of hypermutability rejected by the analysis.

SELF-incompatibility, a major feature of angiosperm evolution and an important determinant of the mating systems of many agricultural crops, has been the focus of a long and distinguished classical genetic tradition. Whereas self-incompatibility may derive from a variety of genetic mechanisms for the regulation of fertilization in hermaphroditic plants, the molecular genetic basis and physiological mechanism of the phenomenon is best known in homomorphic systems: sporophytic self-incompatibility (SSI) in the Brassicaceae and gametophytic self-incompatibility (GSI) in the Solanaceae (reviewed by CLARKE and NEWBIGIN 1993; HINATA *et al.* 1993; SIMS 1993). The recent advent of molecular-level information constitutes a major conceptual as well as empirical breakthrough which warrants a fundamental re-examination of theories for the origin of self-incompatibility. I present a phylogenetic analysis of the multigene family associated with the brassicaceous form of sporophytic self-incompatibility. This study provides generalized least-squares estimates for the time of origin of this SSI system and the rate of evolution of *S* alleles.

Classical views of the origin of self-incompatibility: In a very influential paper, WHITEHOUSE (1951) proposed that the origin of homomorphic systems of self-incompatibility is inseparable from the origin of flowering plants, even suggesting that the avoidance of self-fertilization constituted the key evolutionary innovation that permitted the explosive rise of the angiosperms during the Cretaceous. Although recognizing that self-fertilization can be avoided through other mechanisms, including spatial or temporal separation of male and female expression, WHITEHOUSE argued that homomorphic self-incompatibility with large numbers of specificity-determining alleles entails the least restriction of cross-fertilization through pollen. Self-compatibility is regarded under this hypothesis as a loss of function that appeared during the period of relaxed selection that purportedly followed the domination of the angiosperms over the gymnosperms. Other mechanisms for the avoidance of inbreeding, including heteromorphic self-incompatibility and dioecy, then arose in secondarily self-compatible species in response to pressures favoring outbreeding.

PANDEY (1960) proposed that SSI evolved directly from GSI. Both systems prevent fertilization by pollen that express specificities held in common with the seed parent. Pollen specificity is determined during the gametophyte stage (by genes held by individual pollen grains) under GSI and during the sporophyte stage (by genes held by the pollen parent) under SSI. This transition could have occurred through a minor modification of the timing of determination of pollen specificity (before cytokinesis in

Corresponding author: Marcy K. Uyenoyama, Department of Zoology, Box 90325, Duke University, Durham, NC 27708-0325.
E-mail: marcy@daruma.zoo.duke.edu

the pollen mother cell rather than after). Under PANDEY's scenario, SSI evolved directly from GSI and not, in particular, as an independent derivation from self-compatibility. Pandey interpreted the evolution of SSI from GSI as one aspect of a general evolutionary trend promoting accelerated development of microgametophytes.

DE NETTANCOURT (1977, pp. 22–23) further elaborated this theme, hypothesizing the origin of GSI in the ancestor of all angiosperms, followed by the evolution of SSI directly from GSI, heteromorphic SI from homomorphic SI, and dioecy from heteromorphic SI. These developments were characterized as a continuous orthogenetic progression toward outbreeding.

The molecular basis of self-incompatibility: At the nucleotide level, loci controlling SSI in Brassica (NASRALLAH *et al.* 1985) bear no resemblance to loci controlling GSI in the Solanaceae (ANDERSON *et al.* 1986). This lack of homology challenges the hypothesis of monophyly of the two major forms of homomorphic self-incompatibility. In both the brassicaceous and solanaceous systems, the *S* locus belongs to an apparently ancient multigene family. Loci showing homology to the *S* locus of the Brassica system appear in self-compatible as well as self-incompatible species within the family (NASRALLAH *et al.* 1988) and also in maize (WALKER and ZHANG 1990; ZHANG and WALKER 1993), carrot (VAN ENGELEN *et al.* 1993), tomato (MARTIN *et al.* 1993) and petunia (MU *et al.* 1994). MCCLURE *et al.* (1989) discovered that the *S* locus of the solanaceous GSI system shows homology to a family of ribonuclease (RNase) genes expressed in fungi. This family also includes a glycoprotein that serves as both a coat protein and a secretory product of an RNA virus (SCHNEIDER *et al.* 1993). Solanaceous *S* glycoproteins (*S*-RNases) exhibit *in vitro* RNase activity and cause *in vivo* degradation of ribosomal RNA (rRNA) in incompatible but not compatible pollen tubes (MCCLURE *et al.* 1989, 1990). GSI in the Japanese pear (Rosaceae) also appears to derive from *S*-RNase activity (SASSA *et al.* 1992, 1993). Homologous RNases that do not cosegregate with the physiological expression of self-incompatibility occur in the Solanaceae (JOST *et al.* 1991; LEE *et al.* 1992; LÖFFLER *et al.* 1993), the Brassicaceae (TAYLOR and GREEN 1991; TAYLOR *et al.* 1993; WALKER 1993) and the Cucurbitaceae (IDE *et al.* 1991).

While the original functions of the two multigene families are unknown, Dickinson and colleagues (HODGKIN *et al.* 1988; DICKINSON 1994) have drawn analogies between plant-pathogen interactions and pollen-stigma interactions in the Brassica SSI system, and LEE *et al.* (1992) have speculated that *S*-RNases may have been derived from proteins that functioned in defense against pathogens. Indeed, a gene (*Pto*) in tomato that has been shown to confer resistance to *Pseudomonas syringae,* the cause of bacterial speck disease, shows homology to the catalytic kinase domain of the Brassica *SRK* gene (MARTIN *et al.* 1993).

The most parsimonious interpretation of the presence in fungi of homologues to the solanaceous *S* locus and the presence in plant families for which SSI is atypical of homologues to the Brassica *S* locus is that both multigene families predate the function of self-incompatibility. Rather than representing vestiges of a once-functional *S*-locus, homologues that do not cosegregate with self-incompatibility may well descend from lineages that have no history of expression of self-incompatibility.

Balancing selection: Loci of the major histocompatibility complex (MHC) in vertebrates and *S* loci in flowering plants provide some of the best examples of balancing selection. Effective cell-mediated immune responses against highly diverse foreign antigens require highly diverse MHC loci. Self-incompatibility restricts outcrossing as well as self-fertilization, with the consequence that less frequent *S* alleles hold a selective advantage in mating opportunities. Balancing selection is expected to promote the long-term maintenance of polymorphisms at *S* loci as well as within the MHC.

Trans-specific evolution (ARDEN and KLEIN 1982) describes greater genetic similarity in interspecific than intraspecific comparisons, suggesting divergence of genes before divergence of species. Observations of this pattern for both MHC loci (FIGUEROA *et al.* 1988; LAWLOR *et al.* 1988) and *S* loci (IOERGER *et al.* 1990; DWYER *et al.* 1991) support the hypothesis of balancing selection. Lineages of extant MHC alleles appear to have diverged on the order of 30–40 mya (SATTA *et al.* 1991). Lineages of extant *S*-RNases diverged before diversification within the Solanaceae (27–30 mya; IOERGER *et al.* 1990), and extant brassicaceous *S*-alleles before the separation of *Brassica oleracea* and *B. campestris* (DWYER *et al.* 1991).

Does hypermutability contribute significantly to diversity? A question raised soon after the observation of extraordinarily high levels of genetic diversity at the MHC concerns the extent to which hypermutation contributes to this diversity. Molecular clock calibrations based on comparisons that minimize the effect of trans-specific evolution indicate that the rate of synonymous substitution at MHC loci does not in fact differ from other protein-encoding loci (SATTA *et al.* 1993).

Relatively little work has addressed whether new *S*-allele specificities arise at accelerated rates. Extensive studies of spontaneous and radiation-induced mutations succeeded in identifying an array of loss-of-function variants, but failed to produce a single fully functional *S* allele that expressed a novel specificity (reviewed by LEWIS and CROWE 1954). Although the minimum rate such studies could have detected (10^{-8})

is rather high on an evolutionary scale, FISHER (1961) argued that the extreme genetic diversity typical of *S* loci suggests an even higher rate of origin. FISHER proposed a model under which new *S* alleles may arise without causing self-compatibility, thereby escaping detection in laboratory experiments. DE NETTANCOURT *et al.* (1971) reported the repeated generation of a particular new *S* allele through inbreeding, but the structural and functional relationship of the novel allele to existing *S* alleles has not been determined.

To date, the only study (TRICK and HEIZMANN 1992) that addressed the rate of evolution of *S* alleles imposed the assumption that diversity within the SLG/SRK system of sporophytic self-incompatibility arose wholly within the Brassicaceae. Under this assumption, the rate of nonsynonymous substitution for brassicaceous *S* alleles was found to exceed even the rate for pseudogenes by two- to threefold, with even higher rates indicated for the solanaceous system. Although SLG/SRK-mediated sporophytic self-incompatibility has not been reported outside the Brassicaceae, the possibility that this system of SSI predates divergence within the family has not been explored. A rate of substitution typical of non-*S* sequences might be assigned to the *S* locus and the time of origin of self-incompatibility estimated under molecular clock assumptions; however, such an approach could not address whether *S* alleles do in fact evolve at a distinct rate. The present analysis addresses the problem of distinguishing between ancient origin and hypermutability as factors contributing to the extraordinarily high levels of diversity typical of *S* loci.

Phylogenetic analysis of the *SLG/SRK/SLR* multigene family: This phylogenetic study of brassicaceous *S* alleles and their homologues is designed to address the age of this system of SSI and the rate of amino acid substitution at the *S* locus while imposing minimal assumptions regarding either aspect. Information on the functional roles of various genes permits the assignment of *function-dependent* divergence rates. Rate estimates obtained here provide no evidence that *S* alleles undergo nonsynonymous substitution at higher rates than other members of the multigene family.

Estimates for divergence rates and node ages were obtained relative to the time of separation of *B. oleracea* and *B. campestris*. Variation at a locus (SLR_1) that does not cosegregate with self-incompatibility does not exhibit the trans-specific pattern, an observation consistent with the absence of balancing selection. By equating the times of coalescence of species and sequences at this locus, I estimated that the time since the change of function which signaled the origin of SSI exceeds the species divergence time by a factor of four or five. This estimate of the age of sporophytic self-incompatibility (~50 million years) indicates a much more recent origin than expected under the classical view.

LEAST-SQUARES ESTIMATION

RAO (1973, Chapter 4) has provided a lucid exposition of the least-squares method. My objective here is to describe the application of this approach to the estimation of evolutionary divergence times and rates and the testing of phylogenetic hypotheses.

Statistical basis

Given a set of observations ($\mathbf{Y}$) with Gaussian distribution, a set of parameter estimates ($\mathbf{b}$) is sought such that the observations correspond in expectation to linear combinations of the estimates:

$$E[\mathbf{Y}] = \mathbf{Mb}. \quad (1)$$

In the present context, $\mathbf{Y}$ represents the $\binom{n}{2}$-dimensional vector of all pairwise distances among n sequences, $\mathbf{b}$ the $(2n - 3)$-dimensional vector of lengths of branches in an unrooted, bifurcating tree and $\mathbf{M}$ the incidence matrix in which 1 indicates the presence and 0 the absence of a given branch in the path connecting a given pair of sequences. Matrix $\mathbf{M}$ fully specifies the topology.

Ordinary least-squares estimation: Under the ordinary least-squares (OLS) approach, the dispersion (variance-covariance) matrix of $\mathbf{Y}$ is assumed to be diagonal with equal variances:

$$\mathrm{Var}[\mathbf{Y}] = \sigma^2\mathbf{I}. \quad (2)$$

This assignment assumes identical variances (σ^2) for all observations and no correlations among observations. Neither assumption holds for phylogenetic distances, for which the variances of pairwise distances depend on the magnitudes of the distances and covariances reflect common ancestry.

