

# A 15-Myr-Old Genetic Bottleneck

Timothy Paape,\* Boris Igc,† Stacey D. Smith,‡ Richard Olmstead,§ Lynn Bohs,|| and Joshua R. Kohn\*

\*Section of Ecology, Behavior and Evolution, Department of Biological Sciences, University of California, San Diego; †Department of Biological Sciences, University of Illinois at Chicago; ‡Department of Biology, Duke University; §Department of Biology and Burke Museum, University of Washington; and ||Department of Biology, University of Utah

Balancing selection preserves variation at the self-incompatibility locus (S-locus) of flowering plants for tens of millions of years, making it possible to detect demographic events that occurred prior to the origin of extant species. In contrast to other Solanaceae examined, SI species in the sister genera *Physalis* and *Witheringia* share restricted variation at the S-locus. This restriction is indicative of an ancient bottleneck that occurred in a common ancestor. We sequenced 14 S-alleles from the subtribe Iochrominae, a group that is sister to the clade containing *Physalis* and *Witheringia*. At least 6 ancient S-allele lineages are represented among these alleles, demonstrating that the Iochrominae taxa do not share the restriction in S-locus diversity. Therefore, the bottleneck occurred after the divergence of the Iochrominae from the lineage leading to the most recent common ancestor of *Physalis* and *Witheringia*. Using cpDNA sequences, 3 fossil dates, and a Bayesian-relaxed molecular clock approach, the crown group of Solanaceae was estimated to be 51 Myr old and the restriction of variation at the S-locus occurred 14.0–18.4 Myr before present. These results confirm the great age of polymorphism at the S-locus and the utility of loci under balancing selection for deep historical inference.

## Introduction

Theoretical population genetic studies of balancing selection predict that it will greatly increase the coalescence time of allelic polymorphism relative to neutral variation (Takahata 1990; Vekemans and Slatkin 1994). This prediction has been confirmed by studies of self-recognition loci such as the MHC loci of jawed vertebrates (Klein et al. 1993) and the mating compatibility loci of both fungi (Muirhead et al. 2002) and plants (Ioerger et al. 1990; Richman and Kohn 2000; Castric and Vekemans 2004). In all these systems, the time to coalescence of allelic variation is far older than extant species. Loci under balancing selection can therefore provide evidence of historical genetic and demographic events that far predate current species, a utility that has been termed “molecular paleopopulation biology” (Takahata and Clark 1993).

In many flowering plants, self-incompatibility (SI) systems allow hermaphroditic individuals to recognize and reject their own pollen in favor of pollen from other individuals, thus avoiding the deleterious effects of self-fertilization (de Nettancourt 1977). In single-locus gametophytic SI, as found in the Solanaceae (nightshade family) studied here, a match between the S-allele carried by the haploid pollen grain and either of the S-alleles in the diploid style triggers pollen tube rejection, preventing self-fertilization and also cross-fertilization if the cross-pollen grain carries either allele found in the female parent. In such systems, rare alleles have a selective advantage because they are compatible with more mates (Wright 1939). Selection favoring rare alleles is quite strong, even with large numbers of alleles segregating in populations. For instance, a new pollen S-allele entering a population that already contains 20 alleles has an 11.1% male mating advantage (Clark 1993).

Strong negative frequency-dependent selection is responsible for the 2 outstanding features of S-locus

polymorphism. First, dozens of alleles occur in natural populations with alleles accumulating until a balance is reached between selection favoring rarity and drift causing allele loss (Wright 1939; Lawrence 2000). Second, alleles are often very old because, if any allele drifts toward rarity, selection acts to increase its frequency (Ioerger et al. 1990; Clark 1993). In the Solanaceae, the S-locus gene responsible for self-pollen recognition and rejection in the female tissue is an RNase (S-RNase hereafter; McClure et al. 1989). The great age of polymorphism at the S-locus is exemplified by the fact that S-RNase alleles from the same diploid individual of Solanaceae often differ at more than 50% of their amino acid sites. In addition, S-RNase alleles from species in different genera often cluster together in phylogenetic analyses, evidence of broadly shared ancestral polymorphism (Ioerger et al. 1990; Richman and Kohn 2000; Igc et al. 2004, 2006; Savage and Miller 2006). Much of the S-locus polymorphism found in SI Solanaceae was present in their common ancestor, which must also have been SI (Igc et al. 2004, 2006).

A striking contrast exists between the sequence diversity of S-alleles from species of the closely allied genera *Physalis* and *Witheringia*, and nearly all other Solanaceae, whose S-alleles have been sampled (species of *Brugmansia*, *Lycium*, *Nicotiana*, *Petunia*, and *Solanum*). Although the numbers of S-alleles present in *Physalis* and *Witheringia* species are similar to those found in other Solanaceae (Lawrence 2000; Stone and Pierce 2005; Savage and Miller 2006; Igc et al. 2007), all 93 S-RNases sequenced from 3 *Physalis* (Richman et al. 1996a; Richman and Kohn 1999; Lu 2001) and 2 *Witheringia* (Richman and Kohn 2000; Stone and Pierce 2005) species cluster within only 3 S-allele lineages that predate the divergence of *Physalis* and *Witheringia*. For other Solanaceae, even small samples of alleles usually represent many more ancient lineages (reviewed in Richman and Kohn 2000; Castric and Vekemans 2004; see also Savage and Miller 2006; Igc et al. 2007). This finding has been interpreted as evidence of an ancient bottleneck that restricted variation at the S-locus in some common ancestor of the genera *Physalis* and *Witheringia*. No such restriction is evident at the S-locus of any other sampled SI Solanaceae (Richman

Key words: balancing selection, genetic bottleneck, S-locus, self-incompatibility, shared ancestral polymorphism, Solanaceae.

E-mail: jkohn@ucsd.edu.

*Mol. Biol. Evol.* 25(4):655–663. 2008

doi:10.1093/molbev/msn016

Advance Access publication January 21, 2008

et al. 1996b; Richman 2000; Richman and Kohn 2000; Igic et al. 2004; Stone and Pierce 2005; Igic et al. 2006) except for African species of the genus *Lycium* (Miller et al. 2008), whose S-locus shows evidence of a bottleneck associated with colonization of the Old World from America. After the restriction of S-allele diversity in some common ancestor of *Physalis* and *Witheringia*, the remaining S-allele lineages diversified leaving the observed pattern of large numbers of S-alleles representing only a restricted number of ancient S-allele lineages.

