Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 168–172, January 1999
Evolution

Self-incompatibility alleles from *Physalis*: Implications for historical inference from balanced genetic polymorphisms

ADAM D. RICHMAN*† AND JOSHUA R. KOHN‡

*Biology Department, Montana State University, Bozeman, MT 59717-0346; and ‡ Biology Department, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0116

Edited by M. T. Clegg, University of California, Riverside, CA, and approved November 2, 1998 (received for review July 4, 1998)

ABSTRACT Balanced genetic polymorphism has been proposed as a source from which to infer population history complementary to that of neutral genetic polymorphism, because genetic polymorphism maintained by balancing selection permits inferences about population size over much longer spans of time. However, empirical data for both *S* genes and major histocompatibility complex genes do not fit expectations of coalescent theory. Species-specific gene genealogies have longer terminal branches than expected, indicating an apparent slowdown in the origination of new alleles. Here, we present evidence that divergent *S* alleles were selectively maintained in *Physalis cinerascens* during a reduction in population size, generating longer terminal branches in the *S* gene genealogy relative to the congener *Physalis crassifolia*. Retention of divergent alleles during reduction in the number of alleles violates assumptions of the coalescent model used to estimate effective population size. Recent theoretical and empirical results are consistent with the proposition that nonrandom sorting is a general property of balanced genetic polymorphisms, suggesting that studies of balanced polymorphism that infer the absence of population bottlenecks may overestimate effective population size.

Genetic self-incompatibility presents several advantages for the study of balanced genetic polymorphism. The selective basis for the maintenance of polymorphism at the *S* gene is well understood; the *S* gene prevents self-fertilization, because pollen carrying an allele also expressed in the style is rejected. The gametophytic self-incompatibility system, found for example in the Solanaceae, is particularly mathematically tractable (1) because of the absence of complicating dominance relationships among alleles found in sporophytic systems. The fitness of an *S* allele is negatively frequency-dependent, because low frequency alleles have increased access to compatible mates, and alleles are expected to be maintained in equal frequency. The number of alleles maintained is determined by the equilibrium between the rate of origination and the loss of alleles caused by drift, a function of effective population size (N_e; refs. 2–5). The expectation of equal frequencies makes it straightforward to obtain the likelihood estimate of the number of alleles in the population (6) and the corresponding asymmetric likelihood interval (7). Finally, the inference of the genealogy of *S* alleles is simplified greatly by the apparent absence of recombination at the *S* gene. Both phylogenetic methods and coalescent models commonly assume the absence of recombination. Absence of recombination at the *S* gene is supported by evidence for extreme heterogeneity among different alleles in noncoding regions flanking the *S* gene (8, 9) and by the failure to find evidence of recombination within the *S* gene (10).

We studied the diversity in sequence and in the number of *S* alleles in *Physalis cinerascens* (Solanaceae), a weedy, self-incompatible herb of midwestern North America. *S* allele sequences were obtained by using reverse transcription–PCR. This study was motivated by two previous findings. First, genealogical analysis indicated that the large number of *S* alleles found in the congener *Physalis crassifolia* were of relatively recent origin, apparently the result of a bottleneck event that caused the loss of most *S* alleles in the species and that was followed by population recovery (11). We wished to determine whether *S* allele diversity in other *Physalis* taxa shared the same history of bottleneck and rediversification as a means of estimating the time depth of the bottleneck event. Second, studies of *S* allele diversity in different species have found significant differences in the number of alleles estimated to occur species-wide, and it has been proposed that interspecific variation in life history and its consequent effects on population size and structure determine the number of *S* alleles maintained at equilibrium. For example, a sample of *S* alleles from *P. crassifolia* was shown to contain significantly more *S* alleles than the number estimated from two widely separate population samples of *S* alleles from *Solanum carolinense* (11). *P. cinerascens* is a weedy species similar to *S. carolinense*, raising the question of whether *P. cinerascens* also had a low number of *S* alleles relative to its congener *P. crassifolia*. Here, we report evidence that the bottleneck event predates the divergence of the two *Physalis* species, because *S* alleles in both taxa show rediversification from the same, limited number of *S* allele lineages. Further, analysis of the genealogy of *S* allele sequences indicates that *P. cinerascens* descended from an ancestor with significantly more *S* alleles, similar to its congener at present. We suggest that this loss of *S* alleles was caused by an evolutionary shift in life history to a weedy habit, causing a reduction in species N_e and therefore the number of *S* alleles that can be maintained at equilibrium. Surprisingly, the loss of alleles did not occur at random; more divergent alleles were retained preferentially, resulting in a genealogy with significantly longer terminal branches relative to its congener.

