Molecular Biology and Evolution

Out of America to Africa or Asia: Inference of Dispersal Histories Using Nuclear and Plastid DNA and the *S-RNase* Self-incompatibility Locus

Jill S. Miller,*‡,[1] Ambika Kamath,[1] Julian Damashek,†,[1] and Rachel A. Levin‡,[1]

[1]Department of Biology, Amherst College

†Present address: Department of Environmental Earth System Science, Stanford University, Stanford, CA 94305.

‡These authors contributed equally to the present study.

***Corresponding author:** E-mail: jsmiller@amherst.edu.

Associate editor: Naoki Takebayashi

Abstract

The plant genus *Lycium* (Solanaceae) originated in the Americas and includes approximately 85 species that are distributed worldwide. The vast majority of Old World species occur in southern Africa and eastern Asia. In this study, we examine biogeographic relationships among Old World species using a phylogenetic approach coupled with molecular evolutionary analyses of the *S-RNase* self-incompatibility gene. The phylogeny inferred from nuclear granule-bound starch synthase I (GBSSI), nuclear conserved ortholog set II (COSII) marker C2_At1g24360, and plastid spacer data (*trnH-pbsA*, $trnD^{GUC}$–$trnT^{GGU}$, *rpl32*–$trnL^{UAG}$, and *ndhF*–*rpl32*) includes a clade of eastern Asian *Lycium* nested within the African species, suggesting initial dispersal from the Americas to Africa, with subsequent dispersal to eastern Asia. Molecular dating estimates suggest that these dispersal events occurred relatively recently, with dispersal from the Americas to Africa approximately 3.64 Ma (95% highest posterior density [HPD]: 1.58–6.27), followed by subsequent dispersal to eastern Asia approximately 1.21 Ma (95% HPD: 0.32–2.42). In accordance, the S-RNase genealogy shows that S-RNases isolated from Old World species are restricted to four lineages, a subset of the 14 lineages including S-RNases isolated from New World *Lycium* species, supporting a bottleneck of *S-RNase* alleles concomitant with a single dispersal event from the Americas to the Old World. Furthermore, the S-RNase genealogy is also consistent with dispersal of *Lycium* from Africa to Asia, as eastern Asian alleles are restricted to a subset of the lineages that also include African alleles. Such a multilocus approach, including complementary data from GBSSI, COSII, plastid spacer regions, and S-RNase, is powerful for understanding dispersal histories of closely related species.

Key words: Bayesian estimation of species trees (BEST), biogeography, *Lycium*, molecular dating, self-incompatibility, Solanaceae, *S-RNase*.

Research article

Introduction

Genetically controlled self-incompatibility (SI) has evolved many times in angiosperms and is an important mechanism to avoid self-fertilization and inbreeding depression. In gametophytic SI, two genes at the *S*-locus control specificity, the pistil-expressed *S-RNase* gene and the pollen-expressed *SLF* (or *SFB*) gene. When the SLF specificity of the haploid pollen grain matches either of the S-RNase specificities in the pistil of the maternal parent, pollen tube growth is terminated; thus, all individuals are heterozygous at the *S*-locus. Strong negative frequency dependence operates within populations, both protecting alleles from extinction and creating strong selection for novel alleles. Balancing selection results in the long-term maintenance (often over tens of millions of years) of large numbers of *S-RNase* alleles within self-incompatible populations (Richman and Kohn 2000; Igic et al. 2004; Savage and Miller 2006). In fact, the time of coalescence of S-RNase lineages often predates the age of the genera from which the *S-RNase* alleles are isolated, resulting in the occurrence of transgeneric allelic lineages within S-RNase genealogies (Ioerger et al. 1990).

Recently, studies have demonstrated the utility of molecular evolutionary analyses of loci under balancing selection for inferring historical demographic events (Bos et al. 2008; Miller et al. 2008; Paape et al. 2008). In Solanaceae, data from the *S-RNase* gene have been used to demonstrate genetic bottlenecks at both the generic (Richman and Kohn 1999; Paape et al. 2008) and species levels (Miller et al. 2008). Miller et al. (2008) documented a reduction in transgeneric allelic diversity at the *S-RNase* locus in southern African *Lycium* (Solanaceae) as compared with New World *Lycium* and interpreted these data as evidence for a genetic bottleneck concomitant with long-distance dispersal from the Americas to the Old World. However, the allelic sampling in Miller et al. (2008) was biased toward sequences isolated from American species (41 and 24 American and African sequences, respectively). Although resampling techniques were employed to attempt to control for unequal allelic sampling, differential inclusion of

alleles could potentially influence the inference of a bottleneck in the *S-RNase* gene en route to Africa.

The genus *Lycium*, with approximately 85 species, is one of the largest and most widely distributed genera in Solanaceae, with centers of diversity in southern South America, southwestern North America, southern Africa, and eastern Asia. *Lycium* are woody shrubs that inhabit mainly subtropical desert or coastal areas. Mating system studies involving controlled crosses have documented the presence of strong SI in several American species (Bianchi et al. 2000; Aguilar and Bernardello 2001; Miller and Venable 2002). Controlled crosses have also confirmed strong SI in four species of southern African *Lycium* (Miller et al. 2008; Miller JS. unpublished data). Furthermore, evolutionary analyses of the *S-RNase* gene have inferred gametophytic SI as the ancestral mating system in both *Lycium* and Solanaceae (Richman and Kohn 2000; Igic et al. 2004; Savage and Miller 2006).

Phylogenetic evidence from the nuclear *granule-bound starch synthase I* (GBSSI) gene and the plastid spacer $trnT^{UGU}$-$trnF^{GAA}$ strongly supports a South American origin of the genus, multiple dispersals between North and South America in the New World, and a single dispersal from the New World (likely South America) to the Old World (Levin and Miller 2005; Levin et al. 2007). However, these genomic regions provide limited resolution of relationships among the approximately 35 species of Old World *Lycium*. Within the Old World, *Lycium* is most species rich in southern Africa (24 species) and eastern Asia (7 species). The historical dispersal pathways of *Lycium* within the Old World remain equivocal, in part due to the limited phylogenetic resolution provided by the genetic markers used thus far, and also as a result of limited sampling of Asian species. Recently, rapidly evolving plastid markers in angiosperms and nuclear conserved ortholog set II (COSII) markers in Solanaceae have shown considerable utility in the resolution of species level relationships within *Lycium* (Levin et al. 2009; Miller et al. 2009). Thus, a phylogenetic approach, which has already proved useful in inferring global biogeographic patterns in *Lycium* (Levin and Miller 2005; Levin et al. 2007), can potentially be extended to resolving fine-scale biogeographic patterns within the Old World, specifically relationships among the African and Asian species.

