reduction in chemical potential by forming the crystalline phase. In the second scenario, we assume that oxygen has first adsorbed to the LS interface, driven by the reduction in interface energy (*24*, *28*), having diffused along the interface from the triple junction. Once oxygen adsorbs, a critical-size nucleus (in the form of a step) forms at the triple junction and sweeps across the LS (0001) interface, forming a new (0006) layer of α-Al_2O_3 in its wake. Oxygen then again adsorbs to the new LS interface, supplied from the dissolving outer rim of the nanowire. The nucleation rate of steps is relatively low under the present experimental conditions, leading to single steps that sweep across the interface (no defects resulting from the intersection of steps or multiple nucleation events were detected). Both of these reactions would be very fast; the formation of a new (0006) layer after oxygen diffusion has started requires only ~0.16 s. Our experimental observations cannot distinguish between these two scenarios.

Our in situ HRTEM observations have revealed an oscillatory growth process that is not obvious from ex situ observations of grown nanowires. These observations provide an explanation for the (0001) growth mechanism and why it is not continuous. What distinguishes this reaction from most VLS processes is that the catalyst droplet is not directly involved in supplying all the necessary components for growth. Instead, oxygen for the new (0006) layer is supplied by transient sacrificial dissolution of the top rim of the nanowires via facet formation. The components of interface tensions at the triple junction also promote formation of the facets and probably drive the nucleation of steps of solid (0006), which sweep across the LS (0001) interface, resulting in growth of a new (0006) layer. Formation of the (0006) layer is relatively fast because the Al atoms on the liquid side of the LS surface are already in positions similar to those in the underlying bulk sapphire. The oxygen collected on the LV surface is used to eradicate the facets at the outer, top rim of the nanowires once a complete (0006) layer is formed. It is this process that is the rate-limiting step.

References and Notes

1. E. I. Givargizov, *J. Cryst. Growth* **31**, 20 (1975).
2. F. M. Ross, J. Tersoff, M. C. Reuter, *Phys. Rev. Lett.* **95**, 146104 (2005).
3. Y. Hao, G. Meng, Z. L. Wang, C. Ye, L. Zhang, *Nano Lett.* **6**, 1650 (2006).
4. L. E. Jensen *et al.*, *Nano Lett.* **4**, 1961 (2004).
5. S. Kodambaka, J. Tersoff, M. C. Reuter, F. M. Ross, *Science* **316**, 729 (2007).
6. V. Valcárcel, A. Souto, F. Guitián, *Adv. Mater.* **10**, 138 (1998).
7. W. F. Li *et al.*, *Phys. Stat. Sol. A* **203**, 294 (2006).
8. T. Kuykendall *et al.*, *Nat. Mater.* **3**, 524 (2004).
9. E. A. Stach *et al.*, *Nano Lett.* **3**, 867 (2003).
10. Y. Wu, P. Yang, *J. Am. Chem. Soc.* **123**, 3165 (2001).
11. S. Hofmann *et al.*, *Nat. Mater.* **7**, 372 (2008).
12. G. Lee *et al.*, *Angew. Chem.* **48**, 7366 (2009).
13. B. Tian, P. Xie, T. J. Kempa, D. C. Bell, C. M. Lieber, *Nat. Nanotechnol.* **4**, 824 (2009).
14. S. H. Oh, Y. Kauffmann, C. Scheu, W. D. Kaplan, M. Rühle, *Science* **310**, 661 (2005); published online 6 October 2005 (10.1126/science.1118611).
15. See supporting material on *Science* Online.
16. V. Laurent, D. Chatain, C. Chatillon, N. Eustathopoulos, *Acta Metall.* **36**, 1797 (1988).
17. D. Chatain, W. C. Carter, *Nat. Mater.* **3**, 843 (2004).
18. E. J. Siem, W. C. Carter, D. Chatain, *Philos. Mag.* **84**, 991 (2004).
19. H. Wang, G. S. Fischman, *J. Appl. Phys.* **76**, 1557 (1994).
20. H. Saka, K. Sasaki, S. T. Tsukimoto, S. Arai, *J. Mater. Res.* **20**, 1629 (2005).
21. D. J. Siegel, L. G. Hector Jr., J. B. Adams, *Phys. Rev. B* **65**, 085415 (2002).
22. W. D. Kaplan, Y. Kauffmann, *Annu. Rev. Mater. Res.* **36**, 1 (2006).
23. A. Hashibon, J. Adler, M. W. Finnis, W. D. Kaplan, *Interface Sci.* **9**, 175 (2001).
24. G. Levi, W. D. Kaplan, *Acta Mater.* **50**, 75 (2002).
25. The reaction volume was calculated by assuming a ring of triangular cross section. Calculation of the volume yields ~1.7×10^{-20} cm^3. Given that the number of oxygen atoms per unit volume of α-Al_2O_3 is 8.63×10^{22} atoms cm^{-3}, the number of oxygen atoms contained within the reaction volume is ~1.46×10^3 atoms. The number of oxygen atoms needed to deposit a monolayer of close-packed oxygen at the LS (0001) interface was calculated by assuming a square and a circular shape of the interfaces. The area density of a hexagonal close-packed oxygen layer is 1.6×10^{15} atoms cm^{-2}. The amount of oxygen needed to deposit a monolayer of close-packed oxygen is ~1.6×10^3 and ~1.2×10^3 atoms for the square and the circular interfaces, respectively.
26. S. M. Roper *et al.*, *J. Appl. Phys.* **102**, 034304 (2007).
27. K. W. Schwarz, J. Tersoff, *Phys. Rev. Lett.* **102**, 206101 (2009).
28. G. Levi, W. D. Kaplan, *J. Am. Ceram. Soc.* **85**, 1601 (2002).