The inference of nonrandom loss of alleles as an explanation for the observation of long terminal branches represents an important insight into the cause of what is apparently a general observation for balanced allele genealogies: terminal branches are longer than expected. The observation of widely divergent alleles at the major histocompatibility complex (MHC) class II *Ab* gene in *Mus musculus* suggested the hypothesis that selection may favor divergent alleles that, in combination, confer an ability to recognize a broader spectrum of antigens (12). Most

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.

© 1999 by The National Academy of Sciences 0027-8424/99/96168-5$2.00/0
PNAS is available online at www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office.
Abbreviations: MHC, major histocompatibility complex; TTBL, total terminal-branch length.
Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF058930–AF058941).
†To whom reprint requests should be addressed. e-mail: arichman@montana.edu.

recently, the observation of extreme allele divergence at the *S* gene motivated statistical comparison of the shape of species-specific *S* gene genealogies to expectations from coalescent theory. *S* allele sequences were found to be significantly more divergent than expected (13). We apply this approach to the MHC *Ab* data for *M. musculus* and find that MHC allele sequences are also more divergent than expected. Recent empirical results suggest that nonrandom assortment is the most plausible explanation for the general observation of terminal branches longer than expected for balanced genetic polymorphisms, with the implication that a coalescent model that assumes random assortment will overestimate N_e.

MATERIALS AND METHODS

Sample Collection. Styles from *P. cinerascens* were collected along roadsides in Bastrop County, TX. *P. cinerascens* exhibits vegetative growth, and sampling was widely spaced to minimize the chance of sampling the same individual twice. A pair of adjacent samples proved to consist of identical, heterozygous genotypes (see below) and were considered to be the same individual. Styles were placed immediately on dry ice and stored at −80°C until analyzed. Sequences from *P. cinerascens* are compared with those of *P. crassifolia* sampled from a 1-km^2 area near Palm Desert, CA (13).

***S* Allele Sequence Acquisition and Analysis.** Isolation of stylar RNA and reverse transcription–PCR using *S* locus-specific primers were performed as described (14). Amplified PCR products contained two alleles which were separated by cloning, identified by restriction fragment length polymorphism analysis, and then sequenced. Different sequences were assumed to encode different *S* allele specificities. This assumption was not examined in *P. cinerascens*, but it has been examined in several other Solanaceae, including *P. crassifolia* (15).

The number of alleles in the population was estimated from the number of different alleles in the sample by using Paxman's maximum likelihood estimator (6). This estimator and the corresponding likelihood interval (7) assume equal allele and genotype frequencies, as expected under negative frequency-dependent selection. The validity of this assumption was examined by using Mantel's test (16).

Phylogeny of *S* allele sequences was estimated by using the NJ (neighbor-joining) algorithm implemented in PHYLIP, version 3.5 (17). DNA sequences were aligned to a previous alignment of *P. crassifolia* sequences, and pairwise distances were calculated by using the Kimura two-parameter model implemented in the program DNADIST in PHYLIP (17). Residues included in the calculation of pairwise distances correspond to the amino acid positions 1–129 (shown in figure 1 in ref. 11), with the following exceptions, for which data corresponding only to positions 1–59 were available: P12, P15, P18, P19, P23, and P24.

Testing for Long Terminal Branches at the *S* Gene by Using Resampling Statistics. We investigated whether the *S* allele genealogy of *P. cinerascens* deviated from a random sample of *P. crassifolia* alleles in terminal-branch length. Tree statistics (i.e., total terminal-branch length, TTBL; total genealogy length; genealogy depth; and coefficient of variation in TTBL) of the *P. cinerascens* sample were compared with random samples from *P. crassifolia* of equivalent size, generated by resampling (Fig. 2). TTBL is the sum of the lengths of all terminal branches. Total length of the genealogy is the sum of all branch lengths. The depth of the genealogy is the sum of the branches from the base to any terminal tip. Bootstrap replicates that sample a single allele from a transspecific lineage (16 of 100 replicates) overestimate terminal-branch length for this allele as the depth of the entire genealogy. To correct for this bias, we used the conservative procedure of substituting the largest value of terminal-branch length observed for that clade