The best choice for $\mathbf{b}$ under the least-squares criterion is the vector that minimizes the residual sum of squares (R_0^2):

$$R_0^2 = (\mathbf{Y} - \mathbf{Mb})^{\mathrm{T}}(\mathbf{Y} - \mathbf{Mb}), \quad (3)$$

in which T denotes the transpose. Under the OLS approach, the vector of parameter estimates,

$$\mathbf{b} = (\mathbf{M}^{\mathrm{T}}\mathbf{M})^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{Y}, \quad (4)$$

has a Gaussian distribution with dispersion matrix

$$\mathrm{Var}[\mathbf{b}] = (\mathbf{M}^{\mathrm{T}}\mathbf{M})^{-1}\mathbf{M}^{\mathrm{T}}\mathrm{Var}[\mathbf{Y}][(\mathbf{M}^{\mathrm{T}}\mathbf{M})^{-1}\mathbf{M}^{\mathrm{T}}]^{\mathrm{T}}$$
$$= (\mathbf{M}^{\mathrm{T}}\mathbf{M})^{-1}\sigma^2 \quad (5)$$

[using (2)]. The residual sum of squares R_0^2 has a chi-square distribution and is proportional to σ^2, the unknown common variance of the observations:

$$R_0^2 \sim \sigma^2\chi^2_{[r]}, \quad (6)$$

in which r is the rank of the matrix $\mathbf{M}$. In the context of phylogenetic analysis, r corresponds to the difference

between the number of pairwise distances observed and the number of branch lengths estimated [$r = \binom{n}{2} - (2n - 3) = (n - 2)(n - 3)/2$]. Because σ^2, the common variance of the observations, is unknown, (6) does not directly provide a test of overall fit. However, CAVALLI-SFORZA and EDWARDS (1967) suggested that the *relative* magnitudes of the residual sums of squares associated with alternative topologies can be used as a basis for comparing fit. Tests of other hypotheses require elimination of σ^2, through division of test statistics by R_0^2.

Generalized least-squares estimation: The generalized least-squares (GLS) approach accommodates general dispersion matrices:

$$\mathrm{Var}[\mathbf{Y}] = \Sigma. \tag{7}$$

An upper triangular, square-root matrix $\mathbf{R}$ exists for all nonsingular Σ:

$$\mathbf{R}^T\mathbf{R} = \Sigma.$$

Under the change of variables

$$\mathbf{Z} = (\mathbf{R}^T)^{-1}\mathbf{Y}, \tag{8}$$

the GLS reduces to the OLS framework:

$$E[\mathbf{Z}] = (\mathbf{R}^T)^{-1}\mathbf{M}\mathbf{b} \tag{9}$$

$$\mathrm{Var}[\mathbf{Z}] = (\mathbf{R}^T)^{-1}\mathrm{Var}[\mathbf{Y}][(\mathbf{R}^T)^{-1}]^T = \mathbf{I} \tag{10}$$

[*cf.* (1) and (2)]. The GLS estimate $\mathbf{b}$ and its dispersion matrix Var[$\mathbf{b}$] follow directly from (4) and (5) with $\mathbf{M}$ replaced by $(\mathbf{R}^T)^{-1}\mathbf{M}$ and Var[$\mathbf{Y}$] by $\Sigma = \mathbf{R}^T\mathbf{R}$:

$$\mathbf{b} = (\mathbf{M}^T\Sigma^{-1}\mathbf{M})^{-1}\mathbf{M}^T\Sigma^{-1}\mathbf{Y} \tag{11}$$

$$\mathrm{Var}[\mathbf{b}] = (\mathbf{M}^T\Sigma^{-1}\mathbf{M})^{-1}. \tag{12}$$

As under the OLS approach, the GLS estimate $\mathbf{b}$ minimizes the residual sum of squares

$$R_0^2 = [\mathbf{Z} - (\mathbf{R}^T)^{-1}\mathbf{M}\mathbf{b}]^T[\mathbf{Z} - (\mathbf{R}^T)^{-1}\mathbf{M}\mathbf{b}],$$

which corresponds in the original scale to

$$R_0^2 = (\mathbf{Y} - \mathbf{M}\mathbf{b})^T\Sigma^{-1}(\mathbf{Y} - \mathbf{M}\mathbf{b}). \tag{13}$$

This expression indicates that in the original scale the residual sum of squares is weighted by the inverse of the dispersion matrix of the observations. FITCH and MARGOLIASH's (1967) "standard deviation" corresponds to weighting by the reciprocal of the squared estimates, or replacement of Σ in (13) by a diagonal matrix with nonzero elements given by the squared values of $\mathbf{b}$ (Diag[$\mathbf{b} \circ \mathbf{b}$]).

In the context of phylogenetic analysis, errors in the estimation of smaller distances (which have smaller variances) receive more weight than those of larger distances. Correlations among distances cause correlations among error estimates. Accordingly, the off-diagonal elements of Σ^{-1}, representing covariances among distances, also affect the residual sum of squares, reducing the contribution of positively correlated errors and increasing the contribution of negatively correlated errors.

As the GLS estimate $\mathbf{b}$ (11) and its dispersion matrix (12) do not involve the unknown common variance σ^2 of the OLS estimate, they can be used directly for hypothesis testing and parameter estimation. In particular, the residual sum of squares R_0^2 (13) provides a test of overall fit [see (6) with σ^2 set to unity].

Application to phylogenetic analysis

Distance estimation: Pairwise distances, together with their variances and covariances, may be obtained directly from sequence information. Let p_{ij} correspond to the proportion among sites shared by sequences i and j that differ between members of the pair. The estimated number of substitutions per site, estimated from p_{ij} by correcting for multiple substitutions at a site (see NEI 1987, Chapter 4), provides a measure of distance between pairs of sequences. The proportion of differences per site has a binomal variance:

$$\sigma_{ij}^2 = \frac{p_{ij}(1 - p_{ij})}{n_{ij}}, \tag{14}$$

in which n_{ij} represents the number of sites shared by sequences i and j (KIMURA and OHTA 1972). Covariances between distances reflect common ancestry (NEI *et al.* 1985). For two sequence pairs considered simultaneously ($\{i, j\}$ and $\{k, l\}$), let p_{ijkl} represent the proportion among sites shared by all four sequences that differ between members of both pairs. The covariance between the proportions of differences for the two pairs is

$$\sigma_{ijkl} = (p_{ijkl} - p_{ij}^* p_{kl}^*)\frac{n_{ijkl}}{n_{ij}n_{kl}}, \tag{15}$$

in which n_{ijkl} denotes the number of sites shared by all four sequences, with p_{ij}^* the proportion of those sites that differ between i and j and p_{kl}^* the corresponding proportion for k and l (BULMER 1991).

In the present study, amino acid sequences were aligned pairwise, by comparison to one sequence using the Wilbur-Lipman algorithm, and simultaneously, using CLUSTAL (DNASTAR 1992). Discrepancies were resolved individually, using available nucleotide sequences. A listing of the final alignment will be provided on request. Distances, variances and covariances were computed, and a Poisson correction applied (see BULMER 1991).

Technical difficulties associated with inversion of the dispersion matrix (Σ) may restrict the implementation of the GLS approach (see RZHETSKY and NEI 1992). If two pairs of sequences ($\{i, j\}$ and $\{k, l\}$) show agreements and differences in identical positions, then the

corresponding variances and covariances will also be identical, resulting in a singular dispersion matrix. Such a situation may occur if, for example, sequences i and k, although not identical, are sufficiently similar so that they differ from a third sequence $j = l$ at the same sites. Recognition of more kinds of differences (*e.g.*, transitions and transversions) reduces the chances of singularity.

For sequence pairs associated with identical or very similar arrays of agreements and differences, elimination of any sequence in the set removes the corresponding linearly dependent rows of Σ. The loss of information this procedure entails is expected to be minimal, as the discarded sequences show close similarity to other members of the set. In the present analysis, a Gram-Schmidt orthogonalization procedure (see HEALY 1986, Chapter 6) was applied to the dispersion matrices to identify linearly dependent rows, and the corresponding sequences eliminated.

Estimation of times and rates of divergence: In estimating lineage ages and rates of amino acid substitution, I first applied a multivariate relative rates test to examine homogeneity of evolutionary rates within lineages. A function-dependent molecular clock was then imposed, with divergence along branches associated both with similar functions and with rate homogeneity constrained to proceed at identical rates. Finally, the entire array of rates and times was estimated relative to the age of a particular node, a rough estimate for which is available from external information.

Relative rates test: Lineages within multigene families may well evolve at different rates, reflecting the diversity of constraints imposed by the diversity of functions performed. Whereas the existence of a *global* molecular clock seems unlikely on a *priori* grounds, sequences may evolve according to a *function-dependent* molecular clock.

Tests of relative rates in wide usage (*e.g.*, WU and LI 1985; MUSE and WEIR 1992) entail comparing lengths of branches linking two sequences to their first common node. Examination of relative rates among several sequences requires multiple comparisons, with appropriate adjustments of significance level (see LI and BOUSQUET 1992). Within the least-squares framework, all comparisons can be made in a single test. Any hypothesis concerning linear combinations of branch lengths can be represented by

$$(\mathbf{Pb} - \mathbf{v}) = 0, \tag{16}$$

in which $\mathbf{b}$ denotes the vector of branch length estimates, $\mathbf{P}$ a matrix the rows of which describe linear combinations of branch lengths and $\mathbf{v}$ the vector of values to which the combinations of branch lengths are to be compared. The hypothesis in (16), involving l linearly independent combinations of branch lengths, may be tested using

$$(\mathbf{Pb} - \mathbf{v})^{\mathrm{T}}\mathrm{Var}[\mathbf{Pb}]^{-1}(\mathbf{Pb} - \mathbf{v}) \sim \chi^2_{[l]}, \tag{17}$$

for $\mathrm{Var}[\mathbf{Pb}] = \mathbf{P}\mathrm{Var}[\mathbf{b}]\mathbf{P}^{\mathrm{T}}$ (see RAO 1973, p. 238).

To test the equality of evolutionary rates within a subset of $l + 1$ sequences, the sum of the lengths of branches joining each sequence to the first common ancestor of the subset may be compared with the mean over the subset. A row of matrix $\mathbf{P}$ describes the difference between these two quantities and all elements of $\mathbf{v}$ in (16) are set to 0. Because the mean branch length to the common node is computed from the branch lengths joining all of the sequences to this node, this test involves l independent comparisons.

Log approximation of rates and times: Biological hypotheses concerning function and rate associate products of divergence rate and time with branch lengths:

$$\mathbf{b} = \mathbf{Br}, \tag{18}$$

in which $\mathbf{r}$ denotes a vector of products of rates (a) and times (T), and $\mathbf{B}$ describes the assignment of rate-time products to the branches. Equation 18 describes only a change in naming convention, and replacement of $\mathbf{M}$ in (11) and (12) by $\mathbf{MB}$ provides GLS estimates of $\mathbf{r}$ and its dispersion matrix $\mathbf{V}_\mathbf{r}$. A log transformation,

$$\mathbf{l} = \log\mathbf{r}, \tag{19}$$

provides estimates of the form $\log a + \log T$, with an approximate dispersion matrix (obtained by δ-approximation),

$$\mathbf{V}_\mathbf{l} = \mathrm{Diag}[\tfrac{1}{\mathbf{r}}]\mathbf{V}_\mathbf{r}\mathrm{Diag}[\tfrac{1}{\mathbf{r}}], \tag{20}$$

in which $\mathrm{Diag}[1/\mathbf{r}]$ denotes a diagonal matrix with nonzero elements given by the reciprocals of the elements of $\mathbf{r}$.