29. The in situ TEM experiments were carried out using the Stuttgart high-voltage microscope (JEM-ARM 1250) at the Max-Planck-Institut für Metallforschung. We thank F. Phillipp, R. Höschen, and U. Salzberger for technical support, and D. Chatain and the participants of the 3 Phase, Interface, Component Systems (PICS3) meeting for stimulating discussions. Supported by the German Science Foundation via the Graduiertenkolleg Innere Grenzflächen (GRK 285/3), the German-Israel Fund (grant I-779-42.10/2003), the Nano and Steel Fusion Research Program funded by the POSCO (S.H.O.) Basic Science Research Institute grant 2.0006028 (S.H.O.), the Basic Science Research Program through the National Research Foundation funded by the Ministry of Education, Science and Technology (grant 2010-0005834) (S.H.O.), the excellence cluster Nanosystems Initiative Munich (C.S. and S.H.O.), and the Materials Sciences and Engineering Division of the U.S. Department of Energy (M.F.C. and W.L.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6003/489/DC1
SOM Text
Figs. S1 to S5
References
Movies S1 to S3

7 April 2010; accepted 16 September 2010
10.1126/science.1190596

Downloaded from www.sciencemag.org on October 21, 2010

Species Selection Maintains Self-Incompatibility

Emma E. Goldberg,[1] Joshua R. Kohn,[2] Russell Lande,[3] Kelly A. Robertson,[1] Stephen A. Smith,[4] Boris Igić[1]*

Identifying traits that affect rates of speciation and extinction and, hence, explain differences in species diversity among clades is a major goal of evolutionary biology. Detecting such traits is especially difficult when they undergo frequent transitions between states. Self-incompatibility, the ability of hermaphrodites to enforce outcrossing, is frequently lost in flowering plants, enabling self-fertilization. We show, however, that in the nightshade plant family (Solanaceae), species with functional self-incompatibility diversify at a significantly higher rate than those without it. The apparent short-term advantages of potentially self-fertilizing individuals are therefore offset by strong species selection, which favors obligate outcrossing.

The astonishing diversity of flowering plants has led to numerous hypotheses linking the propensity for diversification to morphological, life history, and physiological characteristics (*1*–*4*). Although it is clear that no single character can entirely explain the variation in plant diversity (*2*, *4*), many traits potentially associated with an increase in species richness may have been overlooked because of methodological shortcomings (*5*, *6*).

The overwhelming majority of flowering plants function as both males and females, but despite the potential for self-fertilization in hermaphrodites, most of these plants predominantly or exclusively mate with other individuals (outcross) (*7*, *8*). Outcrossing is often enforced by self-incompatibility (SI): the ability of individual plants to recognize and reject their own pollen. Each SI system characterized to date relies on

[1]Department of Biological Sciences, University of Illinois at Chicago, 840 West Taylor Street, M/C 067, Chicago, IL 60607, USA. [2]Division of Biological Sciences, University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA. [3]Division of Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire, SL5 7PY, UK. [4]National Evolutionary Synthesis Center 2024 West Main Street, Durham, NC 27705, USA.