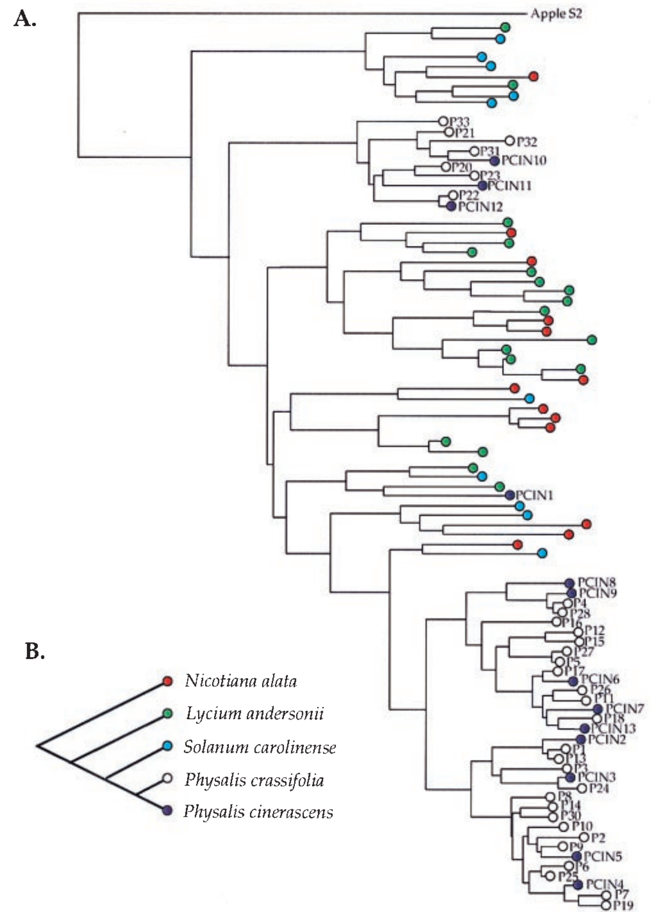

FIG. 1. (*A*) Neighbor-joining topology for *S* allele DNA sequences in the Solanaceae. Citations for published *S* sequences are given in ref. 11, with the exceptions of the *Malus* (apple) sequence (26) and the *Lycium andersonii* sequences (unpublished data). (*B*) Phylogeny of selected genera in the Solanaceae (27).

in all other bootstrap replicates. Statistics were calculated for *P. cinerascens* sequences excluding PCIN1, which does not fall into either of the main clades of *Physalis* alleles (see Fig. 1). Because PCIN1 is relatively divergent from all other *Physalis* alleles, its exclusion is conservative with respect to testing the hypothesis that *P. cinerascens* has long terminal-branch lengths relative to *P. crassifolia*. Trees for resampled *P. crassifolia* sequences were estimated by using the KITSCH algorithm in PHYLIP. The method assumes a molecular clock, and this assumption was examined by using the computer package LINTRE (18). Alleles P12, P15, P18, P19, P23, and P24 (see Fig. 1) were omitted from resampling, because the available partial sequence information would result in biased estimates of branch lengths. Residues 1–59 span a region implicated in determination of allelic specificity. This region shows elevated sequence variation compared with positions 60–129 (19).

Intraspecific vs. Interspecific Comparisons. Values of replacement substitutions per replacement site (d_N) and silent substitutions per silent site (d_S) for pairwise comparisons of *S* alleles within and between species were determined by using the method of Gojobori and Nei (20). For statistical analysis, we compared molecular sequence variation of the closest intraspecific and interspecific pairs of alleles to determine whether these pairs differed in either the proportion of pairwise differences or the relative proportion of replacement vs. silent substitutions. The five closest pairs of interspecific and intraspecific alleles identified by pairwise distances were chosen for analysis. Interspecific pairs: PCIN3, P24; PCIN12,

