Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 9823–9827, November 1991
Evolution

Excess nonsynonymous substitution at shared polymorphic sites among self-incompatibility alleles of Solanaceae

(molecular evolution/shared polymorphism)

ANDREW G. CLARK*† AND TEH-HUI KAO*‡

*Institute of Molecular Evolutionary Genetics and Departments of †Biology and ‡Molecular and Cell Biology, Pennsylvania State University, University Park, PA 16802

Communicated by Michael T. Clegg, August 7, 1991 (received for review April 10, 1991)

ABSTRACT **The function of the self-incompatibility locus (*S* locus) of many plant species dictates that natural selection will favor high levels of protein diversity. Pairwise sequence comparisons between *S* alleles from four species of Solanaceae reveal remarkably high sequence diversity and evidence for shared polymorphism. The level of amino acid constraint was found to be significantly heterogeneous among different regions of the gene, with some regions being highly constrained and others appearing to be virtually unconstrained. In some regions of the protein, there was an excess of nonsynonymous over synonymous substitution, consistent with the strong diversifying selection that must operate on this locus. These hypervariable regions are candidates for the sites that determine functional allelic identity. Simple contingency table tests show that sites that have polymorphism shared between species have more nonsynonymous substitution than polymorphic sites that do not exhibit shared polymorphism. This is consistent with the idea that adaptive evolution favoring amino acid replacement is occurring at sites with shared polymorphism. Tests of clustered polymorphism reveal that an unusually low rate of recombination must be occurring in this locus, allowing very ancient alleles to preserve their identity.**

The *S* locus of the gametophytic self-incompatibility system encodes a pistil-specific protein, the S protein, which functions in the recognition and rejection of pollen bearing the same allele (1). This mechanism ensures that the only successful pollen bears an allele different from either allele in the pistil, so that all plants in a population must be heterozygous at the *S* locus. Although, in principle, a population could survive with only three self-incompatibility alleles, all species examined have a very high number of *S* alleles (2, 3). The recent cloning and sequencing of cDNAs encoding S proteins from several species of Solanaceae have provided the opportunity to test the molecular consequences of the potent selection acting on this locus. Extraordinary sequence divergence among *S* alleles, both within and among species, and sharing of ancient polymorphisms between species have been documented (4).

There is a rich literature on the population genetic theory of self-incompatibility beginning with Wright's (5) analysis of a single gametophytic *S* locus. Wright's conclusion was that such a locus favors the introduction of new alleles because the probability of avoiding a pistil bearing the same allele as the pollen is greatest for rare alleles. This results in a reduced rate of loss of new alleles from a population (compared to neutral alleles), so that an *S* locus will have many more alleles than a neutral locus at equilibrium. Wright also demonstrated that *S* alleles remain in a population much longer than neutral alleles, which should result in *S* alleles accumulating greater sequence diversity than neutral alleles. Maruyama and Nei (6) found that the number of alleles maintained in a finite population with overdominance greatly exceeds the neutral expectation. These results were extended for the specific case of an *S* locus (which can be thought of as a form of overdominance) by Yokoyama and Hetherington (7), who also determined the theoretical sampling distribution of *S* alleles. Recently, the coalescence properties of alleles at a locus undergoing overdominant selection were investigated by Takahata (8) and Takahata and Nei (9), and they found that the structure of a gene genealogy of alleles at an overdominant locus differ from a neutral locus only in having a much longer time scale. Thus, we have formal theoretical grounds for expecting *S* alleles to be highly diverse.

The two goals of this report are to examine the pattern of the rates of synonymous and nonsynonymous substitutions among the *S* alleles and to test for the occurrence of significant clusters of polymorphic sites. Significant heterogeneity in synonymous substitution would indicate that the neutral substitution rate varies across the gene, whereas heterogeneity in nonsynonymous substitution may indicate that selective constraints vary along the gene. These comparisons may provide a means of identifying putative sites that are involved in the functional identity of allelic products (analogous to the antigen recognition site of major histocompatibility complex; ref. 10). The methods of Stephens (11) and of Sawyer (12) are used to identify departures from a null hypothesis of unclustered polymorphisms, and both tests reveal that an unusually low rate of intragenic recombination has occurred.