In this paper, we date the historical restriction of S-locus variation common to *Physalis* and *Witheringia*. First, we examine S-locus diversity in the South American monophyletic subtribe Iochrominae, which is found to comprise the sister group of the lineage containing *Physalis* and *Witheringia* by Olmstead et al. (forthcoming). We ask whether the Iochrominae share the reduced set of S-allele lineages found in *Physalis* and *Witheringia*. If so, then the restriction of diversity at the S-locus predates the most recent common ancestor (MRCA) of the group containing the subtribe Iochrominae as well as *Physalis* and *Witheringia*. On the other hand, if the Iochrominae harbor a wide diversity of ancient S-allele lineages, then the restriction at the S-locus must have occurred after the divergence of the Iochrominae from the group containing *Physalis* and *Witheringia* but before the MRCA of *Physalis* and *Witheringia*. We then generated a cpDNA phylogeny of Solanaceae using a fossil-anchored Bayesian-relaxed molecular clock approach, to date the branch along which the bottleneck is shown to have occurred.

## Materials and Methods

### Plant Material and Molecular Procedures

Stylar tissue from 1 to 4 individuals from 7 SI (Smith and Baum 2006) species from the subtribe Iochrominae (*Dunalia brachyacantha* Miers, *Eriolarynx lorentzii* (Dammer) Hunz., *Iochroma australe* Griseb., *Iochroma cyaneum* (Lindl.) M. L. Green, *Iochroma gesnerioides* Miers, *Iochroma loxense* Miers, and *Vassobia breviflora* (Sendt.) Hunz.) was collected from plants growing at the University of Wisconsin greenhouse facility. Smith and Baum (2007) determined these species' SI status through manual self- and cross-pollinations. Seeds for these taxa were acquired largely from the Solanaceae Germplasm collection at Radboud University, Nijmegen, The Netherlands and a few from offspring of wild-collected individuals. Voucher numbers and accession information are given in Smith and Baum (2006, 2007) and in the Supplementary Material online. No large population samples were available for any single species within the Iochrominae. Sampling across species should provide an estimate of S-locus diversity within a group with the caveat that occasionally the same functional S-allele (same specificity) may be sampled from more than one species, making the estimate of the amount of S-allele diversity in the group conservative. Total RNA was extracted and reverse transcription performed to amplify S-alleles according to methods described by Richman et al. (1995), except for the application of 3'-RACE as in Igic et al. (2007). The forward degenerate primer PR1 (5'-GAATTCAYGGNYTNTGG-

CCNGA-3') amplifies from the 5' end of the conserved region C2 (Ioerger et al. 1991) to the 3' end of the coding region of the S-RNase cDNA. Products obtained via PCR were cloned using the TOPO TA Cloning Kit (Invitrogen Corp., Carlsbad, CA) to separate alleles at the obligately heterozygous S-locus. Amplified cloned PCR products were screened by restriction digests (10 clones per individual, on average) and sent for automated sequencing by Eton Bioscience Inc. (San Diego, CA).

### Genealogy of S-alleles from Solanaceae

For phylogenetic analysis of S-RNase sequences from the Iochrominae, additional S-alleles were obtained from GenBank for the following species (number of alleles): *Lycium andersonii* (10), *Nicotiana alata* (6), *Petunia integrifolia* (6), *Physalis cinerascens* (12), *Solanum carolinense* (9), and *Witheringia solanacea* (15) (see Supplementary Material online for GenBank accession numbers). We chose the taxa and allele sequences used for the phylogenetic analysis based on 3 criteria. First, we aimed for broad taxonomic representation across the Solanaceae. Second, S-RNase sequences had to cover at least the entire region between conserved regions 2 and 5 as described by Ioerger et al. (1991). Many sequences in GenBank are shorter and were discarded. Third, in order to apply maximum likelihood (ML) and Bayesian methods without prohibitively long computation times, we reduced the number of sequences used in the final data set by first constructing a Neighbor-Joining tree in PAUP\* v4.0b10 (Swofford 2002) using 71 S-allele sequences from GenBank along with our Iochrominae sequences. We then removed one of any intra-specific sister pair of non-Iochrominae alleles with fewer than 10 amino acid differences. Twelve alleles were removed in this manner. This should not affect our goal of determining the number of ancient S-allele lineages represented among alleles recovered from the Iochrominae. Alleles that arose prior to the divergence of the Iochrominae are unlikely to fall between any very closely related pair of alleles from within a given species.

DNA sequences were manually aligned using BioEdit 7.0.1.4 (Hall 1999) and Se-Al 2.0 (Rambaut 1996) for phylogenetic analysis. Because many S-allele sequences in GenBank do not include the 3' end of the gene, this region was removed from all sequences leaving 354 bp in the final alignment used for phylogenetic analysis. This represents approximately 62% of the coding region of the S-RNase gene including the hypervariable regions most frequently implicated as involved in specificity determination (Ioerger et al. 1991; Savage and Miller 2006; Igic et al. 2007).

We generated an ML tree of S-alleles using PAUP\* v4.0b10 (Swofford 2002). ML model parameters were determined using Modeltest 3.0 (Posada and Crandall 1998). The Akaike Information Criterion (Akaike 1974) best-fit model (TVM + I +  $\Gamma$ ) was used to heuristically search for the ML phylogeny. One S-RNase from *Antirrhinum hispanicum* (Plantaginaceae; Xue et al. 1996) was used as the outgroup. Bootstrap values were generated using a ML heuristic search of 1,000 replicates using the same base frequencies found above to produce a 50% majority rule consensus tree.

**Table 1**  
**Prior Probability and Posterior Distribution Estimates for Calibration of the Solanaceae Phylogeny**

Node constrained	Normal prior distribution Mean, SD, (95% CI)	Posterior distribution Mean, (95% HPD)
tMRCA <i>Physalis</i>	10 Myr, 4.0, (3.4, 16.6)	11.9 Myr, (7.9, 15.8)
tMRCA <i>Solanum</i>	10 Myr, 4.0, (3.4, 16.6)	16.1 Myr, (12.2, 20.6)
tMRCA Convolvulaceae and Solanaceae	52 Myr, 5.2, (43, 60)	62.1 Myr, (54.4, 69.7)

NOTE.—Priors are normal distributions based on fossils of seeds (Benton 1993) as follows: *Physalis*-like seeds from the mid-Miocene, *Solanum*-like seeds from mid to upper Miocene, and *Convolvulus*-like seeds from the lower Eocene. Median values of particular epochs were used as mean dates (in Myr) for calibration points and each was issued an associated SD and 95% confidence interval (CI) that was normally distributed based the upper and lower bounds of the epoch. The HPD values and their means from the 2 combined MCMC runs are reported in the third column. Posterior distribution values are the result of 2 runs of 5 000 000 generations each, sampled every 500th generation. These were combined into 1 log file and resampled at a frequency of every 1 000th generation for a total of 10 000 trees.

