Ancient polymorphism reveals unidirectional breeding system shifts

Boris Igic†‡, Lynn Bohs‡, and Joshua R. Kohn*†

*Section of Ecology, Behavior, and Evolution, Division of Biological Sciences, University of California at San Diego, La Jolla, CA 92093-0116; and ‡Department of Biology, University of Utah, 257 South 1400 East, Salt Lake City, UT 84112

Edited by Michael J. Donoghue, Yale University, New Haven, CT, and approved December 1, 2005 (received for review July 22, 2005)

Loss of complex characters is thought to be irreversible (Dollo’s law). However, hypotheses of irreversible evolution are remarkably difficult to test, especially when character transitions are frequent. In such cases, inference of ancestral states, in the absence of fossil evidence, is uncertain and represents the single greatest constraint for reconstructing the evolutionary history of characters. Breeding system character transitions are of particular interest because they affect the amount and distribution of genetic variation within species. Transitions from obligate outcrossing to partial or predominant self-fertilization are thought to represent one of the most common trends in flowering plants. We use the unique molecular genetic properties (manifested as deep persistent polymorphisms) of the locus that enforces outcrossing to demonstrate that its loss is irreversible in the plant family Solanaceae. We argue that current phylogenetic methods of reconstruction are potentially inadequate in cases where ancestral state information is inferred by using only the phylogeny and the distribution of character states in extant taxa. This study shows in a statistical framework that a particular character transition is irreversible, consistent with Dollo’s law.

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: ML, maximum-likelihood; SC, self-compatibility; SI, self-incompatibility; S-locus, self-incompatibility locus; S-allele, SI allele.

© 2006 by The National Academy of Sciences of the USA

Fig. 1. Shared ancestral polymorphism. (a) Evidence for shared ancestral polymorphism. Shown is a hypothetical example in which three self-incompatible species (A, C, and E) each harbor five alleles. In each case an allele from one species (e.g., A) is more closely related to alleles from the other two species (C and E) than to any other allele from its own species. This pattern implies that all five alleles were present in the common ancestor of these three species and passed down with modification to its descendants. (b) Character state inference from shared ancestral S-locus polymorphism. A hypothetical example of evolutionary relationships among eight species is shown. Four species are SI (filled squares) and four are self-compatible (open squares). Arrows indicate the species (A, C, and E) whose SI alleles have been sampled and exhibit shared ancestral polymorphism. The inheritance of polymorphism in these three descendants suggests uninterrupted history of SI. Our analyses use this inference by fixing the character state of all nodes leading from the ancestor to such species (dashed line). Species G is SI, but its S-alleles were not sampled. Nodes between species such as G (gray) and fixed nodes (dashed) were free to assume either state.

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: ML, maximum-likelihood; SC, self-compatibility; SI, self-incompatibility; S-locus, self-incompatibility locus; S-allele, SI allele.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY766240–AY766245).

†To whom correspondence should be sent at the present address: Department of Plant Breeding and Genetics, Cornell University, 252 Emerson Hall, Ithaca, NY 14853. E-mail: bigic@ucsd.edu.

© 2006 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0506283103

PNAS | January 31, 2006 | vol. 103 | no. 5 | 1359–1363
shared among extant species can be taken as evidence that SI was not only present in their common ancestor but also functioned continually from the time of the common ancestor to present day. Shared ancestral polymorphism provides evidence akin to the possession of fossilized breeding systems.