The overall goal of this study is to infer and compare dispersal patterns suggested by genomic regions subject to different selective pressures: nuclear and plastid DNA data that include both noncoding and coding regions and S-RNase data, which are entirely coding and under strong balancing selection. Within this framework, we have several specific objectives. First, we seek to infer the historical patterns of dispersal in Old World *Lycium*. We incorporate multilocus DNA sequence data from four plastid (*trnH-pbsA*, $trnD^{GUC}$–$trnT^{GGU}$, $rpl32$–$trnL^{UAG}$, and *ndhF*–*rpl32*) and two nuclear (GBSSI and COSII) regions known to be highly variable (Levin et al. 2009; Miller et al. 2009). Previous studies including all African taxa and two Asian species demonstrated the monophyly of Old World *Lycium* and the biogeographical inference that *Lycium* dispersed to the Old World a single time. In the present study, we include data from six of the seven described taxa from eastern Asia (Zhang et al. 1994) and focus on 1) the location of the initial dispersal event to the Old World; specifically, whether *Lycium* initially dispersed from the Americas into eastern Asia (and subsequently into Africa) or initially into Africa (and subsequently into eastern Asia), 2) the frequency of dispersal within the Old World; for example, if *Lycium* dispersed to Africa initially, did subsequent dispersal to Asia occur single or multiple times, and 3) the timing of these dispersal events by inferring divergence times.

Second, we strengthen the molecular evolutionary approach of Miller et al. (2008) by sequencing *S-RNase* alleles from both African and Asian species; these new S-RNase sequence data rectify the New World sampling bias of Miller et al. (2008). The present study includes nearly three times as many Old World as New World sequences (78 vs. 27, respectively). Furthermore, the vast majority of Old World S-RNases (68 of 78 sequences) are newly sequenced in the present study; thus, this data set represents an additional test of the bottleneck for all Old World *Lycium* proposed by Miller et al. (2008). Because this Old World data set includes *S-RNase* sequences from eastern Asian taxa in addition to African taxa, we can also compare the pattern of S-RNase diversity for Old World taxa with evolutionary relationships inferred from the phylogenetic approach. Given the single dispersal event of *Lycium* to the Old World, if the path of dispersal was from the Americas to Africa and then Asia, then we expect that the *S-RNase* sequences retrieved from the eastern Asian taxa will be in the same restricted set of lineages (or a subset of these) as the S-RNases recovered from African species.

Third, we investigate patterns of molecular evolution of eastern Asian S-RNases and compare these to the patterns identified in confirmed self-incompatible populations to infer the presence of SI in the Asian taxa. Although Miller et al. (2008) showed that African *Lycium* species are self-incompatible, to date, there have been no such studies in eastern Asian taxa.

Materials and Methods

Nuclear and Plastid DNA Analyses

Twenty-five individuals were included in this study, including 15 *Lycium* species from the Old World (supplementary table S1, Supplementary Material online). Among the Old World taxa, six of the seven described Chinese species were included (*Lycium chinense*, *L. ruthenicum*, *L. barbarum*, *L. truncatum*, *L. dasystemum*, and *L. yunnanense*). When possible, multiple accessions from the Chinese species were included to test species monophyly. *Nolana werdermannii*, a representative of a closely related genus, was included as an outgroup.

DNA sequence data from two low-copy nuclear regions and four plastid spacer regions were used in this study (supplementary table S1, Supplementary Material online).

These data include previously published data for the nuclear GBSSI gene (Levin and Miller 2005; Levin et al. 2007) and COSII marker COS14 (Sol genomics network marker C2_At1g24360; see also Levin et al. 2009). Four plastid spacer regions were also amplified and sequenced for all taxa: *trnH-pbsA*, $trnD^{GUC}$–$trnT^{GGU}$, *rpl32*–$trnL^{UAG}$, and *ndhF*–*rpl32*. A total of 114 nuclear and plastid sequences were generated for this study.

Standard polymerase chain reaction (PCR) conditions follow those in Levin and Miller (2005) and Levin et al. (2007) for GBSSI, Levin et al. (2009) for COS14, and Miller et al. (2009) for *trnH-pbsA*, $trnD^{GUC}$–$trnT^{GGU}$, *rpl32*–$trnL^{UAG}$, and *ndhF*–*rpl32*. Due to allelic length polymorphisms, amplicons for COS14 and occasionally for GBSSI were cloned prior to sequencing. Individual sequences were edited and assembled using Sequencher 4.7 (Gene Codes, Ann Arbor, MI). The sequences for each genomic accession were manually aligned across individuals and species using Se-Al v2.0a11 (Rambaut 1996–2002).

For each data set (GBSSI, COS14, and plastid spacers combined), substitution model parameters were calculated using the Akaike information criterion in Modeltest v3.7 (Posada and Crandall 1998). All three data sets were analyzed simultaneously using Bayesian estimation of species trees (BEST v2.2; Liu 2008), which estimates a posterior distribution of species trees based on distributions of gene trees. Each locus had its own substitution model, and analyses were run for 50 million generations, with four chains and a temperature of 0.15. A majority rule consensus for the estimated distribution of species trees was constructed in BEST using the sumt command and a burn-in of 10% of the trees. Two BEST analyses were conducted: 1) each individual was defined as a separate "taxon" and 2) multiple individuals from the same species were lumped together and defined as one "taxon." Note that two alleles from one individual of *L. dasystemum* and one individual of *L. ruthenicum* were included and defined as separate taxa in Analysis (1), as these alleles were not monophyletic in preliminary analyses. In Analysis (1), a total of 27 taxa were defined; in Analysis (2), 19 taxa were defined, with seven taxa including multiple individuals (or alleles for *L. dasystemum* and *L. ruthenicum*, see above). The BEST topology from Analysis (2) was used for ancestral area optimization as conducted in Mesquite (Maddison and Maddison 2009). Ancestral geographic areas (South America, North America, Africa, and East Asia) were optimized using parsimony. To test the effect on the inferred biogeographic pattern, area was optimized as both an ordered and unordered character.