We also used MrBayes v3.1.1 (Ronquist and Huelsenbeck 2003) to generate a 50% majority consensus tree for comparison with the ML tree. Bayesian analysis was run using 4 simultaneous Markov chain Monte Carlo (MCMC) chains (3 heated and 1 cold) with a GTR+ $\Gamma$  substitution model across sites. The analysis was run for 1 000 000 generations, sampling every 100th tree for a total of 10 000 trees. After determining stationarity, the initial 2 501 trees were discarded from the burn-in phase. The remaining trees represent generations 250 001 to 1 000 000 (7 500 trees) on which posterior probabilities were calculated.

#### Species Phylogeny and Divergence Time Estimation

The chloroplast sequences used for species divergence time estimation represent a subset of a much larger sample (200 species) of Solanaceae (Olmstead et al. forthcoming). To reduce computational time needed for divergence time estimation, we limited our taxonomic sample to 29 Solanaceae representing only genera from which S-alleles have been sampled, or genera that represent basal nodes in the diversification of the Solanaceae (e.g., *Schizanthus* and *Cestrum* [Olmstead and Sweere 1994; Olmstead et al. forthcoming]) but from which no S-locus information is currently available. An alignment of sequences from 2 chloroplast regions, *ndhF* coding (2 116 bases) and *trnL-trnF* coding, and intergenic spacer sequences (1 377 bases) was used for a combined total of 3 488 bases. For outgroup comparison and root placement, we included *ndhF* and *trnL-trnF* sequence information from 2 species of Convolvulaceae (*Ipomoea batatas* and *Convolvulus arvensis*) which is considered to be the sister family to the Solanaceae (Olmstead and Sweere 1994). Because our small taxonomic sample could lead to erroneous estimation of relationships, we assumed the topological constraints (ordering of generic divergences) found in the larger phylogenetic analysis of Olmstead et al. (forthcoming) which all receive  $\geq 90\%$  bootstrap support.

Likelihood ratio tests (Felsenstein 1988) were used to determine whether sequence data conformed to the expectation of a molecular clock. ML models with and without the enforcement of a clock were performed using PAUP\* on the constrained topology for each gene separately and on the combined data set (both genes). The 2-parameter HKY85 model (Hasegawa et al. 1985) was selected with

a 4-category gamma distribution of rates across sites estimated from the data. Base frequencies, the transition/transversion ratio, and the gamma distribution shape parameter were estimated while running the ML analyses. The test statistic null model settings for each partition correspond to HKY85 + I +  $\Gamma$  + c with the alternative model being HKY85 + I +  $\Gamma$  assuming  $N - 2$  degrees of freedom where  $N$  is the number of terminal sequences. The distribution of likelihood ratio test under the hypothesis  $\Lambda = (-2[\ln_{\text{clock}}/\ln_{\text{without clock}}])$  was assumed to be as a  $\chi^2$ .

Because the data do not conform to a strict molecular clock (see Results), a Bayesian method (Thorne et al. 1998; Thorne and Kishino 2002; Drummond et al. 2006) of relaxing this assumption was used to estimate divergence times among species. The program BEAST v1.4 (Drummond and Rambaut 2007) performs both exponential and lognormal uncorrelated rate estimates of nucleotide substitution along lineages of a phylogeny using a MCMC simulation process. The Bayesian method of Drummond and Rambaut (2007) also allows the user to specify uncertainty in fossil dates using softbound priors which is not possible using likelihood methods of divergence times (Yang 2006). Our prior probability parameters were as follows: we assumed the HKY85 +  $\Gamma$  model of nucleotide substitution with a proportion of invariant sites estimated from the sequence data. We fixed the mean substitution rate at the root node to be 0.0007 substitutions per Myr, consistent with estimated coding and noncoding rates of cpDNA evolution (Palmer 1991; Schnabel and Wendell 1998). We assumed an uncorrelated lognormal relaxed model of rate heterogeneity among branches and a Yule prior model of speciation. The software also allows the user to calibrate specific nodes on the phylogeny to estimated fossil dates along with confidence intervals as priors. We used 2 fossil dates within the Solanaceae (*Solanum*-like and *Physalis*-like seeds from mid-Miocene and a Lower Eocene Convolvulaceae fossil, Benton 1993) as prior constraints of particular nodes (table 1). Based on these fossils, we assumed normally distributed priors of 10 Myr (standard deviation [SD] = 4.0 Myr) for the age of both *Solanum* and *Physalis* and a mean of 52 Myr (SD = 5.2 Myr) for the outgroup (Convolvulaceae) divergence (Magallón et al. 1999). The SDs on the priors represent the upper and lower bounds of the geological epochs from which the fossils were obtained.

We constrained the starting tree and all subsequent trees in the MCMC analysis to conform to the topology

estimated by Olmstead et al. (forthcoming). This preserves species relationships but allows for variation in node heights that translate to ages in millions of years. The MCMC was run twice each for 5 000 000 generations, sampling every 500th tree with a burn-in phase of 500 000 generations for each run. The 2 runs were checked for convergence, and the posterior age distributions of the nodes of interest were analyzed using Tracer v1.3 (Rambaut and Drummond 2004). The estimated node ages for both runs were combined and resampled at a frequency of every 1 000th tree, providing a sample of 10 000 trees. The time between the MRCA of *Physalis* and *Witheringia* and the MRCA of those genera plus the Iochrominae was estimated by subtracting the relevant node ages for each of the 10 000 samples. The results of the MCMC procedure are given as the mean and the 95% highest posterior density (HPD) intervals in millions of years. The mean and SD of the duration of this branch were calculated from these values. Trees from both runs were combined to produce an ultrametric consensus tree using FigTree1.0 (Rambaut 2006). It should be noted that although the Bayesian program MULTIDIV-TIME (<http://statgen.ncsu.edu/thorne/multidivtime.html>) does not allow softbound prior distributions on fossil dates, similar estimates for ingroup divergences were achieved using the above priors. We present the analysis using BEAST because it facilitates the estimation of the duration and associated error of the branch during which the restriction of variation at the S-locus occurred (see below).

## Results

### S-Allele Genealogy

A total of 14 different alleles from 15 individuals from the 7 Iochrominae species were successfully amplified and sequenced. The low number of alleles relative to the number of individuals sampled resulted from 2 causes. First, several individuals shared common alleles. For example, 4 individuals of *Iochroma australis* and 4 of *E. lorentzii* possessed only 3 different alleles per species. Our sample of plants was derived from small germplasm collections that likely contain lower S-locus diversity than would be found in nature. Second, only 1 allele was successfully isolated from 2 individuals.