Although it is generally accepted that complex characters are more easily lost than gained, irreversibility, the extreme form of this asymmetry, is controversial (17, 18) and difficult to test (19). We use evidence of shared ancestral polymorphism to minimize error in reconstructions of the history of transitions in this breeding system character. We estimate relationships among 202 taxa of Solanaceae whose incompatibility status has been reported by combining information from a number of phylogenetic studies. We use S-allele sequence data from eight of these taxa to show that each has multiple S-allele lineages that were already present in their most recent common ancestor. Shared ancestral polymorphism implies a continuous history of SI from that ancestor to these extant taxa (Fig. 1). Applying this inference, we fix ancestral nodes leading to these eight taxa as SI before using the methods of Pagel (20) and Sanderson (21) to test the hypothesis that the rate of gain of SI has been zero (losses are irreversible). The results of these analyses are compared with those obtained when only the character states of extant taxa are used and ancestral states are unknown, as is common in com-

**Fig. 2.** Ancient polymorphism at the S-locus of Solanaceae. A phylogenetic tree of S-alleles from eight Solanaceae species shows extensive polymorphism that predates the origins of genera. The best tree obtained during a heuristic ML search in **PAUP** (37) is shown. Posterior probabilities >0.90 derived from the Bayesian analysis (36) are shown above each branch. A representative sample of alleles and taxa was used to simplify the presentation. An exhaustive sample of alleles and species yields similar results. The occurrence of ancient polymorphism at this locus is found in all taxa. (Upper Left) Sketch of the relationships among represented species (modified from ref. 27). The following S-allele sequences were used (letter/number codes are GenBank accessions): *Brugmansia versicolor*, AY766243–AY766245; *Lycium andersonii*, AF105343, AF105344, AF105347–AF105349, AF105353, AF105355, AF105358, AF105359, AF105362, and AF105363; *Nicotiana alata*, U08860, U08861, and U66427; *Nicotiana glauca*, AY766240–AY766242; *Physalis cinerascens*, AY766243, AY766244, AY766245, and AY766246; *Solanum carolinense*, L40539–L40547 and L40551; *Solanum peruvianum*, AB072457–AB072459, AB072466, AB072467, D17324, D17325, S65047, Z26582, and Z26583; *Witheringia solanacea*, AY454103–AY454107, AY454109, AY454112, AY454113, AY454115, and AY454118.
parative studies. Using inference from the S-locus, we find strong evidence for irreversible loss of SI during the diversification of the Solanaceae. Conversely, this conclusion is strongly rejected when evidence of ancestral polymorphism is ignored.

Results
The results of maximum-likelihood (ML) and Bayesian phylogenetic analysis of S-alleles from eight Solanaceae species were very similar. Fig. 2 shows the best topology found during both searches. All eight species whose S-alleles were used in the analysis show evidence of at least three S-allele lineages that date to the most recent common ancestor of these taxa (Fig. 2). This is strong evidence of a continuous history of SI since the most recent common ancestor of these species.

In our data set of Solanaceae, 69 of 202 taxa (34%) were SI (Table 2 and Supporting Text, which are published as supporting information on the PNAS web site), close to the estimate of 39% for the entire family (B.I., unpublished data). SC and SI states are often interdigitated on the family phylogeny (Fig. 3), exactly the circumstance under which reconstruction of ancestral states is ordinarily fraught with uncertainty (19). When nodes of the phylogeny ancestral to the eight sampled species were fixed as SI (Fig. 3), we found that the hypothesis of equal rates of gain and loss of SI is strongly rejected ($p < 0.0001$; Table 1 Upper). Next, we tested the hypothesis that transitions are unidirectional (rate of gain of SI is set to zero), essentially positing that the loss of this form of SI is irreversible. Our analyses fail to reject the irreversibility hypothesis ($p > 0.6$; Table 1 Upper), therefore supporting “Dollo’s law” (22) for this character. In addition, all nodes leading to extant SI taxa that were not fixed are reconstructed as significantly more likely to have been SI than SC (data not shown). Analyses that do not incorporate the information about shared ancestral polymorphism at the S-locus fail to reach these conclusions (Table 1 Lower). When analyses are based only on the phylogenetic relationships and character states of extant taxa, irreversibility is rejected ($p < 0.0001$; Table 1 Lower), nodes ancestral to species that harbor trans-specific polymorphism at the S-locus are often reconstructed as significantly more likely to have been SC than SI, and the ancestral state for the entire group is uncertain (data not shown).