A 3692-bp plastid data set (*trnH-pbsA*, $trnD^{GUC}$–$trnT^{GGU}$, *rpl32*–$trnL^{UAG}$, and *ndhF*–*rpl32*) of 19 taxa across tribe Lycieae (see supplementary table S1, Supplementary Material online, and GenBank nos. FJ444876, FJ444889, FJ189609, FJ189612, FJ189627, FJ189633, FJ189636, FJ189651, FJ189671, FJ189674, FJ189689, FJ189738, FJ189741, FJ189756, HM194961, HM194980–HM194982, HM195000–HM195002, and HM195020–HM195022), two species of *Nolana* (FJ189605, FJ189629, FJ189667, and FJ189734; supplementary table S1, Supplementary Material online), three *Solanum* species (AM087200, DQ231562, DQ347958), and *Nicotiana tabacum* (NC001879) were included in a Bayesian analysis using BEAST v1.5.4 (Drummond and Rambaut 2007) to estimate dispersal dates for the ancestor of all Old World *Lycium* and dispersal events within that clade. Substitution model parameters were calculated as described above, and prior probability parameters assumed a GTR + gamma (four categories) model of nucleotide substitution with empirically determined base frequencies. Following Paape et al. (2008), we fixed the mean substitution rate to 0.0007 substitutions/My (similar to previous estimates for plastid data; e.g., Schnabel and Wendel 1998). The analysis used an uncorrelated lognormal relaxed clock model, which is robust to non–clock like data (data deviate somewhat from a strict clock; standard deviation of uncorrelated lognormal relaxed molecular clock = 1.00; Drummond et al. 2007). A Yule model of speciation was used as the tree prior, with a starting tree including branch lengths given from a maximum likelihood (ML) analysis of the data using PAUP* (Swofford 2002) and model parameters estimated by Modeltest. This starting tree was constrained to include a monophyletic Old World lineage given that although these four plastid regions do not contain sufficient variation to robustly resolve a monophyletic clade of Old World *Lycium* species, addition of plastid indels as separate characters (Miller et al. 2009), as well as nuclear GBSSI and COS14 data (see fig. 1A, this paper; Levin et al. 2007) strongly support the monophyly of this clade. Based on a *Solanum*-like seed fossil (Benton 1993), we followed Paape et al. (2008) in employing a normal prior for the time to the most recent common ancestor (MRCA) of *Solanum*, with a mean of 10 My and a standard deviation of 4.0 My. Given knowledge of phylogenetic relationships within Solanaceae (e.g., Olmstead et al. 2008) and *Lycium* (see above), *N. tabacum* was constrained to be the outgroup, the three *Solanum* species were defined as a monophyletic group, and all Old World *Lycium* species were constrained as a monophyletic group.

Two Markov chain Monte Carlo (MCMC) runs were conducted, each for 5,000,000 generations, saving trees every 500 generations. Data were initially analyzed using Tracer v1.5 (Rambaut and Drummond 2003–2009) to ensure convergence. We used LogCombiner v1.5.4 (Drummond and Rambaut 2007) to combine the trees from both runs using a burn-in of 1,000 trees per run. All post–burn-in trees were summarized in a maximum clade credibility tree by TreeAnnotator v1.5.4 (Drummond and Rambaut 2007). This tree with posterior probabilities, estimates of nodal ages, and 95% highest posterior density (HPD) node heights was viewed using FigTree v1.3.1 (Rambaut 2006–2009).

S-RNase Data and Molecular Evolutionary Analyses

We sequenced S-RNases from 14 of the 18 *Lycium* species included in the phylogenetic study, including all 6 eastern Asian species and 7 African taxa. To maximize S-RNase diversity, sequences from an additional African/Mediterranean species, 5 New World *Lycium* species (this study; Savage and Miller 2006; Miller et al. 2008), as well as 19