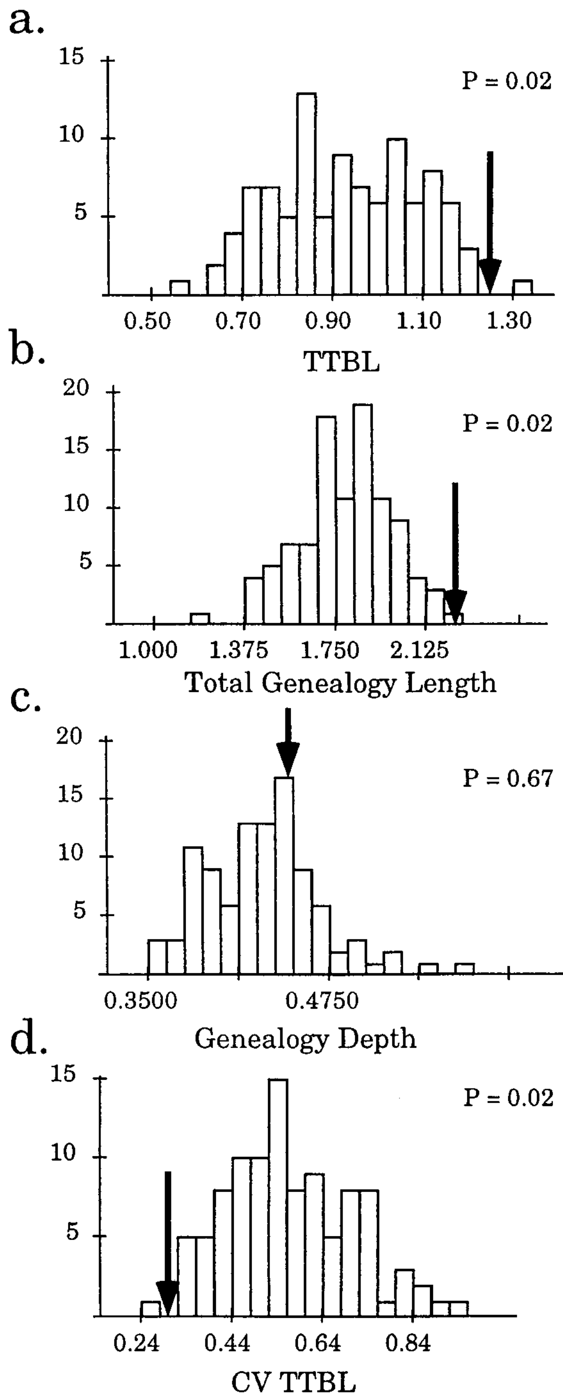

FIG. 2. Frequency distributions for statistics of *S* allele genealogies generated by randomly resampling subsets of *P. crassifolia* sequences. Arrows indicate the value of these statistics for *P. cinerascens*. (*a*) TTBL, the sum of the lengths of all terminal branches. (*b*) Total length of the genealogy, the sum of all branch lengths. (*c*) Depth of the genealogy, the sum of the branches from the base to any terminal tip. (*d*) The coefficient of variation (CV) of TTBL.

P22; PCIN4, P7; PCIN9, P4; and PCIN6, P17. *P. crassifolia* pairs: P1, P13; P7, P19; P6, P25; P5, P27; and P4, P28. *P. cinerascens* pairs: PCIN2, PCIN3; PCIN4, PCIN5; PCIN2, PCIN3; PCIN6, PCIN13; and PCIN10, PCIN12. The number of silent and replacement substitutions and sites for close pairs were determined by using the program SITES (21). d_S values of zero (one intraspecific and one interspecific comparison) were adjusted to 1/(number of silent sites) where necessary to permit calculation of d_N/d_S statistics.

Testing for Long Terminal Branches at an MHC Class II Locus. The fit of the allelic genealogy of 12 *M. musculus Ab* gene exon 2 sequences to theoretical expectation for a balanced allelic genealogy was assessed by using the statistic R_{sd} [$= S(1 - 1/n)/D$], where D is the maximum divergence time (coalescence time of all genes) for the genealogy of n genes, and S is the expected sum of the terminal-branch lengths (13). The statistic is a ratio of total tip length to genealogy depth and has an expected value of 1 for neutrally evolving sequences. Significance was assessed by comparison to results of computer simulation of a balanced genealogy (13). Sequences were obtained from GenBank (accession nos. K01922, M13541, M13540, M13537, M63652, M11357, and M57810–M57815). Pairwise distances were estimated by using the Kimura two-parameter model, and the genealogy was estimated by using the KITSCH algorithm in PHYLIP. The analysis assumes a molecular clock, and this assumption was examined by using LINTRE (18) with *Rattus* MHC class II *Ab* sequences (GenBank accession nos. M57820 and M57821) to root the tree. The assumption of a molecular clock is not rejected for the *Mus Ab* data ($\chi^2 = 15.9$; 11 df; $P > 0.10$).