MATERIALS AND METHODS

Sequences and Alignment. Nucleotide sequences of the 12 *S* alleles examined in this study include S_z (13), S_{2nic} (14), S_{F11} and S_{1nic} (A. Kheyr-Pour and T.-h.K., unpublished data), and S_{3nic} and S_{6nic} (M. A. Anderson and A. E. Clarke, personal communication) from *Nicotiana alata*, an ornamental tobacco species; S_{1pet} and S_{2pet} (Y. Ai and T.-h.K., unpublished data) and S_{3pet} (15) from *Petunia inflata*, a species of wild petunia; S_{2sol} and S_{3sol} from *Solanum chacoense* (16), a wild potato species; and S_{5lyc}, an allele from *Lycopersicon peruvianum* (D.-S. Tsai and T.-h.K., unpublished data), a species of wild tomato. Nucleotide sequence alignments were constructed by first performing an amino acid alignment, then constructing a neighbor-joining tree (17), and finally optimizing the nucleotide alignment to maximize similarity with the algorithm of Lipman *et al.* (18).

Graphical Procedure for Analysis of Synonymous and Nonsynonymous Substitutions. A heuristic graphic approach was taken to observe the substitution events in pairwise comparisons of alleles. For each pair of alleles, a computer program considered a window of 20 codons sliding along the sequences. For each window, the method of Nei and Gojobori (19) was applied to estimate the number of synonymous substitutions per synonymous site (p_s) and the number of

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

nonsynonymous substitutions per nonsynonymous site (p_n). p_s and p_n were averaged among all allele pairs for each window and plotted against the location of the beginning of the window. Such graphs indicate variation across the gene in these proportions and may identify regions with an excess of nonsynonymous substitution. The same procedure was repeated by using the method of Li *et al.* (20) to calculate rates of substitution at nondegenerate, 2-fold, and 4-fold degenerate sites.

Heterogeneity Test for Rates of Synonymous and Nonsynonymous Substitution. One observation from the graphs of p_s and p_n is that these proportions appear to vary widely across the gene. A test of the significance of this heterogeneity was performed as follows. First the variance in p_s and p_n across the gene was calculated, treating each window of 20 adjacent codons as an observation. The null distribution of this variance was then constructed by permuting the sequences and recalculating the variance of the permuted sequences. The shuffling was done such that the *i*th nucleotide in a permutation was the *j*th nucleotide in the original order for all alleles. The observed variance was compared to the distribution of 1000 such variances of the shuffled data. The null hypothesis is that the rates of substitution, both synonymous and nonsynonymous, are homogeneous across the gene.

Test for Excess of Nonsynonymous Substitutions at Shared Polymorphic Sites. In a previous paper, an excess of shared polymorphism (a particular site having the same pair of nucleotides in two or more different species) both at the nucleotide and amino acid levels was established (4). One means of testing the selective neutrality of shared polymorphisms is to ask whether they have a greater chance of being associated with nonsynonymous substitutions than do nonshared polymorphisms. This was tested by first dichotomizing the polymorphic amino acid positions into those that were shared polymorphisms and those that were not. For each group, we then applied method I of Nei and Gojobori (19) to determine the number and proportions of synonymous and nonsynonymous nucleotide substitutions that were represented by these differences. A 2 × 2 contingency table was constructed for counts of shared vs. nonshared polymorphisms and synonymous vs. nonsynonymous substitutions tallied over all pairs of alleles. The null hypothesis that these attributes are independent of one another was tested by a χ^2 statistic.