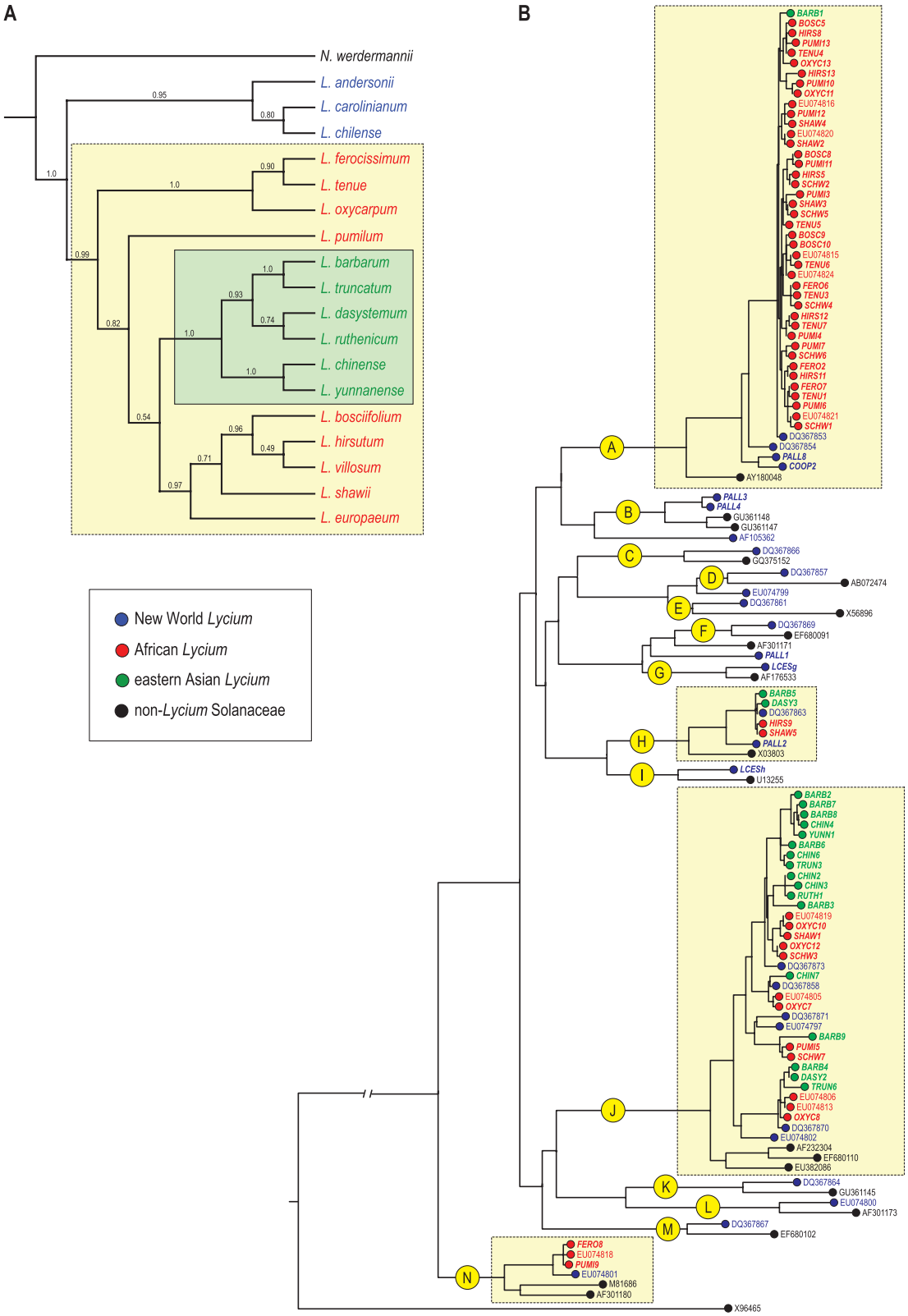

FIG. 1. (*A*) Species relationships inferred from two nuclear regions and four plastid spacer regions using BEST with posterior probabilities indicated above branches. The yellow shading indicates all Old World species and the green shading indicates eastern Asian taxa. (*B*) The most likely S-RNase genealogy, with 14 TGLs for *Lycium* indicated using labeled circles (A–N). All TGLs are supported by Bayesian posterior probabilities $\geq$0.95. For previously generated sequences, GenBank numbers are indicated as taxon labels. In both (*A*) and (*B*), species or *S-RNase* alleles are color coded by provenance.

S-RNases from Solanaceae other than *Lycium* (downloaded from GenBank), were also included (see fig. 1*B*). These 19 S-RNases were identified following BLAST searches (Zhang et al. 2000) using each *Lycium* S-RNase as a query sequence. Following each search, the closest non-*Lycium* sequences were selected. In total, 125 *S-RNase* sequences were included, of which 105 were from *Lycium*; 76 unique *S-RNase* sequences were identified as part of this study (HM195023–HM195098). An *S-RNase* sequence from *Antirrhinum hispanicum* (GenBank X96465) was the outgroup.

Three to ten styles per plant were collected and stored at −20 °C in RNAlater (Ambion, Inc.). Stylar RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Inc.) and cDNA synthesized with the First Strand cDNA Synthesis Kit (EMD Biosciences, Inc.) either using a poly-T primer (see Savage and Miller 2006) or a 3′-rapid amplification of cDNA ends (RACE). When RACE was used in cDNA synthesis, amplifications were carried out with the degenerate primers *PR0* (or *PR1*) and *RaceB*, following Igic et al. (2007). Purified amplification products were cloned using the pSTBlue-1 AccepTor Vector Giga Kit (Novagen, EMD Chemicals Inc.). PCR products were cleaned using ExoSAP, following the protocol in Levin et al. (2009), and sequenced by the Pennsylvania State University Nucleic Acid Facility (University Park, PA) or Retrogen Inc. (San Diego, CA).

Partial *S-RNase* sequences were edited and assembled using Sequencher and manually aligned using Se-Al. Sharing of identical *S-RNase* sequences among closely related taxa occurs in *Lycium* (Miller JS, unpublished data), and several S-RNases sequenced in this study were isolated from multiple species; identical sequences isolated from multiple species were confirmed by PCR of genomic DNA using allele-specific primers, followed by sequencing. A single copy of each of these identical sequences was included in genealogy reconstruction. In two Asian taxa (*L. dasystemum* and *L. yunnanense*) for which we lacked stylar material, lineage-specific primers were used to screen genomic DNA for *S-RNase* sequences. Amplicons were cloned (if necessary) and sequenced as above. For the eastern Asian taxa, all 20 unique *S-RNase* sequences were included in genealogy reconstruction. This reflects sampling of 30 individuals from six species across China, including three to eight individuals per species (one to three populations per species).

A gene tree was inferred using ML as implemented in PAUP*; ML model parameters corresponding to the transversional model + *I* +*G* were determined as above. A Bayesian analysis (MrBayes v3.1.2; Ronquist and Huelsenbeck 2003) was conducted to evaluate support for individual clades. The Bayesian analysis included two independent runs, each with four Markov chains, for 10 million generations, saving trees every 100 generations. MCMC convergence was verified with average standard deviations of split frequencies <0.01 and analysis of the MrBayes output using Tracer v1.5 (Rambaut and Drummond 2003–2009). A majority rule consensus of the estimated distribution of trees was constructed in MrBayes using the sumt command and a burn-in of 10% of the trees.

Average nucleotide and amino acid pairwise distances were calculated in PAUP* for S-RNases from the Old and New World as well as from eastern Asia and southern Africa. To determine whether the pattern of selection expected under the maintenance of SI (i.e., purifying selection on the conserved regions and positive selection on the hypervariable regions of the *S-RNase* gene) is preserved in S-RNases from eastern Asian *Lycium*, we estimated the rate ratio of nonsynonymous to synonymous substitutions (d_N/d_S) for the eastern Asian S-RNases using the *codeml* package in PAML v4.1 (Yang 2007). Specifically, we used likelihood ratio tests to compare null models that assume nearly neutral evolution (M1a and M7) with more complex models that incorporate positive selection (M2a and M8; Yang et al. 2000; Wong et al. 2004). Site-specific d_N/d_S rate ratios were used to determine the number and location of sites under positive selection following significant likelihood ratio tests indicating that models incorporating positive selection explained the data significantly better than null models.