As found in previous studies, the genealogy of Solanaceae S-alleles shows extensive shared ancestral polymorphism among most species (fig. 1B). The S-alleles of each species of *Petunia*, *Nicotiana*, *Lycium*, and *Solanum* represent 5–7 lineages that arose before the divergence of these genera. This is true even though only a subset of available alleles and species were included to simplify the analysis. In contrast, all alleles from *P. cinerascens* and *W. solanacea* fall within only 3 lineages that predate the MRCA of these 2 genera. Previous studies that incorporated additional S-alleles and species have consistently found the same result (Richman et al. 1996b; Richman and Kohn 2000; Lu 2001; Stone and Pierce 2005).

Despite the limited sampling of Iochrominae alleles, several observations can be made. First, in 2 cases, very similar alleles were recovered from different species of Iochrominae. These close pairs (*E.lor1* and *I.lox2*, *I.cya1* and

*I.ges2*) differ by 2 and 3 amino acid residues, respectively, over the region compared, and may represent sequence divergence within a specificity that arose after species divergence. Therefore, the 14 Iochrominae alleles sampled may represent fewer than 14 specificities. Among this set, we recovered at least 6 ancient Iochrominae S-allele lineages 5 of which diverged from one another prior to the origin of the genus *Solanum* (fig. 1). Alleles from group 1 (fig. 1) are more closely related to alleles from *Solanum* than to other Iochrominae alleles. Given uncertainty in the topology in figure 1, this group of alleles could represent either 1 or 2 S-allele lineages that diverged prior to the origin of *Solanum*. Iochrominae S-allele lineages 2, 3, and 5 are each found to be sister to different S-alleles from *Nicotiana* and group 6 is sister to a pair of alleles from *Petunia* and *Lycium*. Iochrominae S-alleles from group 4 are more closely related to alleles from *Physalis* and *Witheringia* than to alleles from other sampled genera. Only 1 Iochrominae S-allele (*I.aus 2*) falls within any of the 3 clades of alleles found in *Physalis* and *Witheringia*. The placement of that allele is uncertain; it may be sister to all other members of clade I (fig. 1). A basal position for this allele would be consistent with diversification of this clade of alleles in *Physalis* and *Witheringia* after divergence of the Iochrominae. Most S-alleles recovered from Iochrominae fall neither within nor sister to the 3 clades of alleles represented in species of *Physalis* and *Witheringia*.

### Species Phylogeny and Divergence Estimates

Likelihood ratio tests strongly rejected the molecular clock for each chloroplast gene individually ( $\chi^2$  distributions—*ndhF*:  $\Lambda = 2[6867.69 - 6830.92] = 73.53$ ,  $P < 0.001$ , 29 degrees of freedom [df]; *trnLF*:  $\Lambda = 2[4319.25 - 4211.59] = 215.32$ ,  $P < < 0.001$ , 28 df) and for both genes combined (*ndhF* + *trnLF*:  $\Lambda = 2[11354 - 11247.94] = 212.84$ ,  $P < < 0.001$ , 31 df). Using the program BEAST (Drummond and Rambaut 2007) we were able to relax the assumption of a strict molecular clock and determine an approximate time interval during which the bottleneck event occurred. The species phylogeny (fig. 2) shows an estimated crown group age of 51 Myr. The mean estimated age of the MRCA of the Iochrominae, *Physalis*, and *Witheringia* was 18.4 MYA (95% HPD: 12.9, 24.2), whereas the mean estimated age of the MRCA of *Physalis* and *Witheringia* was 13.9 MYA (95% HPD: 9.6, 18.9). The mean difference between these 2 divergence times from 10,000 samples from the Bayesian analyses was 4.5 Myr (SD = 2.10).

## Discussion

Even the relatively small sample of S-alleles from the subtribe Iochrominae shows that they do not share the restriction in the diversity of S-allele lineages found in *Physalis* and *Witheringia*. Instead, they have a diverse set of S-alleles that comprise at least 6 lineages that predate the MRCA of the Iochrominae with *Solanum*. Only 1 S-allele sampled from the Iochrominae groups within any of the 3 S-allele clades found in *Physalis* and