External information on any of the rates and times may be incorporated and the estimates of logs of the remaining parameters obtained by the usual GLS procedure by treating the elements of $\mathbf{l}$ as observations. For $\mathbf{c}$ the vector of externally determined values, $\mathbf{l}$ corresponds to

$$\mathbf{l} = \mathbf{Nq} + \mathbf{c}, \tag{21}$$

with $\mathbf{q}$ the vector of logs of rates and logs of times. Replacement of $\mathbf{Y}$ by $\mathbf{l} - \mathbf{c}$, $\mathbf{M}$ by $\mathbf{N}$ and Σ by $\mathbf{V}_\mathbf{l}$ in (11) and (12) provides GLS estimates of $\mathbf{q}$ and its dispersion matrix $\mathbf{V}_\mathbf{q}$. Returning to base 10 produces estimates of the parameters ($\hat{\mathbf{p}}$) and an approximate dispersion matrix ($\mathbf{V}_{\hat{\mathbf{p}}}$):

$$\hat{\mathbf{p}} = e^{\mathbf{q}} \tag{22}$$

$$\mathbf{V}_{\hat{\mathbf{p}}} = \mathrm{Diag}[e^{\mathbf{q}}]\mathbf{V}_\mathbf{q}\mathrm{Diag}[e^{\mathbf{q}}]. \tag{23}$$

Iterative least-squares minimization: Whereas the log transformation approach provides closed form expressions for the parameter and error estimates [(22) and (23)], the approximations involved introduce inaccu-

racies of unknown magnitude. HASEGAWA *et al.* (1985) applied a numerical least-squares procedure to estimate divergence times and rates of transition and transversion substitutions among hominoid mitochondrial DNA sequences. This approach permits assignment of arbitrary functions of parameters to branch lengths. The residual sum of squares is determined as

$$R_0^2 = (\mathbf{Y} - \mathbf{E})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{E}), \qquad (24)$$

in which $\mathbf{E}$ denotes a vector of expected pairwise distances, general functions of parameters determined by a given biological hypothesis. A starting value for R_0^2 is computed for an initial guess for $\mathbf{p}$, the vector of parameters to be estimated, and the least-squares estimate $\hat{\mathbf{p}}$ obtained through progressive reduction of R_0^2 by a numerical multivariate minimization procedure (see PRESS *et al.* 1986). HASEGAWA *et al.* (1985) provided expressions for the dispersion matrix of $\hat{\mathbf{p}}$:

$$\mathbf{V}_{\hat{\mathbf{p}}} = \mathbf{B}^{-1} \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^{\mathrm{T}} \mathbf{B}^{-1}, \qquad (25)$$

in which

$$\mathbf{A} = \left. \frac{\partial^2 R_0^2}{\partial \mathbf{E} \partial \mathbf{Y}^{\mathrm{T}}} \right|_{\hat{\mathbf{p}}} \qquad (26)$$

$$\mathbf{B} = \left. \frac{\partial^2 R_0^2}{\partial \mathbf{E} \partial \mathbf{E}^{\mathrm{T}}} \right|_{\hat{\mathbf{p}}}. \qquad (27)$$

In the present analysis, the log transformation approach provided the initial guess for the parameter values (22); in every case, no further minimization of the residual sum of squares was necessary, indicating that the approximations underlying the closed form solution did not introduce large errors. The standard errors presented here derive from the expressions (25) given by HASEGAWA *et al.* (1985).

PHYLOGENETIC ANALYSIS

Members of the *SLG*/*SRK*/*SLR* family: Soluble *S*-locus glycoproteins (designated SLGs by LALONDE *et al.* 1989) appear abundantly in stigma extracts in parallel with the onset of the expression of self-incompatibility and cosegregate with pollen specificity (see SIMS 1993). Correspondence between these glycoproteins and the *SLG* gene has been confirmed by direct protein sequencing (TAKAYAMA *et al.* 1987). NASRALLAH *et al.* (1991) distinguished between class I and class II SLGs on the basis of amino acid sequence. Class I includes high activity proteins that express strong self-incompatibility reactions and codominance in determination of pollen specificity and class II, weaker, recessive forms.

Recognition of pollen specificity appears to be transduced through an association between extracellular SLGs and a membrane-bound *S*-receptor kinase (SRK). *SRK* genes encode putative serine/threonine protein kinases comprising three domains: an extracellular receptor closely resembling the corresponding SLG, a short, hydrophobic transmembrane region and an intracellular catalytic domain (STEIN *et al.* 1991). The tightly linked *SLG* and *SRK* genes, possibly together with other elements, constitute the classical *S* locus. NASRALLAH and NASRALLAH (1993) suggest that *S*-alleles are more properly designated *S* haplotypes.

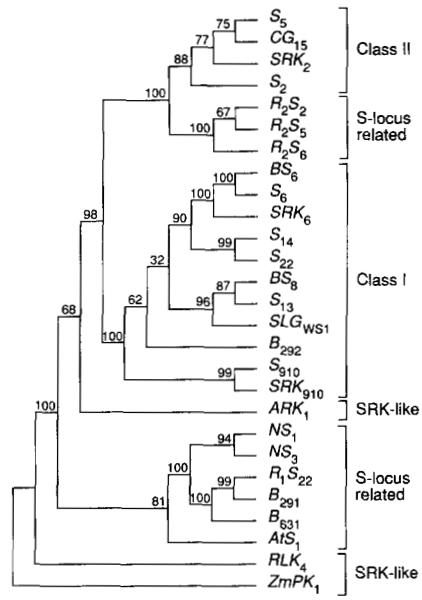

FIGURE 1.—Neighbor-joining phylogeny for 27 *SLG* and related sequences with the maize sequence ($ZmPK_1$) as the outgroup. Bootstrap values for 1000 resamplings appear next to the corresponding nodes.

S-locus-related (*SLR*) genes show high sequence homology to *SLG* genes but do not cosegregate with self-incompatibility. Two groups of sequences, apparently derived from distinct genetic loci, have been characterized: SLR_1 and SLR_2 (LALONDE *et al.* 1989; BOYES *et al.* 1991) or *SRA* and *SRB* (HINATA *et al.* 1993). The loci show tight linkage, with SLR_2 closely resembling the pollen-recessive class II *SLG* genes (BOYES *et al.* 1991).

My estimates of divergence rates and times derive from a phylogenetic analysis of 29 amino acid sequences, including homologues in *Arabidopsis thaliana*, *Lycopersicon esculentum* and *Zea mays* (Table 1). Within Brassica, the sequences were derived from *B. oleracea* and *B. campestris* and also from *B. napus*, an amphidiploid derived from hybridization between *B. oleracea* and *B. campestris*. To address whether ARK_1 from self-compatible Arabidopsis may have descended from an *SRK* gene that expressed self-incompatibility, I also examined the kinase domains of the *SRK* and *SRK*-like genes.

Topology and tests of relative rates: Figure 1 presents the topology, determined by the neighbor-joining (NJ) algorithm (SAITOU and NEI 1987) with Poisson correction for multiple hits as implemented in the MEGA package (KUMAR *et al.* 1993), for 27 sequences,

TABLE 1

***SLG/SRK* family and homologues**

Source	*SLG*	*SRK*	*SLR*	*SRK*-like
Brassica oleracea	S_2,[1] S_5,[2] S_6,[3] S_{13},[6] S_{14},[3] S_{22},[4] B_{292},[5] CG_{15}[9]	SRK_2,[1] SRK_6[1]	R_1S_{22},[4] B_{291},[5] B_{631},[7] R_2S_2,[8] R_2S_5,[10] R_2S_6[8]	
campestris napus	BS_6,[11] BS_8,[6] S_{910}[12] SLG_{WS1}[15]	SRK_{910}[12]	NS_1,[13] NS_3[14]	
Arabidopsis thaliana			AtS_1[16]	ARK_1,[17] RLK_4,[18] TMK_1[19]
Lycopersicon esculentum				*Pto*[20]
Zea mays				$ZmPK_1$[21]

[1] CHEN and NASRALLAH 1990, [2] SCUTT and CROY 1992, [3] NASRALLAH *et al.* 1987, [4] LALONDE *et al.* 1989, [5] TRICK and FLAVELL 1989, [6] DWYER *et al.* 1991, [7] TRICK 1990, [8] BOYES *et al.* 1991, [9] GAUDE *et al.* 1993, [10] SCUTT *et al.* 1990, [11] TAKAYAMA *et al.* 1987, [12] GORING *et al.* 1992, [13] ISOGAI *et al.* 1991, [14] WATANABE *et al.* 1992, [15] ROBERT *et al.* 1994, [16] DWYER *et al.* 1992, [17] TOBIAS *et al.* 1992, [18] WALKER 1993, [19] CHANG *et al.* 1992, [20] MARTIN *et al.* 1993, [21] WALKER and ZHANG 1990.

including *SLG* genes, *SLG*-like genes, the *SLG*-like extracellular domain of *SRK* and *SRK*-like genes and the *S*-locus-related genes SLR_1 and SLR_2. The extracellular domain of TMK_1 from Arabidopsis shows no homology to *SLG* and is unknown for *Pto* from tomato. The indicated bootstrap values suggest four major groups: the outgroup sequences from Zea ($ZmPK_1$) and Arabidopsis (RLK_4), *S*-locus-related sequences (SLR_1 and ARK_1), pollen-recessive (class II) *S* alleles and SLR_2 and pollen-dominant (class I) *S* alleles. The relationship of ARK_1 to the *S* allele and non-*S* clusters is equivocal.

The clustering of the pollen-recessive *S* alleles with sequences from SLR_2, which does not cosegregate with SI, suggests that the SLR_2 locus was derived secondarily by a process such as gene conversion from pollen-recessive *S*-allele sequences or the converse. Both possibilities were considered in the assignment of evolutionary rates.

To permit inversion of the dispersion matrix as required under the GLS approach, seven sequences were removed from consideration: three SLR_1 sequences (B_{291}, B_{631} and NS_1), three SLR_2 sequences (CG_{15}, R_2S_5 and R_2S_6) and one pollen-dominant *S*-allele (BS_6 from *B. campestris*, which differs from the *B. oleracea* S_6 gene by four amino acid substitutions among 405 residues).

Figure 2 presents the NJ topology and bootstrap values for the remaining 20 sequences. Asterisks indicate branches for which GLS estimates of length did not differ significantly from 0. Both the close relationship between the SLR_2 gene and the class II *S* alleles and the unresolved status of ARK_1 were confirmed. Figure 3 presents the same phylogeny with branch lengths drawn proportional to their GLS estimates and non-significant branches eliminated.

To explore whether ARK_1 may have diverged before the AtS_1/SLR_1 cluster, I determined GLS branch lengths for that topology, obtaining a negative value for the length of the branch separating the *S* alleles and the SLR_1 cluster from the remaining sequences. Although this result suggests error in the topology, it cannot be unequivocally rejected as the branch length estimate does not differ significantly from 0 and the residual sum of squares indicates an acceptable fit ($X^2_{[153]} = 170$).

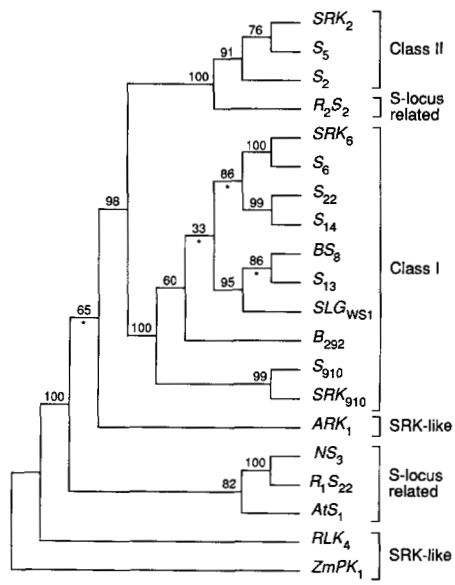

FIGURE 2.—Neighbor-joining phylogeny for the 20 *SLG* and related sequences remaining after reduction to ensure invertibility of the dispersion matrix of the pairwise distances. GLS branch lengths and bootstrap values for 1000 resamplings are shown. Asterisks indicate the four branch lengths that do not differ significantly from zero. From the 190 pairwise distances among the 20 sequences, 37 branch lengths were estimated, leaving 153 degrees of freedom. The value obtained for the residual sum of squares (168) did not exceed the approximate 5% significance bound $\{[1.645 + \sqrt{2(153) - 1}]^2/2 = 183\}$, indicating an acceptable fit.

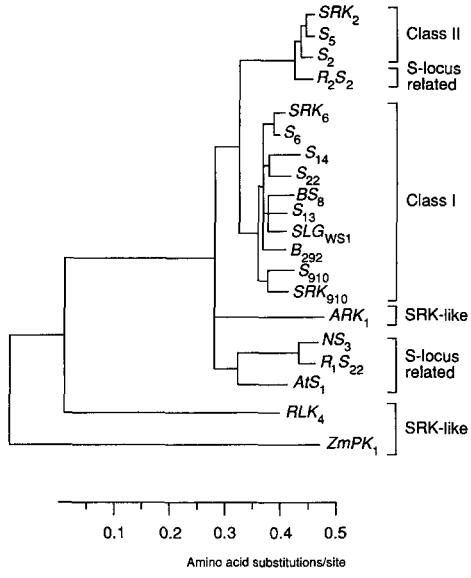

FIGURE 3.—Neighbor-joining phylogeny with branch lengths drawn proportional to the numbers of substitutions per site. The four branches with estimated lengths not significantly different from 0 (asterisks in Figure 2) were eliminated and their lengths added to the lengths of the immediate descendant branches.