Results

Results of the two BEST analyses were concordant; the topology from the analysis that constrained all individuals within a species to belong to a single terminal taxon (fig. 1*A*) was nearly identical to the topology from the analysis that included individuals as separate terminal taxa (tree not shown). The only difference between these topologies is that where multiple individuals within a species were defined as separate terminal taxa, both individuals of *L. barbarum* were strongly included in a clade of eastern Asian taxa, but *L. barbarum* was not a monophyletic species. The species tree from the BEST analysis (fig. 1*A*) shows strong support for a monophyletic Old World clade (posterior probability = 0.99) and a monophyletic eastern Asian clade (posterior probability = 1.0) nested within the Old World clade.

Results of the BEAST analysis indicate that the mean estimated age of the MRCA of tribe Lycieae is 5.11 Ma (95% HPD: 2.50–8.78). Dispersal to the Old World is estimated at 3.64 Ma (95% HPD: 1.58–6.27). Within the Old World, dispersal to eastern Asia is estimated at 1.21 Ma (95% HPD: 0.32–2.42).

Lycium S-RNases occurred in 14 transgeneric lineages (TGLs), defined as the most recent node including both *Lycium* and non-*Lycium S-RNase* sequences. All 14 of these TGLs included S-RNases isolated from New World species (fig. 1*B*). S-RNases isolated from Old World *Lycium* were present in four TGLs; all four included African S-RNases, but only three included eastern Asian S-RNases. Average pairwise amino acid distances were 0.54 for New World, 0.31 for Old World, 0.31 for African, and 0.29 for eastern Asian S-RNases. Models of positive selection fit the S-RNase data isolated from eastern Asian *Lycium* significantly better than the null models (M1a vs. M2a: LRT = 95.337,

Table 1. Amino Acid Positions for Eastern Asian S-RNases Inferred to be Positively Selected under Model M8 as Implemented in PAML.

S-RNase Region	AA Range	Positively Selected AA Positions
—	1–9	None
Hypervariable region a	10–27	11**, 13**, 15**, 16**, 17**, 20**, 22**, 24*, 27**
—	28–38	None
Hypervariable region b	39–49	40**, 41**, 42**, 43*
—	50–56	None
Conserved region 3	57–62	None
—	63–77	65**
Conserved region 4	78–85	None
—	86–129	100*, 125**

NOTE.—Positions followed by two asterisks identify those sites for which the posterior probability was ≥0.99; one asterisk indicates posterior probabilities ≥0.95. The hypervariable and conserved regions were determined by comparison to Ioerger et al. (1991); dashes indicate regions that have not been formally named.

$P < 0.0001$ and M7 vs. M8: LRT = 93.833, $P < 0.0001$). Approximately 15% of all amino acid sites were found to be under positive selection as estimated by both the M2 and the M8 models, and 81% of the positively selected sites were located in the hypervariable regions of the *S-RNase* gene (table 1).

Discussion

There is strong support for the hypothesis that *Lycium* dispersed initially from the Americas to Africa, followed by a single dispersal into eastern Asia. The eastern Asian taxa are monophyletic (fig. 1A) and sister to a clade that includes southern African taxa. Although the clade including the eastern Asian taxa and *L. bosciifolium–L. europaeum* is only weakly supported (posterior probability = 0.54), it is clear that a clade of southern African species (*L. ferocissimum–L. oxycarpum*) is at the base of the Old World clade. Thus, collapsing the node with 54% support affects neither eastern Asian monophyly nor the pattern that the eastern Asian species are nested within the southern African taxa.

Few studies have estimated the divergence times of clades within Solanaceae, likely due to a limited fossil record. However, Paape et al. (2008) used three fossils (Benton 1993), including a *Solanum*-like seed fossil, to calibrate a Solanaceae-wide phylogeny. Although we followed Paape et al. (2008) in using the same date for *Solanum* based on this *Solanum*-like seed fossil, our age estimates are considerably younger. For example, our estimated age of the MRCA of *Solanum* and *Lycium* is 14.24 Ma (95% HPD: 7.53–21.34), whereas the Paape et al. (2008) estimate is approximately 32 Ma. The older age estimated by Paape et al. (2008) is likely due to differential taxon sampling and inclusion of two additional fossil dates, as Paape et al. (2008) included sampling of numerous genera across Solanaceae as well as two taxa in the distantly related Convolvulaceae. Interestingly, a recent study (Dillon et al. 2009), with similar taxon sampling to that of Paape et al. (2008), yielded age estimates more concordant with those reported in the present study (see chronogram in fig. 3, Dillon et al. 2009). Our estimates also concur with those of the angiosperm-wide study of Wikström et al. (2001), as their estimate of the MRCA of *Solanum* and *Lycium* is 13 Ma (bootstrap estimate of standard error = 3 Ma). Perhaps most notably, our age estimates are robust to both data type (nuclear vs. plastid) and analysis method, given that our estimates concur with those from nuclear GBSSI data that were analyzed using r8s v1.71 (Sanderson 2003) (results not shown). The only other genera in Solanaceae for which divergence dates have been independently estimated are *Nolana* and *Nicotiana*. *Nolana* is estimated to have arisen approximately 4.02 Ma (Dillon et al. 2009), which is within the 95% HPD estimated in the present study (mean age estimate for *Nolana* of 2.08 Ma; 95% HPD: 0.25–5.07). *Nicotiana* is thought to have arisen approximately 15.3 Ma (Clarkson et al. 2005), and this date is also consistent with the present study (mean age estimate of MRCA of *Nicotiana* and *Solanum* is 16.47 Ma; 95% HPD: 8.42–25.35).

Lycium is currently widespread in American deserts, and previous phylogenetic work suggests that tribe Lycieae may have originated in the deserts of South America (Levin and Miller 2005; Levin et al. 2007), which is also the range of its closest relative, *Nolana*. Hartley and Chong (2002) suggest that the Atacama Desert in northern Chile arose in the late Pliocene (ca. 2–4 Ma), and our estimate for the diversification of American *Lycium* (mean age of 3.40 Ma) is consistent with such a recent event.