Tests of Clustered Polymorphic Sites. The tests of Stephens (11) and of Sawyer (12) were applied to these data in order to determine the degree of clustering of polymorphic sites. Both tests use information on the clustering to make inferences about recombination (including classical crossing-over and gene conversion). The test of Stephens (11) examines all ways in which the polymorphic sites partition the alleles into groups and uses combinatorial algebra to calculate the probability of partitions. The null hypothesis is that sites associated with a particular partition are no more clustered than they would be by chance (in the absence of exchange events). The Stephens test makes use of the phylogenetic relations among alleles. The Sawyer test examines all alleles in a pairwise fashion and tests clustering with a statistic, SSCF (sum of the squared lengths of condensed fragments), formed by summing the squared lengths of runs of exact matches. The significance of this test statistic is determined by constructing its null distribution from randomly permuted sequences.

RESULTS

Rates of Nonsynonymous Substitution Are Heterogeneous. Plots of the proportion of synonymous substitutions per synonymous site and nonsynonymous substitutions per nonsynonymous site (p_s and p_n) were constructed for windows

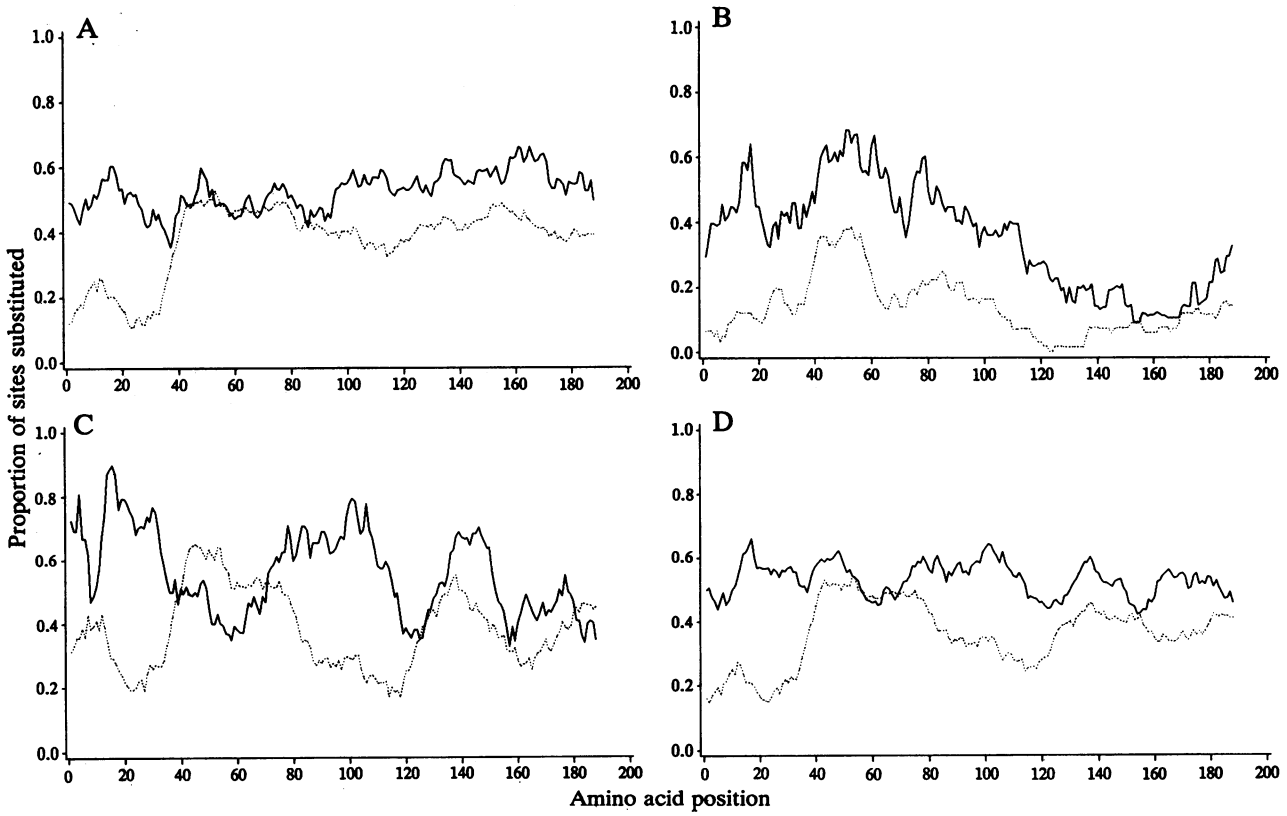

FIG. 1. Proportion of synonymous substitutions per synonymous site (p_s, solid line) and nonsynonymous substitutions per nonsynonymous site (p_n, dotted line) for sliding windows of 20 codons. (*A*) Average of these statistics for all pairwise comparisons of the six *Nicotiana* alleles. (*B*) The three *Petunia* alleles. (*C*) The two *Solanum* alleles. (*D*) All 12 alleles, including 1 tomato allele.

of 20 codons sliding along the sequence alignment for each species and for the entire data set. Fig. 1*A* shows the plot obtained for *Nicotiana*. Note that there is a region from about codons 40–80 in which p_n gets as high, but not higher than, p_s. Elsewhere in the sequence, p_s is substantially greater than p_n, consistent with the hypothesis that selective constraints operating on this part of the protein result in fewer nonsynonymous substitutions. The *Petunia S* alleles are more similar to one another, and the corresponding graph for these alleles gives the appearance that the entire molecule is constrained (Fig. 1*B*). In *Solanum*, the region