Evolutionary divergence rates may vary according to function. To test whether a molecular clock holds among the class I (pollen dominant) *S* alleles, I applied the GLS relative rates test described in (17), obtaining a very highly significant value ($X^2_{[9]} = 30.3$). Inspection of the branch lengths suggested that two sequences (S_6 and S_{13}) have evolved more slowly, and comparison among the remaining pollen-dominant *S* alleles reduced the significance to the 0.025 level ($X^2_{[7]} = 16.4$). Similarly, a rate homogeneity test among the genes that do not cosegregate with self-incompatibility (SLR_1, SLR_2, AtS_1, ARK_1 and RLK_4) was significant ($X^2_{[5]} = 13.9$), but returned a nonsignificant value ($X^2_{[4]} = 2.9$) after the slowly evolving AtS_1 sequence was removed from the comparison. The class II (pollen recessive) *S* alleles showed nonsignificant rate heterogeneity ($X^2_{[2]} = 0.3$).

Biological hypotheses: Function-specific divergence rates were assigned to branches in accordance with biological hypotheses concerning the functional role of ancestral lineages.

Assignments of divergence rates and times: The relative rate comparisons support the assignment of a common rate for the non-*S* SLR_1 and SLR_2 sequences (*a*), a pollen-recessive *S*-allele rate (*b*) and a pollen-dominant *S*-allele rate (*c*), with possibly different rates for AtS_1 (*d*), S_6 (*e*) and S_{13} (*f*). In addition, ARK_1 from self-compatible Arabidopsis was assigned a separate rate (*g*) because ambiguity in its position in the topology does not permit exclusion of the possibility that it is derived from a functional *SRK* sequence. Table 2 summarizes the assignments of divergence rates.

TABLE 2

Function-specific divergence rates

Parameter	Sequence
a	SLR_1 and SLR_2 (non-*S*)
b	Class II *S* alleles
c	Class I *S* alleles
d	AtS_1 (Arabidopsis *S*-locus related)
e	S_6 (*Brassica oleracea* Class I *S* allele)
f	S_{13} (*B. oleracea* Class I *S* allele)
g	ARK_1 (Arabidopsis *SRK*-like)
x	Ancestral non-*S* (=*a*, *d*, or *g*)
y	Class II/SLR_2 progenitor (=*a* or *b*)
z	Ancestral *S* allele (=*b* or *c*)

Figure 4 shows the divergence times and rates assigned to nodes and branches under this biological hypothesis. $ZmPK_1$, RLK_4 and the SLR_1 sequences are assumed to represent the ancestral multigene family, having no history of self-incompatibility function. Rate *x*, corresponding to the ancestral rate prior to the origin of self-incompatibility, was set equal to one of the non-*S* rates (*a*, *d* or *g*) in the scenarios considered. Rate *y*, corresponding to the lineage immediately ancestral to SLR_2 and the class II *S* alleles, was assigned as *a* (implying the derivation of the pollen-recessive *S* alleles from SLR_2) or as *b* (implying the converse). The original *S*-allele lineage was assumed to have evolved at rate *z*, assigned as either the pollen-recessive (*b*) or the pollen-dominant (*c*) rate. The possible combinations of assignments of ancestral rates (see Table 2: 3 for *x*, 2 for *y* and 2 for *z*) determine a maximum of twelve scenarios. Because the hypothesis that the pollen-recessive lineage represents the original *S* alleles ($z = b$) conflicts with the hypothesis that the pollen-recessive alleles derive from SLR_2 ($y = a$), the latter assignment was considered only under the assignment of the pollen-dominant rate to the original *S*-allele lineage ($z = c$). This restriction reduces the number of plausible scenarios to nine.

Origin of self-incompatibility: Two placements of the origin of self-incompatibility were considered (origins 1 and 2 in Figure 4). Under origin 1, the time since the origin of self-incompatibility (T_S) lies between T_2 and T_5, with some period of expression of self-incompatibility in the lineage leading to ARK_1. Because the overall rate of divergence of ARK_1 would then reflect *S* as well as non-*S* function, the assignment of the ARK_1 rate as the ancestral non-*S* rate ($x = g$) was not considered under origin 1. Under origin 2, T_S falls between T_5 and T_6, with the divergence of ARK_1 predating the expression of self-incompatibility. All nine possible scenarios, corresponding to assignments to the ancestral non-*S* lineages (*x*), the immediate common ancestor of SLR_2

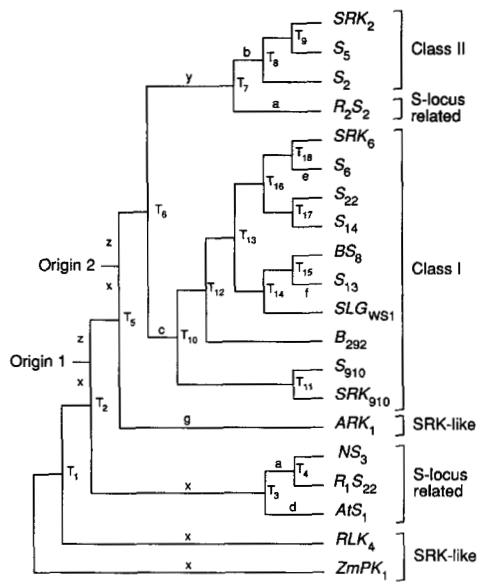

FIGURE 4.—Function-dependent rates of substitution. Rate x is associated with the ancestral non-S sequences and y, the lineage from which SLR_2 and the pollen-recessive S alleles descend. Under each placement of the origin of self-incompatibility (origin 1 and origin 2), substitutions along the branch segment preceding the point of origin occur at rate x and after rate z.

and the pollen-recessive lineages (y), and the original S alleles (z), were considered under origin 2.

Under both origin hypotheses, the time since the origin of self-incompatibility (T_S) depends on the joint assignments of x, y and z. Because the branch containing the origin of self-incompatibility has an expected length of $xT_U - (x - z)T_S - zT_L$ (for T_U and T_L the ages of the nodes flanking the origin), estimation of T_S requires that the ancestral divergence rate before the expression of self-incompatibility differ significantly from the ancestral S-allele rate ($x \neq z$). Tests using the full dispersion matrix for the array of estimates indicated that x and z differed significantly in only two of the scenarios considered ($x = a$, $y = a$ or b, $z = c$); in these cases, an extremely large error variance was associated with the estimate of T_S, which in one case lay outside the interval bounded by the ages of the flanking nodes. Because the data do not permit direct estimation of T_S and because the estimates for T_U and T_L did not differ significantly, the time since the origin of self-incompatibility was estimated as the average of the ages of the flanking nodes [$(T_U + T_L)/2$].

Anchoring the phylogeny: Branch lengths correspond to products of divergence rates and times, with the separate estimation of rates and times entailing anchoring the phylogeny on at least one externally specified parameter value. All rates and times were estimated relative to the coalescence of the SLR_1 sequences at time T_4 (Figure 4).

TABLE 3

Origin- and scenario-independent divergence rates and times

Substitution rates[a]	
SLR_1/SLR_2 (a)	3.5 ± 0.5
Class II S alleles (b)	3.1 ± 0.9
AtS_1 (d)	2.2 ± 0.5
Divergence times[b]	
SLR_1/AtS_1 (T_3)	39.4 ± 6.5
SLR_2/S_2 (T_7)	10.0 ± 2.5
S_2/SRK_2 (T_8)	6.4 ± 2.0
SRK_2/S_5 (T_9)	4.3 ± 1.6

Values are means ± SE.
[a] 10^{-9} substitutions/site/year.
[b] 10^6 years.

Inspection of the SLR_1 sequences indicated that most of the variation occurs between *B. oleracea* and *B. campestris* rather than within species. Of the 27 sequences analyzed, five (R_1S_{22}, B_{291}, B_{631}, NS_1 and NS_3) have been designated SLR_1 on the basis of sequence similarity and the presence of characteristic insertion/deletion regions relative to S alleles (see Dickinson 1990). Among the 438 sites aligned across the five sequences, variation occurred at 53. Of the 29 sites that were informative in the sense used in parsimony analysis, only 4 suggested a grouping incongruent with the species designation. This highly significant value supports the view that variation at this locus has arisen since the divergence of the two species. The time since the species divergence was assigned to T_4.

An accurate estimate of the age of the *B. oleracea*/*B. campestris* divergence appears to be unavailable. The earliest occurrence of brassicaceous pollen in the fossil record appears to date to about five million years before the present (MULLER 1981). Because the origin of plant taxa may considerably predate their appearance in the fossil record (see BRANDL *et al.* 1992), a rough figure of 10 million years was assigned to T_4, the coalescence time of the SLR_1 sequences. Underestimation of T_4 would generate lower divergence times and higher substitution rates, indicating a more recent origin of self-incompatibility.

Estimates of rates and times: Given T_4, the estimates of rates a, b and d, and times T_3 and T_7 through T_9 are virtually independent of the position of the origin of self-incompatibility and the scenarios considered. Table 3 presents these estimates and their standard errors, under the assignment of T_4 as 10 million years. The similarity between T_4 and the estimate for T_7 suggests that the event that marked the divergence between class II S-alleles and the SLR_2 locus occurred close to the speciation event between *B. oleracea* and *B. campestris*.

Under both origin 1 and origin 2, the divergence rate (y) corresponding to the lineage immediately ancestral to the class II S alleles and the SLR_2 sequences

TABLE 4

Origin-independent divergence rates and times

	$y = a$	$y = b$
Substitution rates		
Class I *S* alleles (c)	2.2 ± 0.4	2.0 ± 0.6
S_6 (e)	1.0 ± 0.4	1.0 ± 0.4
S_{13} (f)	1.1 ± 0.4	1.0 ± 0.4
Divergence times		
Class I/II *S* alleles (T_6)	37.9 ± 7.2	41.1 ± 12.7
Class I *S* alleles (T_{10})	23.5 ± 5.0	25.5 ± 8.1
S_{910}/SRK_{910} (T_{11})	17.2 ± 4.2	18.7 ± 6.5
SLG_{WS1}/B_{292} (T_{12})	20.1 ± 4.1	21.8 ± 6.8
S_6/S_{13} (T_{13})	19.7 ± 3.9	21.3 ± 6.6
S_{13}/SLG_{WS1} (T_{14})	15.0 ± 3.2	16.3 ± 5.1
S_{13}/BS_8 (T_{15})	13.8 ± 3.0	15.0 ± 4.8
S_6/S_{14} (T_{16})	18.1 ± 3.9	19.6 ± 6.3
S_{14}/S_{22} (T_{17})	16.7 ± 3.9	18.1 ± 5.9
S_6/SRK_6 (T_{18})	12.2 ± 2.9	13.3 ± 4.4

Values are means ± SE. Rates and times as in Table 3.

affects rates c, e and f and times T_6 and T_{10} through T_{18} (Table 4). These estimates indicate that the age of extant *S* allele lineages (T_6) significantly exceeds the divergence of *B. oleracea* and *B. campestris* (T_4). Because the change of function that led to expression of SSI necessarily occurred before the coalescence of all extant *S*-alleles, T_6 provides a lower bound for the time since the origin of SSI.

Table 5 presents estimates for the remaining parameters, which depend upon the position of the origin of self-incompatibility expression as well as the rates assigned to the internal branches. Estimates for the age of sporophytic self-incompatibility function (Table 6) represent averages of the ages of the nodes flanking the point of origin (T_2 and T_5 for origin 1; T_5 and T_6 for origin 2).

Rate comparisons: A series of pairwise comparisons were made among the divergence rate estimates. Under all scenarios and both origin hypotheses considered, the *S*-allele rates lie *below* the non-*S* SLR_1/SLR_2 rate, with the pollen-dominant *S*-allele rate significantly lower ($c < a$). This result contradicts TRICK and HEIZMANN's (1992) conclusion that *S* alleles evolve at much higher rates, a result that follows directly from the externally imposed assumption that the *SLG/SRK* system of sporophytic self-incompatibility arose within the Brassicaceae.

AtS_1, the homologue of *SLG* in self-compatible Arabidopsis, appears to have evolved significantly slower than SLR_1 and SLR_2 sequences ($a > d$), and pollen-dominant *S*-alleles S_6 and S_{13} significantly slower than all other sequences ($a, b, c, d, g > e, f$). Although not corrected for multiple comparisons, these trends consistently appeared in all scenarios considered.

Because AtS_1 is the only extant sequence associated with rate d and because this rate differs significantly from other non-*S* sequences, scenarios that assign d as the ancestral non-*S* rate ($x = d$) may be unrealistic.