In a previous analysis of S-RNase diversity in *Lycium*, Miller et al. (2008) included 24 S-RNases from five African species and 41 S-RNases from three New World taxa and reported lower allelic and transgeneric diversity in Africa compared with the New World. However, this result could have been due to greater sampling of S-RNases from New World species. In the present study and in contrast to Miller et al. (2008), our sampling strategy was more powerful with regard to testing a bottleneck for Old World *Lycium*; that is, of the 105 *Lycium S-RNase* sequences in the present analysis, 78 (74%) were isolated from Old World taxa, whereas the remaining 27 (26%) were from New World species. The pattern of species sampling was similar; *S-RNase* sequences were isolated from nearly three times as many Old World species as American species (14 vs. 5, respectively). Thus, our sampling would be expected to recover more allelic (and transgeneric) diversity in the Old World compared with the New World. Despite the inclusion of three times as many S-RNases from the Old World, the S-RNase genealogy shows evidence for a bottleneck at the dispersal of *Lycium* to the Old World. S-RNases isolated from New World species occur in 14 TGLs; those isolated from Old World species were found in only 4 of these 14 TGLs (fig. 1B). We note that the data set in the present study includes only ten Old World S-RNases that were also included in Miller et al. (2008), with the vast majority (68 of 78) of *S-RNase* sequences newly generated. Thus, to date, there are 92 S-RNases (68 sequences in the present study plus 24 from Miller et al. 2008) known from

Old World taxa (19 species), all of which are restricted to four TGLs.

Although there was considerable sampling in eastern Asia (i.e., both species and geographic diversity were high), Asian S-RNases were limited to just three of the four TGLs that also included African S-RNases. That *S-RNase* sequences from eastern Asia are restricted to a subset of those recovered for African taxa is consistent with the phylogenetic result obtained here that dispersal progressed from Africa to eastern Asia (fig. 1*A*). The alternative biogeographic possibility is that the initial dispersal event was from the New World to Asia and then subsequently to Africa and that allelic turnover in Asia is responsible for the reduced number of S-RNase lineages in Asian *Lycium*. This alternative deserves attention, although it is unlikely for several reasons. First, as has been pointed out by numerous authors (Wright 1939; Lawrence 2000; Igic et al. 2008; Paape et al. 2008), the loss of alleles (let alone entire lineages of alleles) is exceedingly rare given that the *S-RNase* locus is under strong negative frequency–dependent selection. Second, in a previous paper (Miller et al. 2008), we sampled 15 individuals from five species of African *Lycium* and recovered 24 *S-RNase* sequences; these 24 S-RNases were included in four ancient TGLs. In the present study, we sample twice as many individuals ($n = 30$ individuals) from six eastern Asian species and recover 20 unique *S-RNase* sequences; these 20 Asian S-RNases are included in three of the four lineages described previously and recovered here (lineages A, H, J, and N in fig. 1*B*). Thus, our sampling is conservative on this point; despite twice the sampling effort in Asia, including increased species diversity and coverage over a wider geographic range (compared with Africa), not only did we fail to recover the fourth TGL in Asia, but, critically, we recovered no novel S-RNase lineage diversity in Asia.

Site-specific molecular evolutionary analyses (as implemented in PAML) have proven useful for studying the molecular evolution of the *S-RNase* mating system gene (Takebayashi et al. 2003; Savage and Miller 2006; Igic et al. 2007; Vieira et al. 2007; Miller et al. 2008). Indeed, *S-RNase* alleles from populations with functional SI have a characteristic molecular signature. For example, Savage and Miller (2006) documented positive selection for amino acid positions in hypervariable regions and purifying selection in conserved regions of the *S-RNase* gene for self-incompatible *Lycium parishii* (Miller and Venable 2002). Likewise, Igic et al. (2007) used site-specific molecular analyses, in conjunction with testcrosses, to document a similar pattern of S-RNase molecular evolution in *Solanum chilense*. At larger scales, site-specific methods have been used to document the locations of positively selected sites at the *S-RNase* locus across three plant families (Plantaginaceae, Rosaceae, and Solanaceae; Vieira et al. 2007) and for sets of related species (see Miller et al. 2008). Models of evolution incorporating positive selection fit the S-RNase data from Asian *Lycium* significantly better than models of nearly neutral evolution. Furthermore, positive selection was found to act overwhelmingly on amino acid sites in regions of the *S-RNase* gene controlling allele specificity (hypervariable regions, table 1) and not on sites in regions maintaining protein stability (conserved regions). In fact, at 11 of the 14 amino acid positions in the conserved regions (table 1), the d_N/d_S rate ratio was estimated to be less than 0.32; rate ratios for the other three regions ranged from 0.90 to 0.98. These data indicate strong purifying selection at these positions, as expected if the *S-RNase* gene experienced constraints on function. Thus, despite low transgeneric diversity, SI appears to have been maintained in eastern Asian *Lycium*.

Although SI has likely been maintained in eastern Asian *Lycium*, the reduction in allelic diversity is notable. In particular, a majority of the Asian S-RNases occurred in a single TGL (TGL "J," fig. 1*B*); however, this is not representative of the frequency of allele recovery, given that many species shared identical S-RNases. For example, the "BARB1" (lineage A in fig. 1*B*) and "BARB6" (lineage J) alleles were isolated from four and three Asian species, respectively. Although the Asian species are relatively young (1.21 Ma [95% HPD: 0.32–2.42]), this sequence similarity is notable. We sequenced approximately 70% of the *S-RNase* gene, such that differences in these identical sequences may occur outside of the regions analyzed here. However, this seems unlikely, especially given that comparisons of the intronic sequences for several of the shared sets of alleles were also identical. Future investigations focused on obtaining full-length sequences for individuals with identical S-RNases could address this issue. It would also be interesting to isolate and compare the pollen-expressed *SLF* sequences in these individuals, especially as Surbanovski et al. (2007) have reported identical *S-RNase* alleles, but divergent pollen-expressed *SFB* sequences in *Prunus*. These authors argue that the pollen-expressed *SFB* gene is more tolerant of mutations and may play a role in the evolution of new allelic specificities (for discussion of the evolution of novel specificities, see Uyenoyama et al. 2001; Newbigin and Uyenoyama 2005).