from codons 40–80 again shows an elevated p_n (Fig. 1*C*). Fig. 1*D* shows the p_s and p_n statistics averaged across all pairwise comparisons (including interspecific comparisons) and yields a pattern rather like that of *Nicotiana* alone. The plots from individual species give the appearance of high heterogeneity in p_n across the gene, motivating the heterogeneity test described below.

Histograms of variance across the gene in p_s and p_n for random permutations of the sequences appear in Fig. 2. The permutations were done in such a way that codons were preserved, but their order across the gene was shuffled. The arrows in Fig. 2 indicate the observed variances of p_s and p_n. Note that in the case of p_s, the observed variance falls within the middle of the range of the permuted distribution, and we conclude that there is no excess variance in observed p_s. p_n, on the other hand, had a variance well above that expected from a random permutation, indicating an excess heterogeneity across the gene.

Shared Polymorphic Sites Have an Excess of Nonsynonymous Substitutions. Contingency tables were constructed for all species pairs for which intraspecific polymorphism data were available and for the complete set of 12 alleles from four species. For each comparison, every codon was classified into a 2 × 2 table according to whether a codon had a shared polymorphism or not and according to whether the substitutions within the respective codon resulted in an amino acid replacement or not. The results, summarized in Table 1, show that in all cases the χ^2 statistic was significant, with a deviation in the direction of shared polymorphisms having an excess proportion of nonsynonymous substitutions. A related analysis following the methods of ref. 20 showed that nondegenerate and 2-fold degenerate sites had a higher rate of substitution if they were associated with a shared polymorphism than if they were not shared.

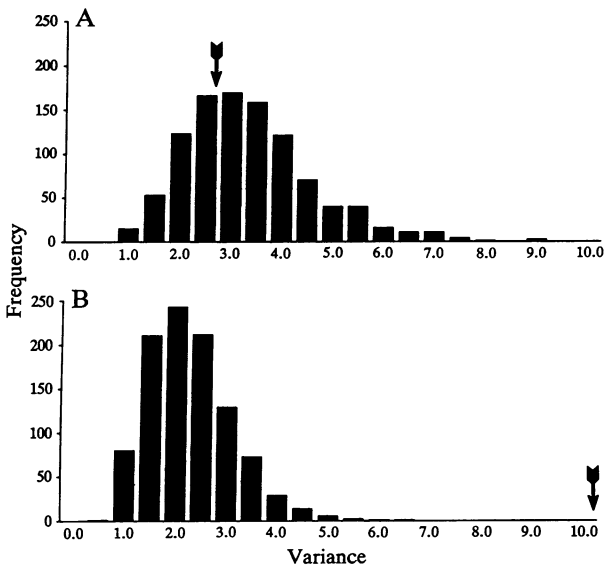

FIG. 2. Histograms of the variance (× 10^3) of p_s (*A*) and p_n (*B*) across the gene obtained for 1000 random permutations of the 12 *S* allele sequences. Arrow in *A* represents the observed variance in p_s and shows that the observed variance is not different from the random expectation. Arrow in *B* shows that the observed variance in p_n is significantly greater than expected because no random permutation had as big a variance.