RESULTS

Phylogeny and Diversity of *S* Alleles. Phylogenetic analysis of *S* allele sequences from both *P. crassifolia* and *P. cinerascens* indicates that the inferred bottleneck event predates the origin of these taxa, because both species show allelic diversification within the same limited number of ancestral *S* lineages (Fig. 1). The *S* alleles sampled from species in other genera of Solanaceae are interspersed throughout the tree, indicating that many *S* allele lineages predate generic divergence, but the alleles in *Physalis* (with the exception of PCIN1) cluster together into just two extensive clades. Despite this recent common history, the sample of *S* alleles from *P. cinerascens* differs from that of its congener in that *S* alleles in *P. cinerascens* are significantly more divergent from each other than are alleles in *P. crassifolia*. The median terminal-branch length of the genealogy of *S* alleles from *P. cinerascens* (excluding PCIN1) is significantly greater than that found in *P. crassifolia* (Mann–Whitney *U* test; $z = 3.6245$; $P = 0.0003$). Differences in evolutionary rate within or between taxa could contribute to this difference in mean pairwise divergence. However, there is no significant deviation from clock-like evolution for the *Physalis* sequences as determined by using LINTRE (ref. 18; $\chi^2 = 47.25$; 37 df; $P > 0.25$).

The observation of significantly longer terminal branches in *P. cinerascens* compared with *P. crassifolia* could result from undersampling *S* alleles from *P. cinerascens*. Recovery of 13 different alleles from 14 diploid individuals gives a maximum likelihood estimate of 15–16 alleles in the population, with a corresponding 95% likelihood interval of 15–21 alleles. The assumption of equal allele and genotype frequencies, expected under gametophytic self-incompatibility, is not rejected for the data ($\chi^2 = 14.7$; 12 df; $P > 0.25$). By comparison, 28 alleles were recovered in a sample of 22 individuals at one locality for *P. crassifolia*, giving an estimate of 43–44 alleles in the population (11). The assumption of equal allele and genotype frequencies was also examined in *P. crassifolia*, where one allele was found to be significantly more common than expected. This bias had little effect on the estimate of the number of *S* alleles in the population when an alternative, less biased estimator was used (11). Thus, the number of *S* alleles in *P. cinerascens* estimated from the population sample is significantly lower than the number found in *P. crassifolia*, indicating that the recovery of fewer alleles is not the result of differences in sampling.

Resampling Statistics. Other populations of *P. cinerascens* could contain additional alleles, which could contribute to interspecific differences in terminal-branch length. To inves-

Table 1. π and d_N/d_S values for the five closest pairs of alleles within *P. crassifolia*, between species, and within *P. cinerascens*

Pairwise comparison	Mean π (SD)	Mean d_N/d_S (SD)
P. crassifolia	0.0393 (0.008)	1.64 (0.89)
P. crassifolia vs. *P. cinerascens*	0.0408 (0.015)	1.81 (1.01)
P. cinerascens	0.2076 (0.069)*	1.80 (0.97)

See *Materials and Methods* for the identity of sequences used for pairwise comparisons.
*Mean value differs from other comparisons at the 0.01 level when using multiple *t* tests with Bonferroni correction.

tigate the effect of this potential source of undersampling, we compared tree-shape statistics calculated from the *S* allele genealogy for *P. cinerascens* to the distribution of these statistics generated by resampling subsets of alleles of equivalent size from the larger sample of *P. crassifolia* alleles (Fig. 2). The *P. cinerascens* sample differs significantly from the resampled data in TTBL (Fig. 2*a*; $P = 0.02$) and total length of the genealogy (Fig. 2*b*; $P = 0.02$) but not in depth of the genealogy (Fig. 2*c*; $P = 0.67$), indicating that the sample of *P. cinerascens* has longer terminal branches but does not differ from resampled genealogies in the sum of lengths of branches internal to the genealogy. Last, the coefficient of variation of TTBL for the *P. cinerascens* sample is significantly lower than expected for the resampled data (Fig. 2*d*; $P = 0.02$) because of the combination of significantly higher TTBL and significantly lower variation in TTBL compared with resampled genealogies. In summary, these results indicate that the significant difference in median terminal-branch length between *Physalis* species is not caused by sampling and that the lengths of terminal branches are much less variable in *P. cinerascens* than in *P. crassifolia*.