Sharing of identical or nearly identical *S-RNase* sequences (sometimes even including introns) among species has been reported in *Prunus* (Surbanovski et al. 2007; Sutherland et al. 2008), *Physalis* (Lu 2001), and *Lycium* (Savage and Miller 2006). Lu (2001) and Sutherland et al. (2008) suggest that sharing of highly similar or identical alleles (particularly for intronic regions) could indicate recent species divergence, such that there is incomplete lineage sorting. For maintenance of these similar alleles across species, Lu (2001) suggests that there must also be an absence of selection pressure for differentiation. Consequently, allele sharing should be less likely between distantly related species. Indeed, Surbanovski et al. (2007) demonstrate that two species of *Prunus* in different subgenera share identical *S-RNase* proteins, yet they have divergent introns; these data are consistent with the early divergence of species but maintenance of the coding sequence. Vieira et al. (2008) estimate that this early divergence between *Prunus* subgenera occurred a maximum of 5 Ma. Clades within *Lycium* are much younger (e.g., dispersal to East Asia

occurred ca. 1.21 Ma; see above); thus, species divergence is much more recent and, consequently, it is not surprising that the sharing reported in *Lycium* also includes identical introns.

Alternatively, it is possible that recent gene flow among species and introgression at the *S-RNase* gene could be responsible for some instances of allele sharing. Castric et al. (2008) observed similar alleles in closely related *Arabidopsis lyrata* and *A. halleri* (Brassicaceae) and demonstrated that introgression was 5-fold more likely at the *SRK* gene compared with genomic background levels. Importantly, these authors argue that introgression at the *SRK* gene is adaptive because it can facilitate the maintenance of high allelic diversity at this locus in diverging lineages. Geographically, introgression may also be a reasonable explanation for allele sharing among Asian *Lycium*. As *L. barbarum* and *L. chinense* are cultivated, they have been anthropogenically moved around China as well as introduced abroad. Furthermore, *L. barbarum*, *L. dasystemum*, and *L. truncatum* all occur together in areas of north central-northwestern China. Identical allele sharing among African species also appears common (Miller JS, unpublished data); as with the Asian species, geographic distributions suggest that introgression may be a likely explanation. Hybridization is known among South African *Lycium*, with morphological hybrids observed between several species (Spies et al. 1993; Venter 2000). However, these species are also quite recent (dispersal to Africa occurred <4 Ma and species pairs are considerably younger than that), such that recent species divergence could also explain these high levels of allele sharing.

Conclusions

In this study, we used a combination of phylogenetic analysis based on nuclear and plastid DNA data and molecular evolutionary analyses using S-RNase data to examine biogeographical relationships within *Lycium*. The S-RNase data concur with the phylogenetic data in suggesting a single dispersal event from the Americas to the Old World. This dispersal event resulted in a bottleneck of *S-RNase* alleles, with subsequent rediversification within a limited number of TGLs (fig. 1*B*). Inclusion of S-RNase data from both African and eastern Asian species of *Lycium* also concur with phylogenetic data in suggesting dispersal from Africa to eastern Asia, given that alleles retrieved from Asian taxa only occur in a subset of TGLs that also include alleles from African species. Although Asian sequences are restricted to three TGLs, selection analyses and similarity of Chinese S-RNases to those from *Lycium* species with demonstrated SI suggest that incompatibility has been maintained. A natural follow-up to this conclusion would be controlled crossing studies of Asian *Lycium*.

To date, similar studies have been conducted at the population level in animals using the major histocompatibility complex (MHC). MHC genes are important for immune system function, facilitating recognition of self versus foreign proteins. Under negative frequency–dependent selection, these genes evolve in a similar manner to the *S-RNase* gene in SI systems. As in the present study, these studies demonstrate that a multilocus approach, including both noncoding and coding DNA data, can provide complementary information about phylogeographic patterns (Aguilar et al. 2004; de Bellocq et al. 2005; Bos et al. 2008). Thus, for both plants and animals, a combination of DNA data under different selective regimes can be a valuable tool for understanding dispersal histories.

Supplementary Material

Supplementary table S1 is available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank M. Dhond, N. Feliciano, and J. Shak for laboratory contributions; N. Ambar for Israeli collections; South African provincial permitting agencies; and Y. Wang and T. Tu for assistance with Chinese *Lycium*. This work was supported by US National Science Foundation grants DEB-0343735 and DEB-0843364 to R.A.L. and J.S.M. Additional support was provided by the Schupf Scholars Program, Howard Hughes Medical Institute Fellowship Program, and Department of Biology at Amherst College.

References

Aguilar R, Bernardello G. 2001. The breeding system of *Lycium cestroides*: a Solanaceae with ovarian self-incompatibility. *Sex Plant Reprod.* 13:273–277.

Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK. 2004. High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. *Proc Natl Acad Sci U S A.* 101:3490–3494.

Benton MJ. 1993. The fossil record 2. London: Chapman and Hall.

Bianchi MB, Gibbs PE, Prado DE, Vesprini JL. 2000. Studies on the breeding systems of understorey species of a Chaco woodland in NE Argentina. *Flora* 195:339–348.

Bos DH, Gopurenko D, Williams RN, DeWoody JA. 2008. Inferring population history and demography using microsatellites, mitochondrial DNA, and major histocompatibility complex (MHC) genes. *Evolution* 62:1458–1468.

Castric V, Bechsgaard J, Schierup MH, Vekemans X. 2008. Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genet.* 4:e1000168.

Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR. 2005. Long-term genome diploidization in allopolyploid Nicotiana-section Repandae(Solanaceae). *New Phytol.* 168:241–252.

de Bellocq JG, Delarbre C, Gachelin G, Morand S. 2005. Allelic diversity at the M *hc*-DQA locus of woodmouse populations (*Apodemus sylvaticus*) present in the islands and mainland of the northern Mediterranean. *Glob Ecol Biogeogr.* 14:115–122.

Dillon MO, Tu T, Xie L, Quipuscoa Silvestre V, Wen J. 2009. Biogeographic diversification in *Nolana* (Solanaceae), a ubiquitous member of the Atacama and Peruvian Deserts along the western coast of South America. *J Syst Evol.* 47:457–476.

Drummond AJ, Ho SYW, Rawlence N, Rambaut A. 2007. A rough guide to BEAST 1.4. [cited 2010 Oct 1]. Available from: http://beast-mcmc.googlecode.com/files/BEAST14_Manual_6July2007.pdf.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 7:214.

Hartley AJ, Chong G. 2002. Late Pliocene age for the Atacama Desert: implications for the desertification of western South America. *Geology* 30:43–46.