Differences among species in codon usage may be a complicating factor in comparing the proportions of nonsynonymous substitutions, so tables of codon usage were examined with χ^2 contingency tests (data not shown). Although there were significant codon usage biases, the four species were consistent in the bias (the heterogeneity χ^2 was not significant). Furthermore, the codon usage of the *S* locus was not found to differ significantly from that of the *RBCS* gene. This negative result must be interpreted cautiously because the power of these tests is not as high as those that examine codon usage in large numbers of genes (21).

Tests of Clustered Polymorphic Sites Reveal a Lack of Intragenic Recombination. The Stephens test (11) places polymorphic sites into phylogenetic partitions and determines whether the sites corresponding to each partition are significantly clustered. In the absence of recombination, the test indicates a lack of significant clustering. When the Stephens test (11) was applied to the *Nicotiana* alleles alone, it was found that 41 sites support the partition that puts S_{F11} and S_z together and lumps the remaining alleles in a second group (Table 2). This partitioning is seen as a deep branch in the phylogeny given by Ioerger *et al.* (4). Altogether, the *Nicotiana* alleles had five significantly clustered partitions, none of which was split by a significant gap. The *Petunia* alleles had no significant partitions, but two of the three partitions listed had significant gaps, indicative of cold spots of substitution. When all alleles of the four species were considered together, there were two significant partitions and one significant gap. The observation of significant partitions implies that there are sets of sites that split the alleles into the same groups, and this may be due either to recombination or to certain types of selection or other means of altering the substitution rate locally. When the test was repeated on only the silent sites, which greatly reduces the chance that selection is responsible for the clustering, no significant partitions were found.

The Sawyer test (12) gave very much the same results (Table 3). This test is very sensitive to clustering, and when all polymorphic sites were used, both alleles from individual species and groups of alleles across species yielded test statistics that indicated significant clustering. However, when only silent polymorphic sites were used, none of the tests rejected the null hypothesis. Given that the Sawyer test detects rare recombination among bacterial genes (12), our result suggests that the *S* locus has a remarkably low level of intragenic exchange.

DISCUSSION

The lack of significant heterogeneity across the gene in the synonymous substitution rate is expected under a simple neutral model, but it should be noted that synonymous substitution rates differ among genes within an organism and between organisms (22). The cause of heterogeneity among genes in synonymous substitution rates may be differences in the neutral mutation rate or regional biases in substitution. The fact that homogeneity of p_s was observed across the *S* locus allows us to conclude that the heterogeneity in p_n is due to differences in selective constraints acting on different parts of the S protein. The picture is not as clear as that for the class I major histocompatibility complex molecules, in which the antigen recognition site could be determined from the protein structure. In that case, the antigen recognition site had an excess of nonsynonymous substitutions (10) and an increase in charge profile diversity (23). Our failure to find such a

Table 1. Tests of homogeneity of counts of substitutions among sites that do and do not exhibit shared polymorphism

	Synonymous		Nonsynonymous		
	Shared	Not shared	Shared	Not shared	χ^2
Nicotiana vs. *Petunia*	0.564	0.528	0.459	0.392	29.19***
	(642)	(2414)	(3132)	(8,997)	
Nicotiana vs. *Solanum*	0.526	0.511	0.449	0.376	60.34***
	(1101)	(4132)	(5292)	(15,271)	
Petunia vs. *Solanum*	0.35	0.60	0.57	0.53	3.88*
	(6)	(40)	(47)	(127)	
All 12 alleles	0.525	0.513	0.450	0.376	67.97***
	(1151)	(4329)	(5548)	(15,998)	