The genealogy for *P. cinerascens* has significantly longer TTBL compared with resampled genealogies for *P. crassifolia* but does not differ in genealogy depth, suggesting that the shape of the genealogy as measured by the ratio of TTBL to genealogy depth also differs from resampled genealogies. This ratio is proportional to Uyenoyama's R_{sd} statistic, differing only by a correction for sample size, which is the same for both *P. cinerascens* and the resampled data for *P. crassifolia*. However, calculation of this ratio fails to show a significant difference between *P. cinerascens* and the resampled data from *P. crassifolia* because of variability in the estimate of the depth of resampled genealogies (Fig. 2*c*). Because both *Physalis* taxa share a common history of rediversification from the same transspecific lineages, it follows that the depth of the two genealogies must be identical. When the ratio of TTBL to genealogy depth is calculated from the genealogy obtained for bootstrapped datasets, including both the resampled *P. crassifolia* alleles and the *P. cinerascens* sequences, the value of R_{sd} for *P. cinerascens* is significantly larger than for the resampled sequences from *P. crassifolia* (paired *t* test; $P = 0.04$). Thus, the shape of the genealogy for *P. cinerascens* differs significantly from that obtained for random samples of *P. crassifolia* in having significantly longer terminal branches relative to the depth of the genealogy, and this difference is not caused by differences in sampling.

Molecular Sequence Comparisons. One explanation for the observation of more distantly related alleles in *P. cinerascens* is that *P. crassifolia* has undergone allele diversification since its divergence from *P. cinerascens*, whereas *P. cinerascens* has not. However, multiple *P. cinerascens* sequences form close sister pairs with sequences from *P. crassifolia* (Fig. 1), implying that most alleles in both taxa arose before species divergence. In addition, the origination of new *S* alleles in *P. crassifolia* after species divergence indicates that the most closely related pairs of alleles in *P. crassifolia* will show reduced pairwise divergence compared with the closest interspecific pairs. Comparison of the five closest allele pairs within *P. crassifolia* and the five closest interspecific pairs finds no evidence for the predicted difference in mean pairwise divergence (Table 1). Further, diversification in *P. crassifolia* but not in *P. cinerascens* is expected to alter the ratio of d_N to d_S for within-species comparisons vs. between-species comparisons, particularly for close pairs of alleles. A plot of d_N vs. d_S values for pairwise comparisons within and between species (Fig. 3) shows a similar excess of replacement relative to silent substitutions for both within-species comparisons and between-species comparisons, counter to this expectation. Statistical comparison of interspecific vs. intraspecific closest pairs also finds no evidence for a difference in d_N/d_S values (Table 1). In summary, the absence of interspecific vs. intraspecific differences in either sequence divergence or d_N/d_S ratio indicates that the more limited average divergence of *S* alleles in *P. crassifolia* relative to *P. cinerascens* is not caused by a difference in the amount of origination of new *S* alleles in these taxa since species divergence. Instead, these results suggest that most alleles in both species arose before species divergence and that the observation of significantly more divergent alleles in *P. cinerascens* is caused by the nonrandom loss of alleles during a reduction in the number of alleles.

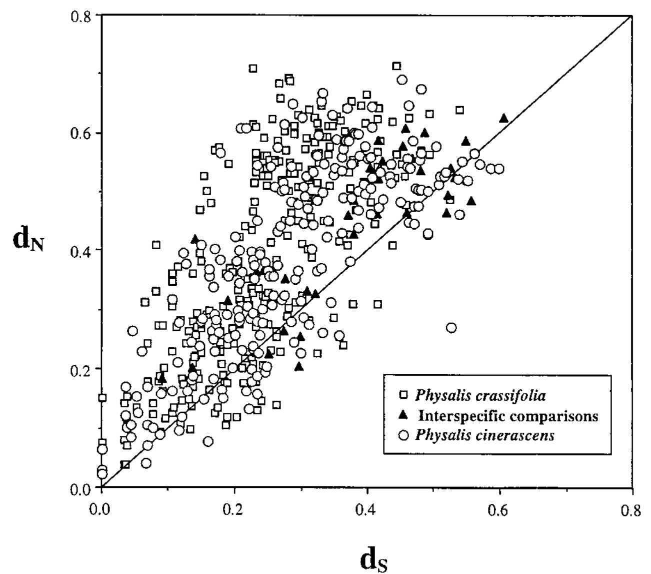

FIG. 3. d_N vs. d_S for pairwise sequence comparisons for *P. cinerascens* (red circles), *P. crassifolia* (black squares), and interspecific pairwise comparisons (blue triangles). Pairwise comparisons between sequences from different transspecific lineages are not shown.