TABLE 5

Origin- and scenario-dependent divergence rates and times

		$x = a$	$x = d$
Origin 1: *SRK* predates ARK_1			
RLK_4/SLR_1 (T_1[a])		118.4 ± 18.4	179.7 ± 38.9
SLR_1/ARK_1 (T_2[a])		50.6 ± 8.7	59.1 ± 12.0
		$y = a$	$y = b$
ARK_1 (g[b])			
$z = b$		nc[c]	3.7 ± 1.1
$z = c$		3.8 ± 0.8	3.5 ± 1.1
ARK_1/S_2 (T_5[†])			
$z = b$		nc	45.7 ± 13.7
$z = c$		44.4 ± 8.6	48.2 ± 14.6
	$x = g$	$x = a$	$x = d$
Origin 2: ARK_1 predates *SRK*			
ARK_1 (g[a])	3.6 ± 0.9	4.0 ± 0.9	3.6 ± 1.1
RLK_4/SLR_1 (T_1[b])	115.7 ± 28.5	118.4 ± 18.4	179.7 ± 38.9
SLR_1/ARK_1 (T_2[b])	52.1 ± 10.2	50.6 ± 8.7	59.1 ± 12.0
ARK_1/S_2 (T_5[b])	46.2 ± 10.8	43.0 ± 9.8	47.4 ± 13.9

Values are means ± SE.
[a] 10^6 years.
[b] 10^{-9} substitutions/site/year.
[c] Not considered.

TABLE 6

Origin of sporophytic self-incompatibility

	$y = a$	$y = b$
Origin 1: *SRK* predates ARK_1[a]		
$x = a, z = b$	nc[b]	48.1 ± 9.4
$x = a, z = c$	47.5 ± 7.8	49.4 ± 9.8
$x = d, z = b$	nc	52.0 ± 10.3
$x = d, z = c$	51.6 ± 9.0	53.3 ± 10.7
Origin 2: ARK_1 predates *SRK*[c]		
$x = a$	40.4 ± 7.3	42.1 ± 9.0
$x = d$	42.5 ± 9.0	43.9 ± 10.2
$x = g$	41.9 ± 7.7	43.2 ± 9.2

Values are means ± SE.
[a] $(T_2 + T_5)/2$, 10^6 years.
[b] Not considered.
[c] $(T_5 + T_6)/2$, 10^6 years.

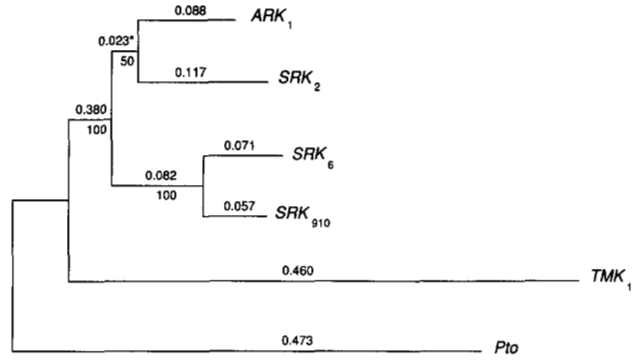

FIGURE 5.—Neighbor-joining phylogeny for the catalytic kinase domain of *SRK* sequences and their homologues. GLS branch lengths and bootstrap values for 1000 resamplings are shown. An asterisk indicates the single branch length that does not differ significantly from 0. The residual sum of squares (5, with 6 degrees of freedom) indicated an acceptable fit.

Similarly, rate *g*, uniquely associated with ARK_1 among extant sequences, is also an unlikely candidate for the ancestral non-*S* rate. Finally, the derivation of the pollen-recessive *S*-allele class from non-*S* SLR_2 sequences ($y = a$) seems less likely than the reverse ($y = b$). Elimination of these possibilities suggests that the most plausible scenario would set the ancestral non-*S* rate to the SLR_1 / SLR_2 rate ($x = a$) and assume that SLR_2 arose from pollen-recessive *S*-allele sequences ($y = b$). Under these assignments, the origin of self-incompatibility would exceed the age of the *B. oleracea*/*B. campestris* divergence by a factor of 4.8 or 4.9 under origin 1 and 4.2 under origin 2 (Table 6).

Examination of the status of ARK_1: In an effort to resolve whether the expression of self-incompatibility arose at origin 1 or origin 2, I conducted a phylogenetic analysis of the putative catalytic kinase domains of SRK-like receptor kinase proteins. The kinase domains of SRK_2 and SRK_6 from *B. oleracea*, SRK_{910} from *B. napus*, ARK_1 and TMK_1 from *A. thaliana* and *Pto* from *L. esculentum* were aligned, both simultaneously, using CLUSTAL as described for the SLG-like domain, and pairwise, against SRK_{910}. Whereas alignments within the ARK_1/*SRK* group were relatively unambiguous (72–88% similarity), the number and positions of gaps in some regions of the alignments between members of this group and the non-SRK proteins (~35% similarity) appeared equivocal.

Figure 5 shows the NJ topology constructed from Poisson-corrected distances, using the tomato sequence *Pto* to root the tree. A GLS relative rates test indicated rate homogeneity throughout the tree.

ARK_1 appears within the *SRK* cluster, suggesting the origin of self-incompatibility before its divergence. Still, this conclusion is not unequivocal, as the length of the branch supporting this node is not significantly positive. To examine further the order of divergence, I obtained GLS estimates for the three possible alternative trees that preserve the positions of the non-*SRK* sequences but place the divergence of ARK_1 before the *SRK* genes. Two returned very highly significant values for the residual sum of squares ($X^2_{[5]} = 30$ and 31), whereas the remaining topology ({ARK_1, [SRK_2, (SRK_6, SRK_{910})]}) contained a branch with significantly negative length. This analysis appears to support origin 1, which hypothesizes that ARK_1 descended from an *SRK* lineage rather than a non-*S* lineage.

DISCUSSION

Polyphyletic origins of self-incompatibility

Multiple recruitments from pre-existing multigene families: Under the classical view, gametophytic self-incompatibility represents the ancestral condition of all flowering plants, with sporophytic systems having evolved directly from gametophytic systems (see Introduction). Both gametophytic and sporophytic elements are reportedly involved in the expression of self-incompatibility in a member of the Brassicaceae (LEWIS *et al.* 1988; ZUBERI and LEWIS 1988), although such factors have not been characterized at either the physiological or genetic levels. Homologues to the *S*-RNases that mediate GSI in the Solanaceae have been detected in self-compatible Arabidopsis (TAYLOR and GREEN 1991; TAYLOR *et al.* 1993), a member of the Brassicaceae for which SSI is typical. Conversely, *SRK*-like genes occur in maize (WALKER and ZHANG 1990; ZHANG and WALKER 1993), a member of the Poaceae for which GSI is typical, and in *Petunia inflata* (MU *et al.* 1994), which expresses the *S*-RNase-mediated form of GSI. Preservation of the hypothesis of monophyly of the two systems of self-incompatibility would appear to require regarding the *S*-RNase homologues in Arabidopsis as vestiges of the ancestral gametophytic system; however, the pres-

ence of *SRK*-like genes in maize and Petunia would indicate the reverse: that GSI descended from SSI. This contradiction suggests that the original assumption of monophyly is false.

More plausible is the interpretation that some lineages within these multigene families have in fact had no history of expression of self-incompatibility, particularly in cases in which the genetic control of self-incompatibility associated with the multigene family departs from that of the plant family. Under this view, GSI expressed in the Solanaceae and SSI in the Brassicaceae represent independent recruitments of the *S* locus from different pre-existing multigene families.

How many times has self-incompatibility evolved? Recent information on the genetic basis and physiological mechanism of self-incompatibility supports the abandonment of the canonical assumption of monophyly of all homomorphic systems and favors a new paradigm hypothesizing multiple, independent origins of SI through recruitment of pre-existing genes. Whereas a definitive test of independence awaits further genetic information, the minimum number of distinct origins may be inferred from the number of distinct multigene families from which the various forms of self-incompatibility were derived. In the absence of sequence-level information, similarity in physiological mechanism or the nature of genetic regulation serves to indicate homology.

Solanaceae: That rRNA in incompatible but not compatible pollen tubes is degraded during the rejection response (McCLURE *et al.* 1990) suggests that the RNase activity exhibited by solanaceous *S* alleles is fundamental to this system of GSI. Yet, the possibility remains that rRNA degradation followed rather than caused cell death (NEWBIGIN *et al.* 1993).

Self-compatibility appears to correlate with low or absent RNase activity. Expression of an *S*-allele transgene associated with self-incompatibility in *Nicotiana alata* failed to confer self-incompatibility in *N. tabacum* (MURFETT *et al.* 1992). Whereas the transgene directed the proper processing and secretion of the expected glycoprotein, which expressed RNase activity, the pattern of tissue expression differed and the concentration of the glycoprotein in styles was two orders of magnitude below that of self-incompatible *N. alata*. KOWYAMA *et al.* (1994) demonstrated the cosegregation of self-compatibility with an allele (S_c) of the *S*-locus in a line derived from a natural population of *L. peruvianum*. The protein product of S_c exhibited much reduced RNase activity compared with *S* alleles that function in self-incompatibility. ANDERSON *et al.* (1993) have determined from cDNA sequences that the S_c allele carries a base-pair change that would direct the replacement of a histidine in a region designated as conserved by IOERGER *et al.* (1991). This histidine may be necessary for *S*-RNase activity, as it lies in the region determined to be the active site of homologous fungal RNases (KURIHARA *et al.* 1992).

Irrespective of whether RNase activity is the direct agent of the cytotoxic aspect of self-incompatibility, it is clearly not *sufficient*. CLARK *et al.* (1990) found comparable levels of *S*-locus mRNA and RNase activity in self-incompatible and pseudoself-compatible varieties of *Petunia hybrida*. AI *et al.* (1992) demonstrated that self-compatibility in a *P. hybrida* line derives from two genes, S_0 and S_x, which had been shown to be alleles of the *S* locus by means of segregation experiments involving crosses with self-incompatible *P. inflata* (AI *et al.* 1991). The protein products of both genes showed RNase activity. Whereas self-compatibility associated with S_0 appears to derive from lesions in S_0 itself, self-compatibility associated with S_x appears to require additional, nonallelic factors (AI *et al.* 1991).

The most convincing demonstration that *S*-RNase proteins and RNase activity are *necessary* for the expression of self-incompatibility comes from recent transformation experiments in Petunia (HUANG *et al.* 1994; LEE *et al.* 1994) and Nicotiana (MURFETT *et al.* 1994). These studies are the first to demonstrate that transformation can confer incompatibility directed specifically against the introduced *S*-allele.

LEE *et al.* (1994) transformed self-incompatible S_1S_2 *P. inflata* individuals with a construct containing the S_3 gene together with 2 KB of its upstream region. Whereas some of the transgenic plants appeared to show cosuppression (self-compatibility owing to the reduced expression of endogenous *S* alleles), four both expressed all three glycoproteins and rejected pollen that expressed any of those specificities. Low levels of expression of the S_3 glycoprotein correlated with partial self-compatibility. Loss-of-function experiments, involving transformation with antisense sequences, caused *S*-allele-specific suppression of incompatibility.

HUANG *et al.* (1994) modified the S_3 sequence, introducing single base-pair changes that directed the replacement of a conserved histidine with either asparagine or arginine. In homologous fungal RNases, the corresponding histidine lies in the active site (KURIHARA *et al.* 1992). Transformation of self-incompatible S_1S_2 individuals using these modified S_3 sequences was accomplished using procedures described by LEE *et al.* (1994). Two out of 162 transgenic plants expressed the modified S_3 protein at levels comparable with those expressed by plants that had acquired the ability to reject S_3 pollen after transformation using the native S_3 sequence. Whereas the two transformants remained self-incompatible, rejecting self-pollen as well as S_1 and S_2 pollen produced by tester lines, seed set with S_3 pollen was comparable with that of compatible pollinations. Accordingly, the modified S_3 protein showed no detectable RNase activity.

MURFETT *et al.* (1994) introduced into self-compati-

ble *N. alata/ N. langsdorfi* hybrids a construct containing the *N. alata S* allele S_{A2} and its 3′ flanking region under the control of a promoter which directs high expression in pollen and pistil. Specific rejection of pollen bearing S_{A2} correlated with the expression of the S_{A2}-RNase protein and with RNase activity.