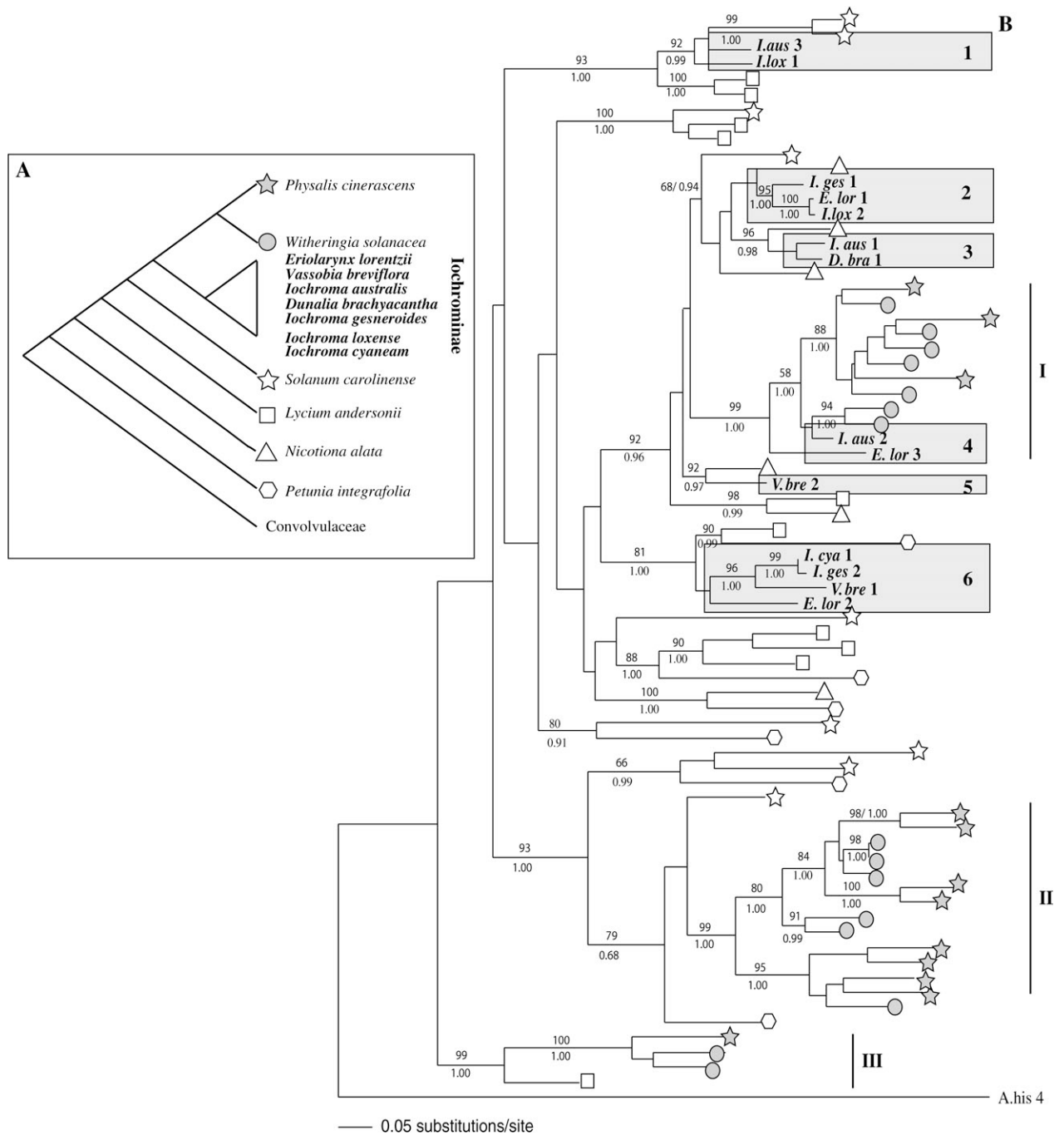


FIG. 1.—ML phylogeny of 72 S-alleles from Solanaceae. Symbols correspond to alleles from taxa in the species phylogeny (inset A). The species phylogeny is redrawn from Olmstead et al. (forthcoming). All nodes in the species phylogeny have >90% bootstrap support. All alleles from the genera *Physalis* and *Witheringia* are restricted to 1 of 3 lineages (indicated by Roman numerals) (B). The 14 alleles from the Iochrominae species (boldface) comprise at least 6 groups (indicated with Arabic numerals and shaded boxes) that predate the Iochrominae. Five of these lineages (all but lineage 4) predate the divergence of *Solanum* from the other genera sampled. The S-allele phylogeny was constructed in PAUP\* v4.0 (Swofford 2002). Bootstrap scores are indicated above branches and posterior probabilities >80% generated by MrBayes v3.0 (Ronquist and Huelsenbeck 2003) are below branches.

*Witheringia* and the position of this S-allele could be basal in that lineage. The diversity of S-allele lineages provides strong evidence that the restriction of S-locus variation common to *Physalis* and *Witheringia* occurred after divergence of the subtribe Iochrominae from the lineage leading to *Physalis* and *Witheringia*. Fossil-calibrated Bayesian-relaxed molecular clock methods estimate the date of the

restriction of diversity at the S-locus to between 14.0 and 18.4 Myr before present.

Our analysis is in remarkable agreement with other evidence concerning the timing of S-locus restriction in *Physalis* and *Witheringia*. Richman (2000) used a lineage-through-time approach to show that diversification within the S-allele lineages found in *Physalis* began

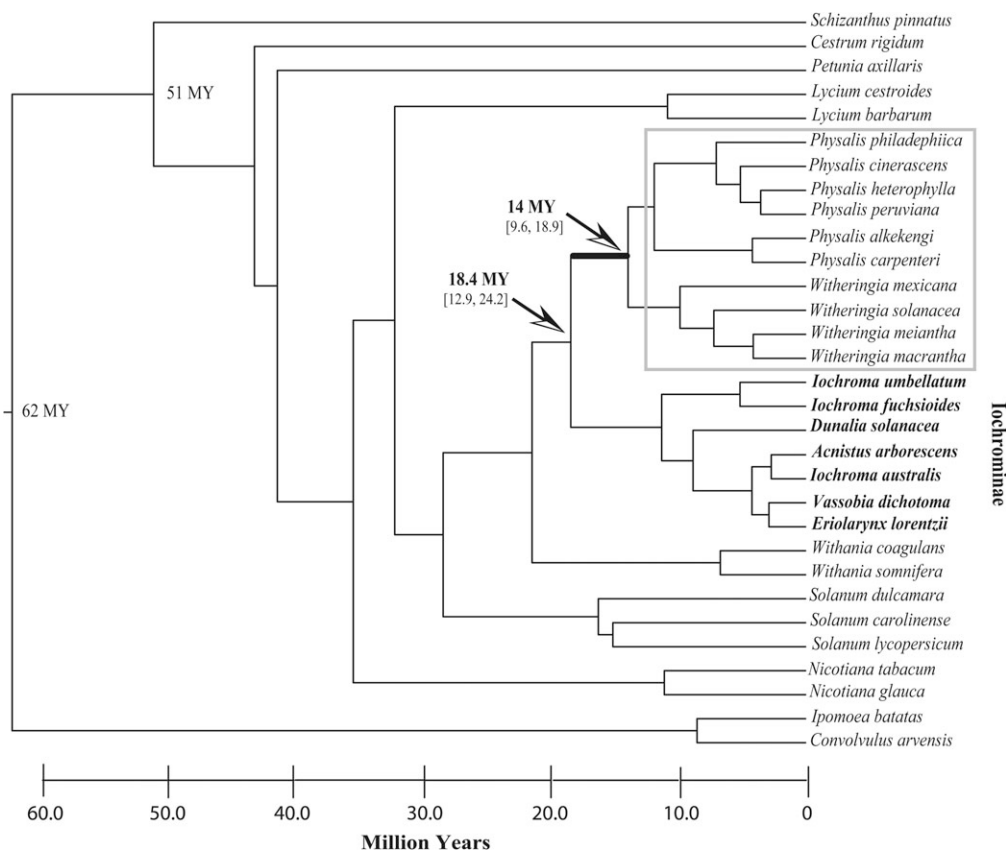


FIG. 2.—Bayesian consensus species phylogeny and divergence time estimates of the Solanaceae based on sequence data from 2 chloroplast genes. The root of the tree was estimated to be 62 Myr (95% HPD: 54.4, 69.7 Myr), and the crown group age was estimated at 51 Myr (95% HPD: 38.6, 63.7) using BEAST v1.4 (Rambaut and Drummond 2007). HPD represents the 95% confidence intervals around the mean in millions of years. Reduction in the diversity of S-lineages in *Physalis* and *Witheringia* (see fig. 1) occurred between the 2 nodes indicated by arrows. The intervening branch is estimated to have a duration of 4.5 (SD = 2.10) Myr.

approximately one-third of the way back from the present to the coalescence of all Solanaceae S-alleles. By comparison, S-alleles drawn from nonbottlenecked Solanaceae show a burst of diversification at the origin of S-allele genealogies, followed by a relative slowdown in diversification toward the present (Uyenoyama 1997; Richman and Kohn 2000; Savage and Miller 2006). Phylogenetic analyses have shown that the common ancestor of the Solanaceae possessed RNase-based SI (Igic and Kohn 2001; Steinbachs and Holsinger 2002; Igic et al. 2004, 2006). Our Bayesian estimate of the time of the MRCA of all extant Solanaceae is 51 MYA (fig. 2), whereas the midpoint of the estimate for the time of the restriction of S-locus diversity is 16.2 MYA, very close to one-third (32%) of the way from the present to the MRCA of extant Solanaceae. Our dating results are also similar to those of Wikström et al. (2001) who used sequence data alone to estimate the age of the divergence of the Solanaceae from the Convolvulaceae as 65 (+/–4) Myr. However, in the absence of fossil data, their estimates of crown group and internal node ages for Solanaceae are somewhat younger than ours.