Testing the Fit of an MHC Gene Genealogy to Coalescent Theory. Last, we examined whether the observation of long terminal branches observed for *S* gene genealogies (13) might also hold for other balanced genetic polymorphisms. We examined the fit of data on the MHC class II *Ab* gene in *M. musculus* to that expected under a model of balancing selection. Branch lengths under the assumption of a molecular clock were estimated by using least squares in LINTRE. The *Mus Ab* genealogy deviates significantly from expectation according to Uyenoyama's *R* statistic ($R_{sd} = 4.20$; $P < 0.01$). The value of R_{sd} exceeds 1, indicating that terminal branches are significantly too long relative to the depth of the genealogy. This result supports a previous analysis of these data that concluded that balancing selection alone was insufficient to explain the maintenance of highly divergent MHC allele sequences (12).

DISCUSSION

Genealogical analysis indicates that *S* alleles sampled from *P. cinerascens* and *P. crassifolia* are derived largely from the same, limited number of *S* allele lineages (Fig. 1), indicating that these species share the same history of bottleneck at the *S* gene. However, the *S* allele sequences from *P. cinerascens* are

significantly more divergent from each other than are those from *P. crassifolia*. This difference in divergence is not caused by differences in the extent of sampling among species, because comparison of the *P. cinerascens* genealogy to resampled genealogies from *P. crassifolia* of equivalent size again shows significantly longer terminal branches in *P. cinerascens*. Molecular-sequence analyses indicate that this difference is caused by selective maintenance of more divergent alleles during a reduction in the number of *S* alleles in *P. cinerascens*, rather than by the origination of new *S* alleles in *P. crassifolia* after species divergence. Many alleles in *P. cinerascens* form close sister pairs with alleles from *P. crassifolia*, indicating that most alleles in both species arose before species divergence. Moreover, close pairs of alleles in *P. crassifolia* do not show more limited pairwise divergence or an excess in the proportion of replacement substitutions compared with close interspecific allele pairs, which would be expected if *P. crassifolia* but not *P. cinerascens* had undergone *S* allele diversification after species divergence. Thus, genealogical and molecular-sequence analyses indicate that the reason for higher allelic divergence in *P. cinerascens* relative to its congener is that more divergent alleles were retained selectively in *P. cinerascens* during a reduction in the number of alleles from a diverse ancestor similar to *P. crassifolia*.

Nonrandom assortment of *S* alleles in *P. cinerascens* is consistent with recent models invoking additional genetic mechanisms for the maintenance of divergent allelic specificities at balanced polymorphic loci. A mechanism for divergent-allele advantage has been proposed for MHC genes in which genotypes consisting of relatively divergent alleles have higher fitness because of the ability to present a broader spectrum of potential antigens to the immune system (12). For the *S* gene, it has been proposed that sheltered deleterious mutations accumulating in tight linkage with the obligately heterozygous *S* gene could lead to lowered fitness of individuals carrying closely related *S* alleles caused by inbreeding depression specific to the *S* locus (13). A different possibility is that the ability to discriminate nonself pollen increases as a function of the degree of sequence divergence between alleles present in the pollen and the style. Either of these models predicts that pairs of closely related alleles would have reduced fitness relative to alleles with no close relatives in the population, causing the preferential loss of one or the other member of the pair. Whatever the mechanism, it seems that genetic factors favoring divergent alleles are weak relative to selection promoting *S* allele diversification caused by an increase in N_e. Alleles accumulated in *P. crassifolia*, despite being closely related, whereas nonrandom assortment in *P. cinerascens* was associated with the loss of *S* alleles, apparently caused by reduced population size and thus weakened selection for new *S* allele specificities.

We propose that changes in the number of *S* alleles maintained within species may often be caused by the evolution of life histories that alter N_e and thus the number of alleles that can be maintained. For example, *S. carolinense* is similar to *P. cinerascens* in exhibiting a significantly lower number of alleles than other gametophytic self-incompatible species (14) and marked allele sequence divergence (13). These species also share a number of life-history characteristics suggesting reduced N_e, including short-lived and partially clonal populations inhabiting disturbed sites. By comparison, *P. crassifolia* has a significantly greater number of *S* alleles (11) and neither propagates clonally nor inhabits disturbed sites. Species with low numbers of *S* alleles but extensive allele divergence may be derived secondarily from species with many alleles if more divergent alleles are selectively maintained, rather than from the accumulation of sequence differences in the absence of allelic diversification as previously proposed (13).