Polymorphic codons were first dichotomized into sites that exhibited a shared polymorphism between species and those that did not. Respective codons were then subjected to analysis by method I of Nei and Gojobori (19). Numbers of synonymous and nonsynonymous sites were tallied, as were the numbers of synonymous and nonsynonymous substitutions for all pairwise comparisons of alleles. Table entries give the proportion of sites that were substituted and counts of respective sites (in parentheses) for each category. A χ^2 test was performed on the 2 × 2 table of substitutions to determine whether the counts were independent of whether a site was a shared polymorphism or not. In all cases, there was an excess of nonsynonymous substitution at shared polymorphic sites. *, $P < 0.05$; ***, $P < 0.001$.

dramatic excess of p_n over p_s may be because of the extreme age of these alleles, such that the regions of importance in identifying self may be nearly maximally diverse. Indeed, the proportions of synonymous substitutions per synonymous site are so high as to preclude the use of the Jukes–Cantor formula in estimating rates of substitution.

The analysis of such highly diverse alleles poses its own set of problems. Lewontin (24) argued that estimation of the number of evolutionary events that separate two sequences requires one approach when amino acid sequences are highly conserved and a much simpler approach when amino acid sequences are unconstrained. For intermediate levels of constraint (measured by the difference between synonymous and nonsynonymous rates of substitution), Lewontin suggested that no inferences can be made concerning the number of evolutionary events, because the level of constraint itself must vary in time. The sequence data from *S* alleles fall into this intermediate level of constraint, in part because there is heterogeneity within the gene. Ad hoc tests that dichotomize sites into shared and nonshared polymorphisms have enabled us to extract information from this highly diverse gene without estimating the number of evolutionary events that separate the sequences. By examining numbers of synonymous and nonsynonymous substitutions within and between species, we have constructed a picture of the evolutionary past of this locus.

The test for differences in numbers of substitutions at synonymous and nonsynonymous sites (or nondegenerate, 2-fold, and 4-fold degenerate sites) between sites that exhibit shared polymorphism and those that do not is based on the following scenario. Nonsynonymous mutations that occur in constrained regions of the gene are generally deleterious and will be underrepresented in the data. Because these mutations are deleterious, they reside in the population for a short time and are thus unlikely to be found among the shared polymorphisms. Hence, polymorphisms that are not shared will occur mostly at synonymous sites. Nonsynonymous mutations that are involved in the function of the S protein through pollen recognition are more likely to be maintained as shared polymorphisms, particularly if they result in a new functionally distinct allele. Hence, the comparison of shared

Table 2. Stephens test on all polymorphic sites

Partition	n_{sites}	d_o	$P(d < d_o)$	g_o	$P(g > g_o)$
		Nicotiana			
112222	41	545	0.024	73	0.110
121112	4	126	0.028	84	0.285
111213	3	70	0.034	64	0.140
112221	4	100	0.014	57	0.444
111123	3	65	0.030	62	0.062
		Petunia			
112	38	584	0.281	108	0.018
121	33	581	0.318	125	0.014
123	10	382	0.053	165	0.087
		All alleles (12 alleles, 4 species)			
123333222422	2	2	0.006	2	—
111122111211	2	1	0.003	1	—
112111111111	3	111	0.083	110	0.001

Each partition represents the grouping of alleles that is consistent with the number of sites (n_{sites}) spanned by d_o nucleotides. The digits describing each partition correspond to the alleles in the same order as they appear in *Materials and Methods*. The probability that the sites for each given partition are significantly clustered, $P(d < d_o)$, is obtained as described by Stephens (11). g_o is the maximum number of adjacent sites observed that do not support a particular partition but are embedded within a set of sites that do support the partition. The significance of g_o is given in the column headed $P(g > g_o)$. When the Stephens test is run on only silent polymorphisms (of which there were 39 among the 12 alleles), no significant partitions are found ($P > 0.05$).