The timing of the loss of S-locus diversity in *Physalis* and *Witheringia* might be further narrowed down by examining S-allele diversity in the genus *Withania*. *Withania* and its close relatives have been placed sister to the group containing *Physalis* and *Witheringia* in a recent molecular

phylogenetic analysis using nuclear loci (Smith and Baum 2006) but this group is placed sister to the clade containing the Iochrominae plus *Physalis* and *Witheringia* using a much larger taxonomic sample and the cpDNA regions used here (Olmstead et al. forthcoming). Strong (>90%) bootstrap support for alternative topologies in these studies may result from incongruity in the histories of nuclear and chloroplast loci or from differences in taxon sampling. So far, no SI species of *Withania* have been reported (Kaul et al. 2005; Anderson et al. 2006). If one were found and shared the restricted number of S-lineages observed in *Physalis* and *Witheringia*, this would indicate both that *Withania* is more closely related to *Physalis* and *Witheringia* than are the Iochrominae and further constrain the window of time during which the restriction at the S-locus took place.

Restriction of sequence variation, but not S-allele number, in *Physalis* and *Witheringia* has been interpreted as resulting from a population bottleneck that severely reduced the number of S-alleles. Following the bottleneck, diversification of the remaining S-allele lineages restored S-allele numbers (Richman et al. 1996b; Richman 2000; Richman and Kohn 2000; Igic et al. 2004; Stone and Pierce 2005). Severe bottlenecks are required to reduce the number of lineages in a group to only 3. Even a population of constant size 100 is expected to maintain 6 alleles at

equilibrium (Wright 1939). Further loss of alleles after demographic recovery from a bottleneck is unlikely due to the strong selection favoring maintenance of alleles when the population is below the equilibrium allele number. Maintenance of small population size over a protracted period decreases the time to coalescence, increasing the rate of loss of S-allele lineages. However, the size and duration of a population restriction needed to cause substantial turnover of S-allele lineages appears somewhat unrealistic. For instance, over the reasonable range of rates of origination of new S-alleles ( $10^{-6}$  to  $10^{-9}$  per gene per generation; Vekemans and Slatkin 1994), the expected time to coalescence of S-allele variation in a population of constant size 100 ranges from  $3 \times 10^5$  to  $1 \times 10^8$  generations, respectively (Vekemans and Slatkin 1994). Thus, either extremely long periods of reduced population size, or a brief but severe bottleneck, would be needed to cause substantial turnover of S-allele lineages.

A founder event in which extremely few, perhaps only 2, individuals began a new population could have reduced S-allele number to the 3 lineages observed. However, if such an event were the cause, the new population would have had to maintain strict isolation from its source population and give rise to the genera *Physalis* and *Witheringia*. Any subsequent gene flow from the ancestral source population would almost certainly have increased the diversity of S-allele lineages above the 3 observed. In addition, any population founded with only 3 alleles would potentially suffer reproductive losses due to the fact that 1 in 3 potential mates would be incompatible.

Miller et al. (2008) provide the only example of a restriction in S-locus diversity in the Solanaceae outside of *Physalis* and *Witheringia*. In that case, the cause is almost certainly a founder event associated with long-distance dispersal. The genus *Lycium* (Solanaceae) is thought to have originated in South America (Levin and Miller 2005; Levin et al. 2007) but also occurs on several oceanic islands as well as southern Africa. African species form a monophyletic clade nested within the genus suggesting a single colonization event. A sample of S-alleles from African *Lycium* species contains significantly fewer lineages that predate the genus than similar samples of from New World species, suggesting a bottleneck associated with the colonization of Africa.

For *Physalis* and *Witheringia*, several nondemographic phenomena must also be considered that could have reduced the number of old S-allele lineages. First, a common ancestor of these 2 genera might have temporarily lost SI over most or all of its range and then regained it after most S-allele diversity had been lost. Loss of SI is expected to lead to the collapse of S-locus polymorphism because it is no longer maintained by negative frequency-dependent selection (Igic et al. 2008). If all functional S-allele polymorphism is lost, the system cannot be regained because 3 alleles are needed for it to function or else all individuals would be mutually incompatible (Wright 1939). In addition, subsequent to the fixation of a mutation causing self-compatibility (SC), loss-of-function mutations in other genes involved in the SI reaction typically arise (reviewed in Stone 2002; Igic et al. 2008), making the loss of SI essentially irreversible (Igic et al. 2006). However, Rick

and Chetelat (1991) found that offspring of crosses among widely separated SC populations of the otherwise SI *Solanum habrochaites* (formerly known as *Lycopersicon hirsutum*) were restored to SI. Apparently, different mutations caused SC in the 2 populations. In situations such as this, it might be possible for a species to revert to SI if the selective forces acting on SI were to reverse before functional polymorphism at the S-locus was lost and before additional loss-of-function mutations accumulated. An additional possibility of this sort would be the restoration of SI in an SC species through interspecific hybridization. Such a scenario would have to involve extremely few hybridization events, however, because of the strong negative frequency-dependent selection favoring interspecific transit of additional S-alleles.

Selective sweeps reduce variation and have been inferred in resistance loci such as the R-loci of plants (Bergelson et al. 2001) and the MHC of vertebrates (de Groot et al. 2002) thought to normally be subject to some form of balancing selection. For resistance loci, selection might at times be directional due to the prevalence of a certain disease for which one or a few alleles confer resistance. For the S-locus, such a scenario does not seem plausible. S-RNases are expressed only in stylar tissue and have no known function outside of incompatibility. Therefore, conversion of balancing selection to directional selection favoring a particular allele seems unlikely.