*To whom correspondence should be addressed. E-mail: boris@uic.edu

complex genetic mechanisms (*9*). Their value is primarily attributed to the avoidance of inbreeding depression (*10*), but the short-term disadvantages of SI, which limits an individual plant's ability to reproduce if outcross pollen is not available, is expected to result in frequent transitions to self-compatibility (SC) (*8*). Mate limitation and the automatic advantage of selfing (*10*, *11*) each favor the spread of mutants that break down SI. Indeed, angiosperm families in which SI is found also contain significant proportions of SC species (*12*). The two states are commonly interspersed within clades, implying frequent evolutionary transitions. Independent data, both genetic and phylogenetic (*12*, *13*), suggest that the transition from SI to SC is one of the most common evolutionary shifts in flowering plants (*14*). Studies of the genetic basis of SI indicate that novel SI systems originate very rarely (*9*, *15*). The combined evidence implies that SI is frequently lost by species and that an identical form of SI is only rarely, if ever, regained (*12*).

In Solanaceae species with SI, strong negative frequency-dependent selection maintains dozens of alleles at the locus controlling self-recognition (S-locus), where it has preserved ancestral polymorphism over 40 million years (*13*, *16*, *17*). After pollination, if the haploid pollen allele matches either allele of the diploid female reproductive organ, fertilization is prevented. Individuals with common alleles therefore have fewer potential mates, and those with rare alleles more, so that alleles ordinarily lost by drift are instead maintained over long time scales (*16*). The outcome is a pattern of ancestral polymorphism shared not only across species boundaries but among distantly related genera. Randomly sampled alleles from different SI species of Solanaceae are often more closely related than those within species (*16*, *17*). This provides strong evidence that a polymorphic and therefore functional S-locus was present in the common ancestor and passed on to descendent SI species in the family (*17*).

Although the SI mechanism has been frequently lost and not regained within Solanaceae (*13*), it occurs at an intermediate frequency. The proportion of SI species is declining with each transition to SC, so SI may be destined for extinction. But the SI mechanism of Solanaceae is ~90 million years old, nearly as old as eudicots (*15*), and SI may instead persist if the diversification rate of SI lineages exceeds that of SC lineages by at least the SI-to-SC transition rate (*17*). In this scenario, the loss of SI may be countered by a greater difference between speciation and extinction rates associated with the presence of SI relative to SC. A strong association between SI and increased diversification would provide unusually compelling evidence for a causal relationship, because frequent character changes to SC uncouple the evolution of the focal character from others and reduce the chance that an inferred connection is spurious (*18*).

We examined the effect of SI on diversification in Solanaceae. The nightshades contain many agriculturally important species (such as tomato, potato, chile, and tobacco), so phylogenetic and breeding system data are extensive. Among the estimated 2700 species in the family, approximately 57% are SC, 41% are SI, and less than 2% have separate male and female plants (*19*).

We constructed a phylogeny through an initial alignment matrix of 998 species and eight genes, and we assigned character states (SI or SC) for 616 species (*19*). Subsequent analyses concentrated on a monophyletic clade termed "x = 12" for its uniform base chromosome number, containing approximately 2300 species, including the subfamilies Solanoideae and Nicotianoideae (*20*) [we also conducted analyses on the whole family (*19*)]. Both character states and phylogenetic data are better sampled in this clade than in the rest of the family, yielding 356 species present in the tree and with known mating systems (Fig. 1). This sample did not statistically differ in SI frequency from the total available data set of 616 species ($P > 0.3$), and our analyses accounted for sampling incompleteness (*21*).

Irreversible loss of SI in Solanaceae will cause the frequency of SI to decline to zero if $r_I \leq r_C + q_{IC}$, where r_I and r_C are net diversification rates (the difference between speciation and extinction rates) for SI and SC lineages, respectively, and q_{IC} is the per-lineage rate of transition from SI to SC. Alternatively, the loss of SI may be offset by more rapid diversification of SI lineages when $r_I > r_C + q_{IC}$. In this latter case, the equilibrium proportion of SI species, p_{SI}, is positive and equal to $1 - q_{IC}/(r_I - r_C)$ (*17*).