Table 3. Sawyer test of clustered polymorphic sites

	n	SSCF	*P*	MCF	*P*
		All polymorphic sites			
Nicotiana	342	9,093	<0.0001	20	0.035
Petunia	148	326	0.083	9	0.017
All alleles	395	60,293	<0.0001	37	0.018
		Silent polymorphic sites			
Nicotiana	44	1,276	0.746	12	0.263
Petunia	42	63	0.908	3	0.697
All alleles	39	6,800	0.825	17	0.723

n is the number of polymorphic sites upon which the test is based. SSCF is the sum of the squared lengths of condensed fragments, where a condensed fragment is a run of exact base matches between a pair of alleles. The sum is taken over all pairwise comparisons within each species and over all alleles (there were not enough alleles to perform the test on *Solanum* or *Lycopersicon*). *P* is the probability of obtaining a larger SSCF by chance and is calculated by randomly permuting the sequences 1000 times. MCF is the maximum length of a condensed fragment, and *P* indicates the fraction of random permutations with a larger MCF. The number of silent sites is smaller when all alleles are included because the additional alleles of other species add gaps to the aligned sequences (see ref. 4), which are not included in the analysis of clusters.

and nonshared polymorphisms is expected to reveal an excess of nonsynonymous substitution among shared polymorphisms, and this is what was found.

The occurrence of recombination is relevant to the observation of excess shared polymorphism. Only when alleles are extremely old can shared polymorphisms be attributable to common ancestry (4), and even low rates of intragenic recombination may obscure the deep branches in the gene genealogy. The sequence diversity in the immediate flanking regions of the *S* gene is remarkably high (C. E. Coleman and T.-h.K., unpublished results), and this observation is consistent with the very low apparent rate of intragenic recombination in this region. While this raises the question of allelism of these sequences, extensive genomic Southern blot analyses continue to support the uniqueness of the *S* gene in these species and show perfect correlation of the allele-specific restriction patterns with respective alleles in genetic crosses (refs. 14 and 25; C. E. Coleman and T.-h.K., unpublished results). The Stephens and Sawyer tests both indicate clustering of nonsynonymous polymorphic sites, but the tests cannot distinguish the effects of recombination from the effects of natural selection favoring diversity in local regions of the gene. The heterogeneity of p_n suggests that such local regions do exist. To avoid confounding inference of recombination with selection, both analyses were repeated on only the silent polymorphic sites. Both tests then fail to reject the null hypothesis of random accumulation of polymorphisms and suggest that an extraordinarily low rate of intragenic recombination is occurring. Further direct tests of suppression of recombination in this region are strongly motivated by these findings.

Analyses of *S* allele sequences have produced results that are consistent with predictions of Wright's classical one-locus population genetic model of gametophytic self-incompatibility (5). In this model, functionally new alleles generated by newly arisen mutations are initially favored over extant alleles because their initial low frequency in the population ensures that pollen of the new allele will be less likely to land on a stigma of the same allele. These functionally new alleles must be the result of nonsynonymous substitutions. Opposing this selectively favored introduction of new variation is the loss of alleles by random genetic drift. The model attains an equilibrium in which many more alleles are maintained in a finite population than could be maintained at a strictly neutral locus. The model therefore predicts an excess of nonsynonymous polymorphism at sites that are involved in the determination of self. Recent theoretical analyses show that the residence times and expected coalescence times of alleles at an *S* locus far exceed those of neutral alleles (9). The inevitable consequence of the extreme age of alleles is that they will have highly divergent sequences. Interestingly, loci involved in mating self-incompatibilities in the fungi *Ustilago* and *Neurospora* also bear highly divergent alleles (26, 27).

The formal mechanism for the rejection of self-pollen must be better understood at the molecular level before we can be too confident of any evolutionary model. Theory suggests that modifier loci must be present to allow invasion of self-incompatibility into a self-compatible species (28), and empirical evidence for this is being marshalled (29). Finally, the consequences on p_s and p_n of selection acting on constrained and diversity-favoring sites in the same gene need to be further characterized.