These transformation experiments together provide the best evidence that *S*-RNase proteins and RNase activity *per se* mediate self-incompatibility.

Papaveraceae: FOOTE *et al.* (1994) reported the cloning and characterization of a gene that cosegregates with gametophytic self-incompatibility in the field poppy (*Papaver rhoeas*). They established that this gene is a component of the *S*-locus by demonstrating that its protein product specifically inhibits germination of pollen that bear the same allele. Nucleotide sequence analysis indicated homology neither to the solanaceous *S*-RNases nor to the brassicaceous *SLG/ SRK* family. This observation, together with the previously demonstrated absence of RNase activity (FRANKLIN-TONG *et al.* 1991) and the mediation of self-incompatibility through distinct physiological mechanisms (FRANKLIN-TONG *et al.* 1993), supports the view that GSI in poppies and in the Solanaceae represent independent recruitments from distinct families of genes (FRANKLIN-TONG and FRANKLIN 1993).

GSI in other families: Other gametophytic systems of self-incompatibility that appear to be distinct from the *S*-RNase system operate in lilies and grasses. Self-incompatibility in *Lilium longiflorum* derives from arrest of pollen tube growth within the hollow lily style, and shows sensitivity to cyclic AMP concentration (TEZUKA *et al.* 1993). Self-incompatibility in *L. martagon* differs with respect to genetic control, with four loci possibly involved (LUNDQVIST 1991). Two unlinked loci (*S* and *Z*) regulate self-incompatibility in those members of the Poaceae for which genetical control has been analyzed in detail (reviewed by LEACH 1983; HAYMAN and RICHTER 1992). HAYMAN (1956) argued against the derivation of the multilocus GSI systems of monocotyledons by duplication of the single-locus systems typical of dicotyledons. First, the independence of expression of the *S* and *Z* genes in determining pollen specificity in monocots contrasts with the epistasis typically expressed in polyploid dicots. Further, self-incompatibility systems that appear to resemble the system in the grasses with respect to both multilocus control and expression have been described in dicot taxa, which show relatively close relationships to monocots (*Ranunculus acris,* OSTERBYE 1975; *Beta vulgaris,* LARSEN 1977). LUNDQVIST (1975) has suggested that the multilocus system of the grasses may in fact have been derived by a *reduction* in the number of controlling loci. Self-incompatibility in rye (*Secale cereale*) may be mediated, not through expression of ribonuclease function, but rather through transduction of signals involving protein phosphorylation (WEHLING *et al.* 1994). These observations are consistent with the view that GSI may derive from at least three independent origins (solanaceous *S*-RNase system, non-*S*-RNase system in poppies, multilocus system in the grasses), with a possible fourth origin in the lilies.

Sporophytic systems: SSI may also have arisen multiple times. Self-incompatibility in Brassica results from the failure of incompatible pollen grains to hydrate at the stigmatic surface (SARKER *et al.* 1988). Although SSI also occurs in the Compositae, the morphological and physiological processes involved in both compatible and incompatible pollination appear quite distinct in sunflowers (*Helianthus annuus,* ELLEMAN *et al.* 1992).

Ipomoea trifida (Convolvulaceae) exhibits classical SSI (KOWYAMA *et al.* 1980). Whereas homologues to the *SLG/ SRK* family have been detected in this species, those elements do not cosegregate with the physiological expression of self-incompatibility (KOWYAMA *et al.* 1993), suggesting a derivation of SSI independent from the system in Brassica. Alternatively, if the *S* locus in Ipomoea is in fact a member of the *SLG/ SRK* family, this system of SSI would appear to have originated before the divergence of superorders Dillenianae and Lamininae.

Statistical approach

Function-specific molecular clocks: Whereas a variety of methods for phylogenetic reconstruction provide estimates and standard errors for branch lengths, the quantities of biological interest in many cases are rates of genetic divergence and node ages. Using estimates of branch lengths, interpreted as products of divergence rates and times, to estimate rates and times separately requires external information. For example, datings of fossils or biogeographical events are often used to determine node ages, with divergence rates inferred under the assumption of a molecular clock. Whereas local, or *lineage-specific,* molecular clocks may operate within closely related groups (O'HUIGIN and LI 1992), variation in evolutionary rate among lineages (particularly the hominoids, GU and LI 1992) precludes imposition of universal molecular clocks.

The generalized least-squares method introduced in the present analysis provides approximate closed-form expressions for estimates and standard errors of *function-specific* divergence rates and node ages, incorporating externally determined values for any number of parameters [see (22) and (23)]. HASEGAWA *et al.* (1985) estimated rates of transition and transversion substitutions in regions subject to rapid evolution (third codon positions) and moderate evolution (first and second codon positions). To confirm and refine the approximate closed-form solutions, I adapted the HASEGAWA *et al.* approach to estimate substitution rates

that vary according to function rather than region. In all cases studied, the closed-form expressions for the parameter estimates required no further improvement, indicating that the approximations involved introduced negligible error. The procedure used in the present analysis provides a general means of determining node ages and divergence rates for phylogenies in which rate inhomogeneity precludes imposition of a global molecular clock.

Calibration under trans-specific evolution: Trans-specific sharing of genetic variation presents difficulties for the estimation of divergence rates and times. The conventional practice of equating the divergence time of species with the divergence time of genes is clearly inappropriate if the age of the genes greatly exceeds the age of the taxa that carry them. Divergence among genes reflects substitutions accumulated both before and after the divergence of species. To minimize the fraction of substitutions that occurred before the species divergence, the minimum-minimum method of SATTA *et al.* (1991, 1993) bases the calibration on the interspecific sequence comparison that shows the least divergence. They cautioned that this method can overestimate the genetic divergence after speciation if all available sequences predate that event, and selecting the lowest observed divergence to represent the expected divergence can result in underestimation.

By analyzing *S*-allele sequences together with homologous sequences that do not cosegregate with the physiological expression of self-incompatibility, the method introduced here calibrates the entire phylogeny on the basis of lineages that are not subject to trans-specific sharing. Maintenance of trans-specific polymorphisms due to the strong balancing selection characteristic of the expression of self-incompatibility is not expected for the non-*S* members of the multigene family. This expectation was verified for the SLR_1 sequences by comparing the number of informative sites that supported gene phylogenies congruent and incongruent with the species phylogeny. Identification of the divergence time of species *B. oleracea* and *B. campestris* with the divergence time among SLR_1 sequences (T_4 in Figure 4) permitted estimation of all divergence rates and node ages.

Tempo and mode of evolution at the *S* locus

Estimation of the origin of sporophytic self-incompatibility: This analysis indicates that the coalescence time of all extant *S* alleles, both pollen dominant and recessive, exceeds the species divergence time by fourfold (T_6, Table 4). Results obtained here confirm the origin of *S*-allele lineages before the *B. oleracea*/*B. campestris* divergence (DWYER *et al.* 1991) and considerably increase the lower bound on the age of the *SLG*/*SRK* system of sporophytic self-incompatibility.

Inferring an upper bound for the origin of SSI depends on whether ARK_1, the *SRK*-like sequence from self-compatible Arabidopsis, descends from a functional *SRK* gene (origin 1) or a lineage having no history of self-incompatibility (origin 2). Analysis of the putative catalytic kinase domain of *SRK* and its homologues (Figure 5) suggests that the lineage ancestral to ARK_1 may in fact have functioned in self-incompatibility. This interpretation would increase the age of the *SLG*/*SRK* system of sporophytic self-incompatibility to five times the age of the *B. oleracea*/*B. campestris* divergence.

Ambiguity in the NJ topology (Figures 3 and 4) leaves open the possibility of an even more ancient origin. Because the placement of the divergence of ARK_1 after the SLR_1 cluster is not well supported (65% NJ bootstrap probability and nonsignificant GLS branch length; see Figure 2), the origin of SSI prior to the divergence of the SLR_1 cluster cannot be excluded. This topology was not considered in detail because it includes a branch of (nonsignificantly) negative length. Placing the origin of SSI before ARK_1 and ARK_1 in turn before SLR_1 would increase the upper bound on the origin of SSI to T_1 (Figure 5). Whereas parameter estimates were not obtained for this origin hypothesis, the present analysis provides a value of ~120 million years for T_1, adopting the very rough estimate of 10 million years for the separation of the species (see Table 5). This more ancient upper bound nevertheless postdates the monocot/dicot divergence, which is now believed to have occurred upwards of 200 mya (MARTIN *et al.* 1989; WOLFE *et al.* 1989; BRANDL *et al.* 1992).

Rates of divergence: TRICK and HEIZMANN (1992) reported that at the amino acid sequence level Brassica *S* alleles evolve an order of magnitude faster than non-*S* members of the *SLG*/*SLR*/*SRK* multigene family. This extraordinarily high divergence rate likely represents an artifact of the assumption that this system of SSI arose since the *B. oleracea*/*B. campestris* speciation. Nevertheless, balancing selection can in fact accelerate divergence rates at sites that determine selectively distinct allelic classes (MARUYAMA and NEI 1981; TAKAHATA 1990; SASAKI 1992). Compared with neutral mutations, mutations that create overdominant alleles are more likely to become common and so are less prone to loss. Selection induced by expression of self-incompatibility is similar in many respects to overdominant selection (YOKOYAMA and NEI 1979). Under both selective regimes, the magnitude of the rate acceleration may be small in very highly polymorphic populations in which any given allele is rare.

Because the region that determines *S*-allele specificity has yet to be identified, no estimates of the rates of divergence within such regions are available. The values obtained here, corresponding to nonsynonymous divergence over the entire coding region, provide no evidence of an acceleration of substitution rate upon the

acquisition of self-incompatibility function. In fact, the analysis indicates that pollen-dominant *S*-allele sequences have evolved significantly *slower* than non-*S* sequences ($c < a$; Tables 3 and 4). The extraordinarily high sequence diversity observed among *S* alleles appears to reflect not hypermutability of the *S* locus but rather the maintenance of *S*-allele lineages over very long periods.

Although such *relative* comparisons of substitution rates or node ages are robust to error in the externally determined value assigned as the species divergence time, the *absolute* estimates given are proportional to this value. The substitution rates presented in Tables 3–5, obtained under the assignment of the species divergence time as 10 million years, appear somewhat high in comparison to other protein-coding genes in plants (WOLFE *et al.* 1987). Assignment of a species divergence time below the true value would cause overestimation of the substitution rates and underestimation of node ages.

Rate of origin of new *S* alleles: A key result due to TAKAHATA (1990) established that the drift process under overdominance closely resembles the process under pure neutrality but for a rescaling of time. CLARK (1993) applied this theory for overdominant selection to gametophytic self-incompatibility loci, and VEKEMANS and SLATKIN (1994) modified Takahata's derivation to obtain the scaling factor (f_s) appropriate for GSI. The expected time for coalescence of i to $i - 1$ *S*-allele classes (t_i) is

$$E[t_i] = \frac{4N_e f_s}{i(i-1)}, \qquad (28)$$

for N_e the effective population size. The scaling factor f_s depends on various factors, including the rate of mutation to new *S*-allele specificities, the effective number of specificities present in the population and the effective population size. The expected ratio of intervals between coalescence events (which are independent) does not involve this unknown quantity. Assuming that the expectation of the inverse of t_{i-1} approximates the inverse of the expectation ($E[1/t_{i-1}] \approx 1/E[t_{i-1}]$), the expected ratio of the coalescence times from i to $i - 1$ and from $i - 1$ to $i - 2$ *S*-allele classes is

$$E\left[\frac{t_i}{t_{i-1}}\right] = \frac{i-2}{i}, \qquad (29)$$

with variance given by the square of (29). This expression reflects that the total rate of appearance of new *S* alleles increases with the number of existing *S* alleles.

The estimates of node ages obtained in the present analysis (Tables 3–5) provide some empirical information on the generation of new *S*-allele specificities. In contrast to expectation, the time between bifurcations shows no indication of declining as the number of *S*-allele lineages increases. The *prima facie* indication is that many *S*-allele specificities were established near the origin of the self-incompatibility system and have been continuously maintained, with relatively few extant specificities having arisen subsequently. VEKEMANS and SLATKIN (1994) described a similar trend in their analysis of solanaceous *S*-RNase sequences associated with GSI. It should be recognized that node ages correspond to coalescence times of *S*-allele *lineages* rather than *specificities*, as the specificity of a lineage may change without altering the balancing selection that maintains the lineage. This distinction could account for the apparent slowdown only if the establishment of a new specificity has resulted in the replacement of the immediately ancestral specificity more often than other specificities. Alternatively, the slowdown in appearance of new *S* alleles, if not artifactual, may reflect constraints on the nature of changes that can both preserve self-incompatibility function and generate new specificities.