Selective sweeps involving loci linked to the site of interest can also cause loss of variation. The S-locus comprises both pollen specificity and stylar specificity-encoding loci, plus at least several other genes in a region of much reduced recombination (Stephan and Langley 1998; Wang et al. 2003; McClure 2006). However, it would be exceedingly difficult for directional selection on a linked locus to work against the strong force of negative frequency-dependent selection acting on the S-locus. Directional selection on a linked locus would have to be strong and linkage extraordinarily tight for one or a few S-allele lineages to diversify and replace all others because of selection favoring a linked gene.

Finally, it is possible that certain S-allele lineages might diversify more rapidly than others leading to the loss of those that diversify more slowly. For instance, fewer amino acid substitutions might be needed to alter the specificity of alleles in some S-allele lineages. This explanation is unlikely, given that the multiple remaining S-allele lineages in *Physalis* and *Witheringia* began diversifying at roughly the same time (Richman 2000; Igic et al. 2004). Uyenoyama (1997, 2003) suggested another factor that might affect the diversification rate of particular S-allele lineages. She noted that enforced heterozygosity at the gametophytic S-locus could shelter deleterious recessive mutations in genes linked to it. When a new allele arises, matings between it and its ancestral allele express this genetic load, lowering the fitness of both alleles until one or the other goes extinct leaving no evidence of diversification. Loss of load linked to certain S-alleles could lead to increased rates of diversification. Some experimental evidence for genetic load linked to the S-locus exists, at least for certain alleles (Stone 2004). However, Uyenoyama (1997) suggests that the loss of load linked to the S-alleles

of an ancestor of *Physalis* and *Witheringia* may itself have been caused by fixation of deleterious recessives held in common by the few alleles remaining after a demographic bottleneck.

No scenario for the reduction in S-allele lineages seen in *Physalis* and *Witheringia* appears particularly persuasive or unequivocal. Whatever caused the loss of S-allele lineages in a common ancestor of *Physalis* and *Witheringia* represents an ancient and rare event during the estimated 50 Myr diversification of the Solanaceae. Among taxa so far sampled, only the African clade of *Lycium* (Miller et al. 2008) suffered a similar restriction. Because of the long duration of polymorphism at this locus, we can infer that no ancestor of any SI Solanaceae whose S-alleles do not show reduced numbers of ancient lineages suffered a historical restriction of S-allele diversity. These results provide strong evidence that events which occurred more than 10 MYA can leave a persistent signature on loci under balancing selection. Inferring the precise cause of this restriction of S-locus diversity appears to be a considerably more difficult problem than documenting and estimating the time of occurrence.

### Supplementary Material

Supplementary material containing the accession numbers of all sequences and plants used are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

### Acknowledgments

Support for this work was provided by National Scientific Foundation DEB-0108173 and DEB-0639984 to J.R.K., DEB-0309184 to B.I. and J.R.K., DEB-0316614 to L.B. and R.O., and DEB-0309310 to S.D.S.

### Literature Cited

- Akaike H. 1974. A new look at the statistical model identification. *IEEE Trans Automat Contr.* 19:716–716.
- Anderson GJ, Bernardello G, Opel MR, Santos-Guerra M, Anderson M. 2006. Reproductive biology of the dioecious Canary Islands endemic *Withania aristata* (Solanaceae). *Am J Bot.* 93:1295–1305.
- Benton MJ. 1993. *The fossil record 2*. London: Chapman & Hall.
- Bergelson J, Kreitman M, Stahl EA, Tian D. 2001. Evolutionary dynamics of plant R-genes. *Science.* 292:2281–2285.
- Castric V, Vekemans X. 2004. Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances. *Mol Ecol.* 13:2873–2889.
- Clark AG. 1993. Evolutionary inferences from molecular characterization of self-incompatibility alleles. In: Takahata N, Clark AG, editors. *Mechanisms of molecular evolution: introduction to molecular paleopopulation biology*. Sunderland (MA): Sinauer Associates. p. 79–108.
- de Groot NG, Otting N, Doxiadis GG, Balla-Jhaghoorsingh SS, Heeney JL, van Rood JJ, Gagneux P, Bontrop RE. 2002. Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. *Proc Natl Acad Sci USA.* 99:11748–11753.
- de Nettancourt D. 1977. *Incompatibility in angiosperms*. Berlin: Springer.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:699–710.
- Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.
- Felsenstein J. 1988. Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet.* 22:521–565.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.
- Hasegawa M, Kishino Y, Yano Y. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol.* 22:160–174.
- Igic B, Bohs L, Kohn JR. 2004. Historical inferences from the self-incompatibility locus. *New Phytol.* 161:97–105.
- Igic B, Bohs L, Kohn JR. 2006. Ancient polymorphism reveals unidirectional breeding system shifts. *Proc Natl Acad Sci USA.* 103:1359–1363.
- Igic B, Kohn JR. 2001. Evolutionary relationships among self-incompatibility RNases. *Proc Natl Acad Sci.* 98:13167–13171.
- Igic B, Lande R, Kohn JR. 2008. Loss of self-incompatibility and its evolutionary consequences. *Int J Plant Sci.* 169:93–104.
- Igic B, Smith WA, Robertson KA, Schaal BA, Kohn JR. 2007. Studies of self-incompatibility in wild tomatoes: I. S-allele diversity in *Solanum chilense* Dun. (Solanaceae). *Heredity.* 99:553–561.
- Ioerger TR, Clark AG, Kao T-h. 1990. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proc Natl Acad Sci USA.* 87:9732–9735.
- Ioerger TR, Gohlke JR, Xu B, Kao T-h. 1991. Primary structural features of the self-incompatibility protein in Solanaceae. *Sex Plant Reprod.* 4:81–87.
- Kaul MK, Kumar A, Sharma A. 2005. Reproductive biology of *Withania somnifera* (L.) Dunal. *Curr Sci.* 1375–1377.
- Klein J, Satta Y, O'hUigin C, Takahata N. 1993. The molecular descent of the major histocompatibility complex. *Annu Rev Immunol.* 11:269–295.
- Lawrence MJ. 2000. Population genetics of the homomorphic self-incompatibility polymorphisms in flowering plants. *Ann Bot.* 85:221–226.
- Levin RA, Miller JS. 2005. Relationships within tribe Lycieae (Solanaceae): paraphyly of *Lycium* and multiple origins of gender dimorphism. *Am J Bot.* 92:2044–2053.
- Levin RA, Shak JR, Miller JS, Bernardello G, Venter AM. 2007. Evolutionary relationships in tribe Lycieae (Solanaceae). *Acta Hort.* 745:225–239.
- Lu Y. 2001. Roles of lineage sorting and phylogenetic relationship in the genetic diversity at the self-incompatibility locus of Solanaceae. *Heredity.* 86:195–205.
- Magallón S, Crane PR, Herendeen PS. 1999. Phylogenetic pattern, diversity, and diversification of eudicots. *Ann Mo Bot Gard.* 86:297–372.
- McClure BA. 2006. New views of S-RNase-based self-incompatibility. *Curr Opin Plant Biol.* 9:639–646.
- McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, Clarke AE. 1989. Style self-incompatibility gene products of *Nicotiana glauca* are ribonucleases. *Nature.* 342:955–957.
- Miller JS, Levin RA, Feliciano NM. Forthcoming. 2008. A tale of two continents: Baker's rule and the maintenance of self-incompatibility in *Lycium* (Solanaceae). *Evolution.*
- Muirhead CA, Glass NA, Slatkin M. 2002. Multilocus self-recognition systems in fungi as a cause of trans-species polymorphism. *Genetics.* 161:633–641.