Discussion
Reconstructions of the ancestry of SI that do not incorporate additional information indicate that loss of SI may not be irreversible. When data from trans-specific S-locus polymorphism are incorporated into the analyses of ancestral states, irreversibility cannot be rejected (Table 1 Upper). While this analysis cannot prove irreversibility per se (absence of evidence is not evidence of absence), we are not aware of any study that can offer more convincing statistical evidence of irreversibility.

The existing reconstruction methods, widely used in evolutionary analyses, may sometimes fail to perform adequately when not informed by additional data. Pagel (23) recognized that likelihood methods were no substitute for fossil evidence of ancestral traits. However, studies that compare analyses with and without additional evidence are rare. Our results show the impact of evidence from trans-specific evolution of S-alleles on the estimation of character state transition rates. Without it, misleading conclusions concerning the likelihood of gain of SI would be reached. We suspect that other studies using only the character states of terminal taxa may find strong statistical support for erroneous conclusions. The current methods assume that character states have no influence on diversification rates, which may be one source of error (B.I., unpublished data).

It seems reasonable that complex traits, which owe their existence to the interaction among many genes, are more likely to be lost than gained. It is widely believed that, unless the genes underlying such traits are maintained by other selective con-
strains, the possibility of reactivation is negligible in the long term (17, 24). SI fits these expectations for a Dollo character. The simultaneous action of many genes, including the co-localized style and pollen components, is required in order for the RNase-based SI system to function (25, 26). There is little evidence for pleiotropic effects of the genes involved in SI. The plurality of species in our analyses are SC and exist with no obvious deleterious effects from the lack of functional copies of S-locus or modifier genes. Furthermore, any gain of SI appears exceedingly difficult, because it requires pre-existing allelic variation and suppressed recombination between the style and pollen components. However, unlike other characters that may obey Dollo’s law, SI is not only directly adaptive as an inbreeding avoidance mechanism but also may increase the levels of genetic variation within populations and species, affecting the rate of evolution of other characters (27).

Although the present study finds strong support for the irreversible loss of RNase-based SI in the Solanaceae, it is clear that there have been cases of evolution of alternative SI mechanisms in other angiosperm groups. RNase-based SI is homologous in the three families known to use it (Solanaceae, Scrophulariaceae, and Rosaceae), indicating that the common ancestor of the majority of dicots possessed this trait (28, 29). Other, nonhomologous, forms of SI have evolved independently (28, 29). We argue that, in this instance, the irreversibility of RNase-based SI follows Dollo’s law, because the character never reverts to its ancestral condition (see ref. 22 for a detailed discussion).

Unambiguous reconstruction of ancestral states based on molecular evidence will aid studies that examine the influence of SI on diversification. Frequent and irreversible losses of RNase-based SI raise the question of how SI persists through time (9). It is either declining toward extinction or persists because of some long-term selective advantage that it confers. With a better understanding of phylogenetic relationships, and more extensive information on the compatibility status of species, our methodology may enable tests of relative diversification rates of SI and SC lineages. Because shared ancestral polymorphism is a common feature of SI systems (2, 30, 31), our approach may be applicable in dozens of angiosperm families.

Previous studies of the evolution of complex characters have been plagued by the lack of reliable information concerning ancestral states. When character states among terminal taxa are interdigitated, investigators have resorted to one of two strategies. They have either searched for the minimal character transition weightings necessary find a single origin of the complex character (32, 33) or contended that multiple gains of a complex character are likely and used indirect evidence to support the plausibility of this view (17, 18). We demonstrate the benefits of direct molecular evidence for ancestral state reconstructions of a complex character. Although genes with ancient polymorphism are unlikely to underlie many traits of interest, we expect that creative uses of molecular genetic data will become a starting point for evolutionary analyses based on reliable reconstructions of ancestral character states. Progress in understanding the molecular genetic basis of phenotypic differences (34, 35) will produce a wealth of data that enable compelling evolutionary inferences.