I thank M. NEI and A. L. HUGHES for introducing methods of phylogenetic analysis to me; X. VEKEMANS and M. L. LAWRENCE for their kind encouragement and for providing manuscripts; and S. KUMAR, A. G. STEPHENSON, N. TAKAHATA, X. VEKEMANS and two anonymous reviewers for very much appreciated comments and suggestions. Support from the Alfred P. Sloan Foundation and U.S. Public Health Service grant GM-37841, which made possible a research leave to the Department of Biology, Pennsylvania State University, University Park, is most gratefully acknowledged. Preliminary results of this study were presented at the Workshop on Molecular Evolution, Rotorua, New Zealand, under the sponsorship of the U.S.-N.Z. Cooperative Science Program of the National Science Foundation and at the XV International Botanical Congress, Yokohama, Japan.

LITERATURE CITED

AI, Y., E. KRON and T.-h. KAO, 1991 *S*-alleles are retained and expressed in a self-incompatible cultivar of *Petunia hybrida*. Mol. Gen. Genet. **230:** 353–358.

AI, Y. J., D. S. TSAI and T.-h. KAO, 1992 Cloning and sequencing of cDNAs encoding two *S*-proteins of a self-compatible cultivar of *Petunia hybrida*. Plant Mol. Biol. **19:** 523–528.

ANDERSON, M. A., E. C. CORNISH, S.-L. MAU, E. G. WILLIAMS, R. HOGGART *et al.*, 1986 Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. Nature **321:** 48–44.

ANDERSON, M. A., A. BACIC, E. NEWBIGIN and A. E. CLARKE, 1993 Self and non-self recognition during pollination in flowering plants. XV International Botanical Congress, Yokohama, Japan.

ARDEN, B. and J. KLEIN 1982 Biochemical comparison of major histocompatibility complex molecules from different subspecies of *Mus musculus:* evidence for trans-specific evolution of alleles. Proc. Natl. Acad. Sci. USA **79:** 2342–2346.

BOYES, D. C., C.-H. CHEN, T. TANTIKANJANA, J. J. ESCH and J. B. NASRALLAH, 1991 Isolation of a second *S*-locus-related cDNA from *Brassica oleracea:* genetic relationships between the *S* locus and two related loci. Genetics **127:** 221–228.

BRANDL, R., W. MANN and M. SPRINZL, 1992 Estimation of the monocot-dicot age through tRNA sequences from the chloroplast. Proc. R. Soc. Lond. Ser. B **249:** 13–17.

BULMER, M., 1991 Use of the method of generalized least squares in reconstructing phylogenies from sequence data. Mol. Biol. Evol. **8:** 868–883.

CAVALLI-SFORZA, L. L., and A. W. F. EDWARDS, 1967 Phylogenetic analysis: Models and estimation procedures. Am. J. Hum. Genet. **19:** 233–257.

CHANG, C., G. E. SCHALLER, S. E. PATTERSON, S. F. KWOK, E. M. MEYEROWITZ *et al.*, 1992 The TMK_1 gene from Arabidopsis codes for a protein with structural and biochemical characteristics of a receptor protein kinase. Plant Cell **4:** 1263–1271.

CHEN, C.-H., and J. B. NASRALLAH, 1990 A new class of *S* sequences defined by a pollen recessive self-incompatibility allele of *Brassica oleracea*. Mol. Gen. Genet. **222:** 241–248.

CLARK, A. G., 1993 Evolutionary inferences from molecular characterization of self-incompatibility alleles, pp. 79–108 in *Mechanisms of Molecular Evolution,* edited by N. TAKAHATA and A. G. CLARK. Sinauer Associates, Sunderland, MA.

CLARK, K. R., J. J. OKULEY, P. D. COLLINS and T. L. SIMS, 1990 Sequence variability and developmental expression of *S*-alleles in self-incompatible and pseudo-self-compatible petunia. Plant Cell **2:** 815–826.

CLARKE, A. E., and E. NEWBIGIN, 1993 Molecular aspects of self-incompatibility in flowering plants. Annu. Rev. Genet. **27:** 257–279.

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

DE NETTANCOURT, D., R. ECOCHARD, M. D. G. PERQUIN, T. VAN DER DRIFT and M. WESTERHOF, 1971 The generation of new *S* alleles at the incompatibility locus of *Lycopersicon peruvianum* Mill. Theor. Appl. Genet. **41:** 120–129.

DICKINSON, H. G., 1990 Self-incompatibility in flowering plants. BioEssays **12:** 155–161.

DICKINSON, H. G., 1994 Simply a social disease? Nature **367:** 517–518.

DNASTAR, 1992 DNASTAR, Inc., Madison, WI.

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.

DWYER, K. G., B. A. LALONDE, J. B. NASRALLAH and M. E. NASRALLAH, 1992 Structure and expression of AtS_1, an *Arabidopsis thaliana* gene homologous to the *S*-locus related genes of Brassica. Mol. Gen. Genet. **231:** 442–448.

ELLEMAN, C. J., V. FRANKLIN-TONG and H. G. DICKINSON, 1992 Pollination in species with dry stigmas: the nature of the early stigmatic response and the pathway taken by pollen tubes. New Phytol. **12:** 413–424.

FIGUEROA, R., E. GÜNTHER and J. KLEIN, 1988 MHC polymorphism pre-dating speciation. Nature **335:** 265–267.

FISHER, R. A., 1961 A model for the generation of self-sterility alleles. J. Theor. Biol. **1:** 411–414.

FITCH, W. M., and E. MARGOLIASH, 1967 A method for estimating the number of invariant amino acid coding positions in a gene using cytochrome c as a model case. Biochem. Genet. **1:** 65–71.

FOOTE, H. C. C., J. P. RIDE, V. E. FRANKLIN-TONG, E. A. WALKER, M. J. LAWRENCE *et al.*, 1994 Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas* L. Proc. Natl. Acad. Sci. USA **91:** 2265–2269.

FRANKLIN-TONG, V. E., and F. C. H. FRANKLIN, 1993 Gametophytic self-incompatibility: Contrasting mechanisms for Nicotiana and Papaver. Trends Cell Biol. **3:** 340–345.

FRANKLIN-TONG, V. E., K. K. ATWAL, E. C. HOWELL, M. J. LAWRENCE and F. C. H. FRANKLIN, 1991 Self-incompatibility in *Papaver rhoeas:* there is no evidence for the involvement of stigmatic ribonuclease activity. Plant Cell Environ. **14:** 423–429.

FRANKLIN-TONG, V. E., J. P. RIDE, N. D. READ, A. J. TREWAVAS and F. C. H. FRANKLIN, 1993 The self-incompatibility response in *Papaver rhoeas* is mediated by cytosolic free calcium. Plant J. **4:** 163–177.

GAUDE, T., A. FRIRY, P. HEIZMANN, C. MARIAC, M. ROUGIER *et al.*, 1993 Expression of a self-incompatibility gene in a self-compatible line of *Brassica oleracea*. Plant Cell **5:** 75–86.

GORING, D. R., P. BANKS, W. D. BEVERSDORF and S. J. ROTHSTEIN, 1992 Use of the polymerase chain reaction to isolate an *S*-locus glycoprotein cDNA introgressed from *Brassica campestris* into *B. napus* ssp. *oleifera*. Mol. Gen. Genet. **234:** 185–192.

GU, X., and W.-H. LI, 1992 Higher rates of amino acid substitution in rodents than in humans. Mol. Phylog. Evol. **1:** 211–214.

HASEGAWA, M., H. KISHINO and T. YANO, 1985 Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22:** 160–174.

HAYMAN, D. L., 1956 The genetic control of incompatibility in *Phalaris coerulescens* desf. Aust. J. Biol. Sci. **9:** 323–331.

HAYMAN, D. L., and J. RICHTER, 1992 Mutations affecting self-incompatibility in *Phalaris coerulescens* Desf. (Poaceae). Heredity **68:** 495–503.

HEALY, M. J. R., 1986 *Matrices for Statistics.* Clarendon Press, Oxford.

HINATA, K., M. WATANABE, K. TORIYAMA and A. ISOGAI, 1993 A review of recent studies on homomorphic self-incompatibility. Int. Rev. Cytol. **143:** 257–296.

HODGKIN, T., G. D. LYON and H. G. DICKINSON, 1988 Recognition in flowering plants: a comparison of the Brassica self-incompatibility system and plant pathogen interactions. New Phytol. **110:** 557–569.

HUANG, S., H.-S. LEE, B. KARUNANADAA and T.-h. KAO, 1994 Ribonuclease activity of *Petunia inflata* S proteins is essential for rejection of self-pollen. Plant Cell **6:** 1021–1028.

IDE, H., M. KIMURA, M. ARAI and G. FUNATSU, 1991 The complete amino acid sequence of ribonuclease from the seeds of bitter gourd (*Momordica charantia*). FEBS Lett. **284:** 161–164.

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.

IOERGER, T. R., J. R. GOHLKE, B. XU and T.-h. KAO, 1991 Primary structural features of the self-incompatibility protein in Solanaceae. Sex. Plant Reprod. **4:** 81–87.

ISOGAI, A., S. YAMAKAWA, H. SHIOZAWA, S. TAKAYAMA, H. TANAKA *et al.*, 1991 The cDNA sequence of NS_1 glycoprotein of *Brassica campestris* and its homology to *S*-locus-related glycoproteins of *B. oleracea*. Plant Mol. Biol. **17:** 269–271.

JOST, W., H. BAK, K. GLUND, P. TERPSTRA and J. J. BEINTEME, 1991 Amino acid sequence of an extracellular, phosphate-starvation-induced ribonuclease from cultured tomato (*Lycopersicon esculentum*) cells. Eur. J. Biochem. **198:** 1–6.

KIMURA, M., and T. OHTA, 1972 On the stochastic model for estimation of mutational distance between homologous proteins. J. Mol. Evol. **2:** 87–90.

KOWYAMA, Y., N. SHIMANO and T. KAWASE, 1980 Genetic analysis of incompatibility in the diploid Ipomoea species closely related to the sweet potato. Theor. Appl. Genet. **58:** 149–155.

KOWYAMA, Y., K. KAKEDA and T. HATTORI, 1993 Genetical and molecular biological aspects of self-incompatibility in the genus Ipomoea. XV International Botanical Congress, Yokohama, Japan.

KOWYAMA, Y., C. KUNZ, I. LEWIS, E. NEWBIGIN, A. E. CLARKE *et al.*, 1994 Self-compatibility in a *Lycopersicon peruvianum* variant (LA2157) is associated with a lack of style S-RNase. Theor. Appl. Genet. **88:** 859–864.

KUMAR, S., K. TAMURA and M. NEI, 1993 MEGA: molecular evolutionary genetics analysis, Version 1.0. The Pennsylvania State University, University Park, PA.

KURIHARA, H., Y. MITSUI, K. OHGI, M. IRIE, H. MIZUNO *et al.*, 1992 Crystal and molecular structure of RNase Rh, a new class of microbial ribonuclease from *Rhizopus niveus*. FEBS Lett. **306:** 189–192.

LALONDE, B. A., M. E. NASRALLAH, K. G. DWYER, C.-H. CHEN, B. BARLOW *et al.*, 1989 A highly conserved Brassica gene with homology to the *S*-locus-specific glycoprotein structural gene. Plant Cell **1:** 249–258.

LARSEN, K., 1977 Self-incompatibility in *Beta vulgaris* L. Hereditas **85:** 227–248.

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 **335:** 268–271.

LEACH, C. R., 1983 Are there more than two self-incompatibility loci in the grasses? Heredity **52:** 303–305.

LEE, H.-S., A. SINGH and T.-h. KAO, 1992 RNase X2, a pistil-specific ribonuclease from *Petunia inflata,* shares sequence similarity with solanaceous *S* proteins. Plant Mol. Biol. **20:** 1131–1141.

LEE, H.-S., S. HUANG and T.-h. KAO, 1994 *S* proteins control rejection of incompatible pollen in *Petunia inflata*. Nature **367:** 560–563.

LEWIS, D., and L. K. CROWE, 1954 Structure of the incompatibility

gene. IV. Types of mutations in *Prunus avium* L. Heredity **8:** 357–363.