- Olmstead RG, Bohs L, Abdel Magid H, Santiago-Valentin E, Collier SM, Garcia VF. Forthcoming. A molecular phylogeny of the Solanaceae. *Taxon*.
- Olmstead RG, Sweere JA. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Syst Biol*. 43:467–481.
- Palmer JD. 1991. Plastid chromosome: structure and evolution. In: Bogorad L, Vasil IK, editors. *The molecular biology of plastids*. San Diego (CA): Academic Press. p. 5–53.
- Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics*. 14:817–818.
- Rambaut A. 1996. Se-AI: sequence alignment editor [Internet]. Available from: <http://evolve.zoo.ox.ac.uk/>
- Rambaut A. 2006. FigTree [Internet]. Available from: <http://evolve.zoo.ox.ac.uk/software/figtree>
- Rambaut A, Drummond AJ. 2004. Tracer. Version 1.1 [computer program] [Internet]. Available from: <http://evolve.zoo.ox.ac.uk/software.html?id=tracer>
- Richman AD. 2000. Evolution of balanced genetic polymorphism. *Mol Ecol*. 9:1953–1963.
- Richman AD, Kao T-h, Schaeffer SW, Uyenoyama MK. 1995. S-allele sequence diversity in natural populations of *Solanum carolinense* (Horsenettle). *Heredity*. 75:405–415.
- Richman AD, Kohn JR. 1999. Self-incompatibility alleles from *Physalis*: implications for historical inference from balanced genetic polymorphisms. *Proc Natl Acad Sci USA*. 96:168–172.
- Richman AD, Kohn JR. 2000. Evolutionary genetics of self-incompatibility in the Solanaceae. *Plant Mol Biol*. 42:169–179.
- Richman AD, Uyenoyama MK, Kohn JR. 1996a. S-allele diversity in a natural population of ground cherry *Physalis crassifolia* (Solanaceae) assessed by RT-PCR. *Heredity*. 76:497–505.
- Richman AD, Uyenoyama MK, Kohn JR. 1996b. Allelic diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. *Science*. 273:1212–1216.
- Rick CM, Chetelat R. 1991. The breakdown of self-incompatibility in *Lycopersicon hirsutum*. In: Hawkes L, Nee M, Estrada N, editors. *Solanaceae III: taxonomy, chemistry, evolution*. Richmond (United Kingdom): Royal Botanic Gardens Kew and Linnean Society of London. p. 253–256.
- Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 19:1572–1574.
- Savage AE, Miller JS. 2006. Gametophytic self-incompatibility in *Lycium parishii* (Solanaceae): allelic diversity, genealogical structure, and patterns of molecular evolution. *Heredity*. 96:434–444.
- Schnabel A, Wendell JF. 1998. Cladistic biogeography of *Gleditsia* (Leguminosae) based on *ndhF* and *rpl16* chloroplast gene sequences. *Am J Bot*. 85:1753–1765.
- Smith SD, Baum DA. 2006. Floral diversification and pollination biology for the Andean clade Iochrominae (Solanaceae). *Am J Bot*. 98:1140–1153.
- Smith SD, Baum DA. 2007. Systematics of Iochrominae (Solanaceae): patterns in floral diversity and interspecific crossability. *Acta Hort*. 745:241–254.
- Steinbachs JE, Holsinger KE. 2002. S-RNase-mediated gametophytic self-incompatibility is ancestral in eudicots. *Mol Biol Evol*. 19:825–829.
- Stephan W, Langley CH. 1998. DNA Polymorphism in *Lycopersicon* and crossing-over per physical length. *Genetics*. 150:1585–1593.
- Stone JL. 2002. Molecular mechanisms underlying the breakdown of gametophytic self-incompatibility. *Q Rev Biol*. 77:17–32.
- Stone JL. 2004. Sheltered load associated with S-alleles in *Solanum carolinense*. *Heredity*. 92:335–342.
- Stone JL, Pierce SE. 2005. Rapid recent radiation of S-RNase lineages in *Witheringia solanacea* (Solanaceae). *Heredity*. 94:547–555.
- Swofford DL. 2002. PAUP\*: phylogenetic analysis using parsimony (\*and other methods). Version 4. Sunderland (MA): Sinauer Associates.
- Takahata N. 1990. A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc Natl Acad Sci USA*. 87:2419–2423.
- Takahata N, Clark AG. 1993. Mechanisms of molecular evolution: introduction to molecular paleopopulation biology. Sunderland (MA): Sinauer Associates.
- Thorne JL, Kishino H. 2002. Divergence time estimation and rate evolution with multilocus data sets. *Syst Biol*. 51:689–702.
- Thorne JL, Kishino H, Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol Bio Evol*. 15:1647–1657.
- Uyenoyama MK. 1997. Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics*. 147:1389–1400.
- Uyenoyama MK. 2003. Genealogy dependent viability among self-incompatibility genotypes. *Theor Popul Biol*. 63:281–293.
- Vekemans X, Slatkin M. 1994. Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics*. 137:1157–1165.
- Wang Y, Wang X, McCubbin AG, Kao T-h. 2003. Genetic mapping and molecular characterization of the self-incompatibility (S) locus in *Petunia inflata*. *Plant Mol Biol*. 53:565–580.
- Wikström N, Savolainen V, Chase MW. 2001. Evolution of the angiosperms: calibrating the family tree. *Proc R Soc B*. 268:2211–2220.
- Wright S. 1939. The distribution of self-sterility alleles in populations. *Genetics*. 24:538–552.
- Xue Y, Carpenter R, Dickinson HG, Coen ES. 1996. Origin of allelic diversity in *Antirrhinum* S locus RNases. *Plant Cell*. 8:805–814.
- Yang Z. 2006. *Molecular clock and species divergence times in computational molecular evolution*. Oxford (UK): Oxford University Press. p. 248–258.

Scott Edwards, Associate Editor

Accepted December 28, 2007