Methods

Trans-Specific Polymorphism. The phylogenetic evidence for shared ancestral polymorphism at the S-locus used in this study came from analysis of 61 S-allele sequences from eight SI taxa: B. versicolor, L. andersonii, N. alata, N. glauca, P. cinerascens, S. carolinense, S. peruvianum, and W. solanacea (Fig. 2).

We used MrBayes v.3.0 (36) to generate a phylogenetic hypothesis for these S-alleles. We implemented Bayesian analyses using Markov chain Monte Carlo sampling with the Metropolis–Hastings–Green algorithm running four chains (three heated, one cold) for 5,000,000 generations. We sampled every 2,500th tree in this analysis. The initial 1,001 trees were discarded in the burn-in, well after stationarity was reached. The posterior probabilities of individual clades were calculated by using the remaining 1,000 trees. We also used PAUP 4.0b10 (37) to heuristically find the best tree in a ML search. For this search, we used MODELEST 3.0 (38) to find the optimal model of evolution (TVM+I+G), selected by using the Akaike Information Criterion (39).

Species Phylogeny. We generated a hypothesis of evolutionary relationships using a composite phylogenetic tree constructed from a number of recent molecular phylogenies of Solanaceae species. For its backbone, our species phylogeny (Figs. 3 and 4, which is published as supporting information on the PNAS web site), relies on broad-scale phylogenetic studies (40, 41). The placement of shallow nodes and tips was obtained by the grafting of trees from detailed phylogenetic studies of well established monophyletic groups (40, 42–51) onto the backbone. Overall, the analysis contained 202 taxa for which both breeding system data and phylogenetic position were available. In most cases, we imposed the branching order found in strict consensus trees of the individual data sets listed above. Bootstrap values are given for branches with >70% support (Fig. 4), except for data derived from two studies. Spooner and Sytsma (43) did not report bootstrap values. Mace et al. (48) performed a distance analysis on amplified fragment length polymorphism data, so neither a consensus analysis nor a metric of support is available.

Species for which both SI and self-compatibility (SC) were reported (e.g., most populations are SI but some are SC) were encoded as separate taxa with infinitesimal branch lengths separating them. Polymorphisms were resolved randomly and assigned infinitesimal values (0.000001), so as to essentially remain unresolved. Different resolutions do not change our results significantly. All other branches were assigned unit lengths.

Tests of Irreversibility. We conducted ML analyses of transition rates between SI and SC in a phylogenetic framework (20, 52). We incorporated the information on the continuous history of SI by fixing all nodes leading from the common ancestor to the eight extant species from six genera of Solanaceae that share ancestral S-locus polymorphism (Figs. 2 and 3). Transition rate estimates were obtained by using DISCRETE 2.0 (20). ML esti-
mates of transition rates and the associated log-likelihood (lnL) values were used to conduct two likelihood ratio tests. The null model estimates both gain and loss parameters freely (two-parameter model). To test whether the transition rates are significantly asymmetric, we posited that the rates of gain and loss of SI were equal and compared the values of this one-parameter model with a free estimate. The resulting value of ΔlnL approximates a χ² distribution with one degree of freedom (20). Next, we found the ML estimates for the irreversibility model in which the rate of gain of SI was fixed to zero, and the distribution with one degree of freedom was used.

We thank R. Olmstead for a critical review, suggestions, and unpublished data used in the paper; A. Angert, J. Bollback, P. Fenberg, B. Fischman, H. Hockstra, E. Goldberg, K. Marhaver, A. Putnam, and S. Weller for helping to improve the early versions of the manuscript and providing valuable discussions; and T. Case, M. Chase, R. Lande, T. Mione, E. Newbiggin, P. Smith, S. Smith, J. Stone, and M. Whitson for generously sharing data or ideas. This work was supported by grants from the National Science Foundation (DEB-0108173 and DEB-0309184) and Sigma Xi (to J.R.K. and B.L.).