LEWIS, D., S. C. VERMA and M. I. ZUBERI, 1988 Gametophytic-sporophytic incompatibility in the Cruciferae— *Raphanus sativus*. Heredity **61:** 355–366.

LI, P., and J. BOUSQUET, 1992 Relative-rate test for nucleotide substitutions between two lineages. Mol. Biol. Evol. **9:** 1185–1189.

LÖFFLER, A., K. GLUND and M. IRIE, 1993 Amino acid sequence of an intracellular, phosphate-starvation-induced ribonuclease from cultured tomato (*Lycopersicon esculentum*) cells. Eur. J. Biochem. **214:** 627–633.

LUNDQVIST, A., 1975 Complex self-incompatibility systems in angiosperms. Proc. R. Soc. Lond. Ser. B. **188:** 235–245.

LUNDQVIST, A., 1991 Four-locus *S*-gene control of self-incompatibility made probable in *Lilium martagon* (Liliaceae). Hereditas **114:** 57–63.

MARTIN, G. B., S. H. BROMMONSCHENKEL, J. CHUNWONGSE, A. FRARY, M. W. GANAL *et al.*, 1993 Map-based cloning of a protein kinase gene conferring disease resistance in tomato. Science **262:** 1432–1436.

MARTIN, W., A. GIERL and H. SAEDLER, 1989 Molecular evidence for pre-Cretaceous angiosperm origins. Nature **339:** 46–48.

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

MCCLURE, B. A., V. HARING, P. R. EBERT, M. A. ANDERSON, R. J. SIMPSON *et al.*, 1989 Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. Nature **342:** 955–957.

MCCLURE, B. A., J. E. GRAY, M. A. ANDERSON, and A. E. CLARKE, 1990 Self-incompatibility in *Nicotiana alata* involves degradation of pollen rRNA. Nature **347:** 757–760.

MU, J.-H., H.-S. LEE and T.-h. KAO, 1994 Characterization of a pollen-expressed receptor-like kinase gene of *Petunia inflata* and the activity of its encoded kinase. Plant Cell **6:** 709–721.

MULLER, J., 1981 Fossil pollen records of extant angiosperms. Bot. Rev. **47:** 1–142.

MURFETT, J., E. C. CORNISH, P. R. EBERT, I. BÖNIG, B. A. MCCLURE *et al.*, 1992 Expression of a self-incompatibility glycoprotein (S_2-ribonuclease) from *Nicotiana alata* in transgenic *Nicotiana tabacum*. Plant Cell **4:** 1063–1074.

MURFETT, J., T. L. ATHERTON, B. MOU, C. S. GASSER and B. A. MCCLURE, 1994 *S*-RNase expressed in transgenic Nicotiana causes *S*-allele-specific pollen rejection. Nature **367:** 563–566.

MUSE, S. V., and B. S. WEIR, 1992 Testing for equality of evolutionary rates. Genetics **132:** 269–276.

NASRALLAH, J. B., and M. E. NASRALLAH, 1993 Pollen-stigma signaling in the sporophytic self-incompatibility response. Plant Cell **5:** 1325–1335.

NASRALLAH, J. B., T.-h. KAO, M. L. GOLDBERG and M. E. NASRALLAH, 1985 A cDNA clone encoding an *S*-locus-specific glycoprotein from *Brassica oleracea*. Nature **318:** 263–267.

NASRALLAH, J. B., T.-h. KAO, C.-H. CHEN, M. L. GOLDBERG and M. E. NASRALLAH, 1987 Amino acid sequence of glycoproteins encoded by three alleles of the *S*-locus of *Brassica oleracea*. Nature **326:** 617–619.

NASRALLAH, J. B., S.-M. YU and M. E. NASRALLAH, 1988 Self-incompatibility genes of *Brassica oleracea:* expression, isolation, and structure. Proc. Natl. Acad. Sci. USA **85:** 5551–5555.

NASRALLAH, J. B., T. NISHIO and M. E. NASRALLAH, 1991 The self-incompatibility genes of Brassica: expression and use in genetic ablation of floral tissues. Annu. Rev. Plant Physiol. Plant Mol. Biol. **42:** 391–422.

NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia Univ. Press, New York.

NEI, M., J. C. STEPHENS and N. SAITOU, 1985 Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. Mol. Biol. Evol. **2:** 66–85.

NEWBIGIN, E., M. A. ANDERSON and A. E. CLARKE, 1993 Gametophytic self-incompatibility systems. Plant Cell **5:** 1315–1324.

O'HUIGIN, C., and W.-H. LI, 1992 The molecular clock ticks regularly in muroid rodents and hamsters. J. Mol. Evol. **35:** 377–384.

OSTERBYE, U., 1975 Self-incompatibility in *Ranunculus acris* L. Hereditas **80:** 91–112.

PANDEY, K. K., 1960 Evolution of gametophytic and sporophytic systems of self-incompatibility in angiosperms. Evolution **14:** 98–115.

PRESS, W. H., B. P. FLANNERY, S. A. TEUKOLSKY and W. T. VETTERLING, 1986 *Numerical Recipes: The Art of Scientific Computing*. Cambridge Univ. Press, Cambridge.

RAO, C. R., 1973 *Linear Statistical Inference and Its Applications*. John Wiley & Sons, New York.

ROBERT, L. S., S. ALLARD, T. M. FRANKLIN and M. TRICK, 1994 Sequence and expression of endogenous *S*-locus glycoprotein genes in self-compatibility *Brassica napus*. Mol. Genet. Genet. **242:** 209–216.

RZHETSKY, A., and M. NEI, 1992 Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. J. Mol. Evol. **35:** 367–375.

SAITOU, N., and M. NEI, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

SARKER, R. H., C. J. ELLEMAN and H. G. DICKINSON, 1988 Control of pollen hydration in Brassica requires continued protein synthesis, and glycosylation is necessary for intraspecific incompatibility. Proc. Natl. Acad. Sci. USA **85:** 4340–4344.

SASAKI, A., 1992 The evolution of host and pathogen genes under epidemiological interaction, pp. 247–263 in *Population Paleo-Genetics*, edited by N. TAKAHATA. Japan Scientific Society Press, Tokyo.

SASSA, H., H. HIRANO and H. IKEHASHI, 1992 Self-incompatibility-related RNases in styles of Japanese pear (*Pyrus serotina* Rehd.). Plant Cell Physiol. **33:** 811–814.

SASSA, H., H. HIRANO and H. IKEHASHI, 1993 Identification and characterization of stylar glycoproteins associated with self-incompatibility genes of Japanese pear, *Pyrus serotina* Rehd. Mol. Gen. Genet. **241:** 17–25.

SATTA, Y., N. TAKAHATA, C. SCHÖNBACH, J. GUTKNECHT and J. KLEIN, 1991 Calibrating evolutionary rates at major histocompatibility complex loci, pp. 51–62 in *Molecular Evolution of the Major Histocompatibility Complex*, edited by J. Klein and D. Klein. Springer-Verlag, Heidelberg.

SATTA, Y., C. O'HUIGIN, N. TAKAHATA and J. KLEIN, 1993 The synonymous substitution rate of the major histocompatibility complex loci in primates. Proc. Natl. Acad. Sci. USA **90:** 7480–7484.

SCHNEIDER, R., G. UNGER, R. STARK, E. SCHNEIDER-SCHERZER and H.-J. THIEL, 1993 Identification of a structural glycoprotein of an RNA virus as a ribonuclease. Science **261:** 1169–1171.

SCUTT, C. P., and R. R. D. CROY, 1992 An S_5 self-incompatibility allele-specific cDNA sequence from *Brassica oleracea* shows high homology to the SLR_2 gene. Mol. Gen. Genet. **232:** 240–246.

SCUTT, C. P., P. J. GATES, J. A. GATEHOUSE, D. BOULTER and R. R. D. CROY, 1990 A cDNA encoding an *S*-locus specific glycoprotein from *Brassica oleracea* plants containing the S_5 self-incompatibility allele. Mol. Gen. Genet. **220:** 409–413.

SIMS, T. L., 1993 Genetic regulation of self-incompatibility. Crit. Rev. Plant Sci. **12:** 129–167.

STEIN, J. C., B. HOWLETT, D. C. BOYES, M. E. NASRALLAH and J. B. NASRALLAH, 1991 Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. Proc. Natl. Acad. Sci. USA **88:** 8816–8820.

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.

TAKAYAMA, S., A. ISOGAI, C. TSUKAMOTO, Y. UEDA, K. HINATA *et al.*, 1987 Sequences of *S*-glycoproteins, products of the *Brassica campestris* self-incompatibility locus. Nature **326:** 102–105.

TAYLOR, C. B., and P. J. GREEN, 1991 Genes with homology to fungal and *S*-gene RNases are expressed in *Arabidopsis thaliana*. Plant Physiol. **96:** 980–984.

TAYLOR, C. B., P. A. BARIOLA, S. B. DEL CARDAYR, R. T. RAINES and P. J. GREEN, 1993 RNS2: a senescence-associated RNase of Arabidopsis that diverged from the *S*-RNases before speciation. Proc. Natl. Acad. Sci. USA **90:** 5118–5122.

TEZUKA, T., S. HIRATSUKA and S. Y. TAKAHASHI, 1993 Promotion of the growth of self-incompatible pollen tubes in lily by cAMP. Plant Cell Physiol. **34:** 955–958.

TOBIAS, C. M., B. HOWLETT and J. B. NASRALLAH, 1992 An *Arabi-*

dopsis thaliana gene with sequence similarity to the *S*-locus receptor kinase of *Brassica oleracea*—sequence and expression. Plant Physiol. **99:** 284–290.

TRICK, M., 1990 Genomic sequence of a Brassica *S*-locus related gene. Plant Mol. Biol. **15:** 203–205.

TRICK, M., and R. B. FLAVELL, 1989 A homozygous *S* genotype of *Brassica oleracea* expresses two *S*-like genes. Mol. Gen. Genet. **218:** 112–117.

TRICK, M., and P. HEIZMANN, 1992 Sporophytic self-incompatibility systems: Brassica *S* gene family. Int. Rev. Cytol. **140:** 485–524.

VAN ENGELEN, F. A., M. V. HARTOG, T. L. THOMAS, B. TAYLOR, A. STURM *et al.*, 1993 The carrot secreted glycoprotein gene *EP1* is expressed in the epidermis and has sequence homology to Brassica *S*-locus glycoproteins. Plant J. **4:** 855–862.

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

WALKER, J. C., 1993 Receptor-like protein kinase genes of *Arabidopsis thaliana.* Plant J. **3:** 451–456.

WALKER, J. C., and R. ZHANG, 1990 Relationship of a putative receptor protein kinase from maize to the *S*-locus glycoproteins of Brassica. Nature **345:** 743–746.

WATANABE, M., I. S. NOU, S. TAKAYAMA, S. YAMAKAWA, A. ISOGAI *et al.*, 1992 Variations in and inheritance of *NS*-glycoprotein in self-incompatible *Brassica campestris* L. Plant Cell Physiol. **33:** 343–351.

WEHLING, P., B. HACKAUF and G. WRICKE, 1994 Phosphorylation of pollen proteins in relation to self-incompatibility in rye (*Secale cereale* L.). Sex. Plant Reprod. **7:** 67–75.

WHITEHOUSE, H. L. K., 1951 Multiple-allelomorph incompatibility of pollen and style in the evolution of the angiosperms. Ann. Bot. **14:** 198–216.

WOLFE, K. H., W.-H. LI and P. M. SHARPE, 1987 Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA **84:** 9054–9058.

WOLFE, K. H., M. GOUY, Y.-W. YANG, P. M. SHARPE and W.-H. LI, 1989 Date of the monocot-dicot divergence estimated from chloroplast DNA sequence data. Proc. Natl. Acad. Sci. USA **86:** 6201–6205.

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

WU, C.-I., and W.-H. LI, 1985 Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82:** 1741–1745.

YOKOYAMA, S., and M. NEI, 1979 Population dynamics of sex-determining alleles in honey bees and self-incompatibility alleles in plants. Genetics **91:** 609–626.

ZHANG, R., and J. C. WALKER, 1993 Structure and expression of the *S* locus-related genes of maize. Plant Mol. Biol. **21:** 1171–1174.

ZUBERI, M. I., and D. LEWIS, 1988 Gametophytic-sporophytic incompatibility in the Cruciferae—*Brassica campestris.* Heredity **61:** 367–377.

Communicating editor: A. G. CLARK