Several recent studies interpret the observation of many widely divergent alleles at balanced polymorphic genes as evidence of the absence of population bottlenecks over millions of years (22–24). An important implication of nonrandom assortment of *S* alleles in *P. cinerascens* is that coalescent estimates of population size will be biased upwards if reduction in the number of alleles is accompanied by selective maintenance of divergent alleles. Although evidence for nonrandom assortment presented here is restricted to the *S* gene in the Solanaceae, the mechanism of divergent allele advantage is a strong possibility at MHC loci for biological reasons (12), and the observation of longer than expected terminal branches for an MHC class II gene reported here is consistent with selective maintenance of divergent alleles. Evidence that most allelic diversity at the MHC locus DRB1 in humans has arisen recently (25) from a limited set of more highly divergent lineages that are serotypically distinct seems consistent with a mechanism for divergent allele advantage, where serotypic lineages persist over long periods of time and periods of reduction in population size result in the elimination of within-serotype allelic variation. The distribution of balanced genetic diversity among closely related taxa, in particular in demographic situations where reduction in effective size is expected, provides a means to investigate the mechanism for such deviations at MHC and other balanced polymorphic loci.

We thank Justin Williams for assistance in sample collection and species identification. This study was supported by National Science Foundation Grants DEB 98-70766 (to A.D.R.) and DEB 95-27834 (to J.R.K. and A.D.R.).

1. Vekemans, X. & Slatkin, M. (1994) *Genetics* **137,** 1157–1165.
2. Wright, S. (1960) *Biometrics* **16,** 61–85.
3. Wright, S. (1965) *Evolution* **18,** 609–619.
4. Yokoyama, S. & Hetherington, L. E. (1982) *Heredity* **48,** 299–303.
5. Yokoyama, S. & Nei, M. (1979) *Genetics* **91,** 609–626.
6. Paxman, G. J. (1963) *Genetics* **48,** 1029–1032.
7. O'Donnell, S. & Lawrence, M. J. (1984) *Heredity* **53,** 495–507.
8. 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.
9. Coleman, C. E. & Kao, T. (1992) *Plant Mol. Biol.* **18,** 725–737.
10. Clark, A. G. & Kao, T. H. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9823–9827.
11. Richman, A. D., Uyenoyama, M. K. & Kohn, J. R. (1996) *Science* **273,** 1212–1216.
12. Wakeland, E. K., Boehme, S. & She, J. X. (1990) *Immunol. Rev.* **113,** 207–226.
13. Uyenoyama, M. K. (1997) *Genetics* **147,** 1389–1400.
14. Richman, A. D., Kao, T. H., Schaeffer, S. W. & Uyenoyama, M. K. (1995) *Heredity* **75,** 405–415.
15. Richman, A. D., Uyenoyama, M. K. & Kohn, J. (1996) *Heredity* **76,** 497–505.
16. Mantel, N. (1974) *Biometrics* **30,** 355–362.
17. Felsenstein, J. (1996) PHYLIP, Phylogenetic Inference Package (Dept. of Genetics, Univ. of Washington, Seattle), Version 3.5.
18. Takezaki, N., Rzhetsky, A. & Nei, M. (1995) *Mol. Biol. Evol.* **12,** 823–833.
19. Ioerger, T. R., Gohlke, J. R., Xu, B. & Kao, T. H. (1991) *Sex. Plant Reprod.* **4,** 81–87.
20. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3,** 418–426.
21. Hey, J. & Wakeley, J. (1997) *Genetics* **145,** 833–846.
22. Vincek, V., O'hUigin, C., Satta, Y., Takahata, N., Boag, P. T., Grant, P. R., Grant, B. R. & Klein, J. (1997) *Proc. R. Soc. London Ser. B* **264,** 111–118.
23. Takahata, N. (1993) in *Mechanisms of Molecular Evolution*, eds. Takahata, N. & Clarke, A. G. (Sinauer, Sunderland, MA), pp. 1–21.
24. Klein, D., Ono, H., O'hUigen, C., Vincek, V., Goldschmidt, T. & Klein, J. (1993) *Nature (London)* **364,** 330–334.
25. Bergstrom, T. F., Josefsson, A., Erlich, H. A. & Gyllensten, U. (1998) *Nat. Genet.* **18,** 237–242.
26. Broothaerts, W., Janssens, G. A., Proost, P. & Broekaert, W. F. (1995) *Plant Mol. Biol.* **27,** 499–511.
27. Olmstead, R. G. & Sweere, J. A. (1994) *Syst. Biol.* **43,** 467–481.