We thank Marilyn A. Anderson and Adrienne E. Clarke for providing the nucleotide sequences of S_{3nic} and S_{6nic}, and Thomas S. Whittam for insightful discussions. This work was supported by grants from the National Science Foundation (T.-h.K.) and the National Institutes of Health (A.G.C.).

1. Haring, V., Gray, J. E., McClure, B. A., Anderson, M. A. & Clarke, A. E. (1990) *Science* **250**, 937–941.
2. de Nettancourt, D. (1977) *Incompatibility in Angiosperms* (Springer, New York).
3. Ebert, P. R., Anderson, M. A., Bernatzky, R., Altschuler, M. & Clarke, A. E. (1989) *Cell* **56**, 255–262.
4. Ioerger, T. R., Clark, A. G. & Kao, T.-h. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9732–9735.
5. Wright, S. (1939) *Genetics* **24**, 538–552.
6. Maruyama, T. & Nei, M. (1981) *Genetics* **98**, 441–459.
7. Yokoyama, S. & Hetherington, L. (1982) *Heredity* **48**, 299–303.
8. Takahata, N. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2419–2423.
9. Takahata, N. & Nei, M. (1990) *Genetics* **124**, 967–978.
10. Hughes, A. L. & Nei, M. (1988) *Nature (London)* **335**, 167–170.
11. Stephens, J. C. (1985) *Mol. Biol. Evol.* **2**, 539–556.
12. Sawyer, S. (1989) *Mol. Biol. Evol.* **6**, 526–538.
13. Kheyr-Pour, A., Bintrim, S. B., Ioerger, T. R., Remy, R., Hammond, S. A. & Kao, T.-h. (1990) *Sex. Plant Reprod.* **3**, 88–97.
14. Anderson, M. A., McFadden, G. I., Bernatzky, R., Atkinson, A., Orpin, T., Dedman, H., Tregear, G., Fernley, R. & Clarke, A. E. (1989) *Plant Cell* **1**, 483–491.
15. Ai, Y., Singh, A., Coleman, C. E., Ioerger, T. R., Kheyr-Pour, A. & Kao, T.-h. (1990) *Sex. Plant Reprod.* **3**, 130–138.
16. Xu, B., Mu, J., Nevins, D. L., Grun, P. & Kao, T.-h. (1990) *Mol. Gen. Genet.* **224**, 341–346.
17. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
18. Lipman, D. J., Altschul, S. F. & Kececioglu, J. D. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4412–4415.
19. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3**, 418–426.
20. Li, W.-H., Wu, C.-I. & Luo, C.-C. (1985) *Mol. Biol. Evol.* **2**, 150–174.
21. Sharp, P. M., Cowe, E., Higgins, D. G., Shields, D., Wolfe, K. H. & Wright, F. (1988) *Nucleic Acids Res.* **16**, 8207–8211.
22. Wolfe, K. H., Sharp, P. M. & Li, W.-H. (1989) *J. Mol. Evol.* **29**, 208–211.
23. Hughes, A. L., Ota, T. & Nei, M. (1990) *Mol. Biol. Evol.* **7**, 515–524.
24. Lewontin, R. C. (1989) *Mol. Biol. Evol.* **6**, 15–32.
25. Clark, K. R., Okuley, J. J., Collins, P. D. & Sims, T. L. (1990) *Plant Cell* **2**, 815–826.
26. Schulz, B., Banuett, F., Dahl, M., Schlesinger, R., Schaefer, W., Martin, T., Herskowitz, I. & Kahmann, R. (1990) *Cell* **60**, 295–306.
27. Glass, N. L., Sollmer, S. J., Staben, C., Grotelueschen, J., Metzenberg, R. L. & Yanofsky, C. (1988) *Science* **241**, 570–573.
28. Uyenoyama, M. K. (1991) *Genetics* **128**, 453–469.
29. Ai, Y., Kron, E. & Kao, T.-h. (1991) *Mol. Gen. Genet.*, in press.