Igic B, Bohs L, Kohn JR. 2004. Historical inferences from the self-incompatibility locus. *New Phytol.* 161:97–105.

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. 1990. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proc Natl Acad Sci U S A*. 87:9372–9375.

Ioerger TR, Gohlke JR, Xu B, Kao TH. 1991. Primary structural features of the self-incompatibility protein in Solanaceae. *Sex Plant Reprod.* 4:81–87.

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, Bernardello G, Venter AM, Miller JS. 2007. Evolutionary relationships in tribe Lycieae (Solanaceae). *Acta Hortic.* 745:225–239.

Levin RA, Whelan A, Miller JS. 2009. The utility of nuclear conserved ortholog set II (COSII) genomic regions for species-level phylogenetic inference in Lycium (Solanaceae). *Mol Phylogenet Evol.* 53:881–890.

Liu L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics*. 24:2542–2543.

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.

Maddison WP, Maddison DR. 2009. Mesquite: a modular system for evolutionary analysis. Version 2.72. [cited 2010 Oct 1]. Available from: http://mesquiteproject.org.

Miller JS, Kamath A, Levin RA. 2009. Do multiple tortoises equal a hare? The utility of nine noncoding plastid regions for species-level phylogenetics in tribe Lycieae (Solanaceae). *Syst Bot.* 34:796–804.

Miller JS, Levin RA, Feliciano NM. 2008. A tale of two continents: Baker's rule and the maintenance of self-incompatibility in Lycium (Solanaceae). *Evolution* 62:1052–1065.

Miller JS, Venable DL. 2002. The transition to gender dimorphism on an evolutionary background of self-incompatibility: an example from Lycium (Solanaceae). *Am J Bot.* 89:1907–1915.

Newbigin E, Uyenoyama MK. 2005. The evolutionary dynamics of self-incompatibility systems. *Trends Genet.* 21:500–504.

Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia VF, Collier SM. 2008. A molecular phylogeny of the Solanaceae. *Taxon* 57:1159–1181.

Paape T, Igic B, Smith SD, Olmstead R, Bohs L, Kohn JR. 2008. A 15-myr-old genetic bottleneck. *Mol Biol Evol.* 25:655–663.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14:817–818.

Rambaut A. 1996–2002. Se-Al: sequence alignment editor. Version 2.0a11. [cited 2010 Oct 1]. Available from: http://tree.bio.ed.ac.uk/software/seal/.

Rambaut A. 2006–2009. FigTree: tree figure drawing tool. Version 1.3.1. [cited 2010 Oct 1]. Available from: http://tree.bio.ed.ac.uk/software/figtree/

Rambaut A, Drummond AJ. 2003–2009. Tracer: MCMC trace analysis tool. Version 1.5.0. [cited 2010 Oct 1]. Available from: http://tree.bio.ed.ac.uk/software/tracer/.

Richman AD, Kohn JR. 1999. Self-incompatibility alleles from *Physalis*: implication for historical inference from balanced genetic polymorphisms. *Proc Natl Acad Sci U S A*. 96:168–172.

Richman AD, Kohn JR. 2000. Evolutionary genetics of self-incompatibility in the Solanaceae. *Plant Mol Biol.* 42:169–179.

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.

Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19:301–302.

Savage AE, Miller JS. 2006. Gametophytic self-incompatibility in *Lycium parishii* (Solanaceae): allelic diversity, genealogical structure, and patterns of molecular evolution at the *S-RNase* locus. *Heredity* 96:434–444.

Schnabel A, Wendel JF. 1998. Cladistic biogeography of *Gleditsia* (Leguminosae) based on *ndhF* and *rpl16* chloroplast gene sequences. *Am J Bot.* 85:1753–1765.

Spies JJ, Minne L, Venter HJT, Venter AM. 1993. A cytogenetic study of the functionally dioecious species in the genus Lycium (Solanaceae). *S Afr J Bot.* 59:535–540.

Surbanovski N, Tobutt KR, Konstantinovic M, Maksimovic V, Sargent DJ, Stevanovic V, Ortega E, Boskovic RI. 2007. Self-incompatibility of *Prunus tenella* and evidence that reproductively isolated species of *Prunus* have different SFB alleles coupled with an identical *S-RNase* allele. *Plant J.* 50:723–734.

Sutherland BG, Tobutt KR, Robbins TP. 2008. Trans-specific S-RNase and SFB alleles in *Prunus* self-incompatibility haplotypes. *Mol Genet Genomics.* 279:95–106.

Swofford DL. 2002. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4. Sunderland (MA): Sinauer Associates.

Takebayashi N, Brewer PB, Newbigin E, Uyenoyama MK. 2003. Patterns of variation within self-incompatibility loci. *Mol Biol Evol.* 20:1778–1794.

Uyenoyama MK, Zhang Y, Newbigin E. 2001. On the origin of self-incompatibility haplotypes: transition through self-compatible intermediates. *Genetics* 157:1805–1817.

Venter AM. 2000. Taxonomy of the genus Lycium L. (Solanaceae) in Africa. [Ph.D. dissertation]. [Bloemfontein (South Africa)]: University of the Orange Free State.

Vieira J, Fonseca NA, Santos RAM, Habu T, Tao R, Vieira CP. 2008. The number, age, sharing and relatedness of S-locus specificities in Prunus. *Genet Res.* 90:17–26.

Vieira J, Morales-Hojas R, Santos RAM, Vieira CP. 2007. Different positively selected sites at the gametophytic self-incompatibility pistil *S-RNase* gene in the Solanaceae and Rosaceae (*Prunus*, *Pyrus*, and *Malus*). *J Mol Evol.* 65:175–185.

Wikström N, Savolainen V, Chase MW. 2001. Evolution of the angiosperms: calibrating the family tree. *Proc R Soc B.* 268: 2211–2220.

Wong WSW, Yang Z, Goldman N, Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics* 168:1041–1051.

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

Yang Z. 2007. PAML 4: a program package for phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.

Yang Z, Nielsen R, Goldman N, Pedersen AM. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol.* 7:203–214.

Zhang ZY, Lu AM, D'Arcy WG. 1994. Solanaceae. In: Wu ZY, Raven PH, editors. Flora of China. Vol. 17. St. Louis (MO): Missouri Botanical Garden Press. p.300–332