We analyzed character evolution and diversification on the maximum likelihood tree (Fig. 1) and also on 100 bootstrap trees to assess robustness to phylogenetic uncertainty (*19*). Rates of speciation (λ_I and λ_C; the subscripts denote SI

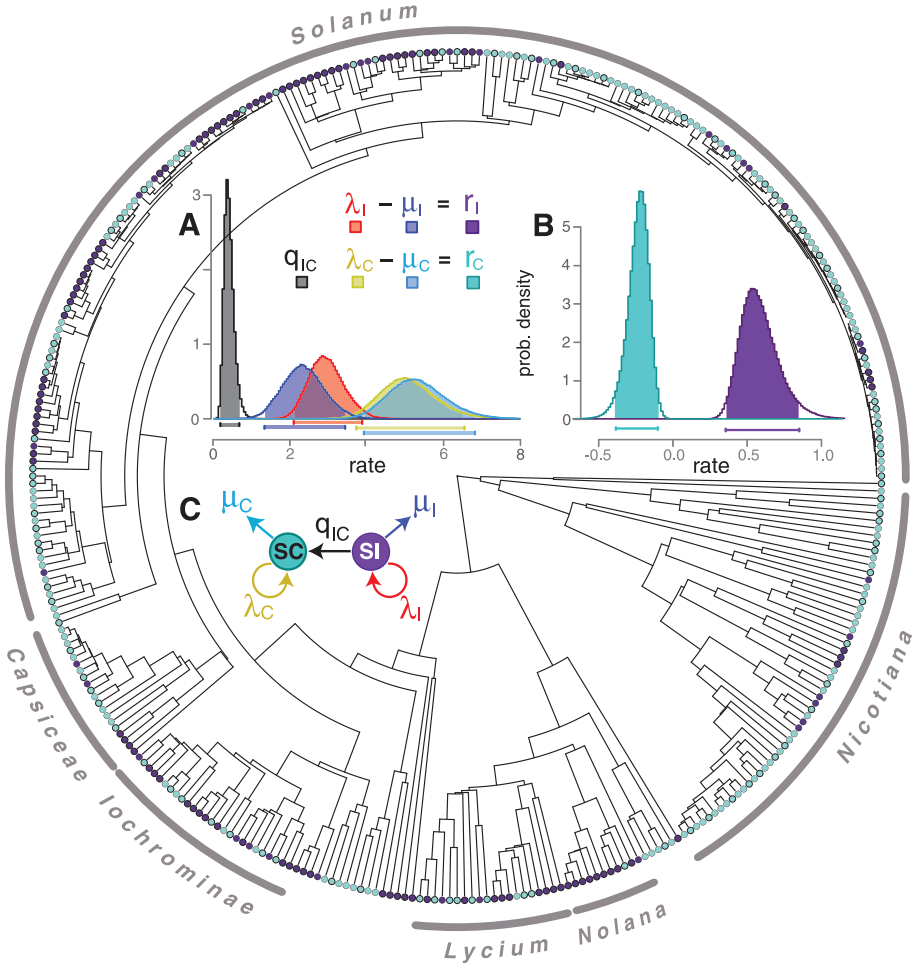

Fig. 1. Maximum likelihood tree of phylogenetic relationships among 356 species of Solanaceae. Higher ranks are indicated around the perimeter of the tree. Purple and turquoise tip colors denote SI and SC extant species, respectively. The root age is 36 million years. Inset panels display posterior probability distributions and 95% credibility intervals of reconstructed rates of character evolution (the time unit is millions of years). (**A**) BiSSE estimates of transition, speciation, and extinction parameters ($q_{IC} << \mu_I < \lambda_I << \lambda_C < \mu_C$). (**B**) Net diversification rate—the difference between speciation and extinction rates—associated with each state. (**C**) Schematic summary of estimated rate parameters. For methods, species names, character states, and further results, see (*19*).

Downloaded from www.sciencemag.org on October 21, 2010

and SC states, respectively), extinction (μ_I and μ_C), and loss of SI (q_{IC}) were estimated with Markov chain Monte Carlo analyses under the binary-state speciation and extinction (BiSSE) model (*19*, *22*), which explicitly incorporates differing speciation and extinction rates associated with the two states of a binary character. Because SI has apparently not been regained in this clade (*13*), we fixed the rate of transitions from SC to SI, q_{CI}, to be zero in our main analyses [although we also considered bidirectional transitions (*19*)].

We find a higher rate of speciation for SC than SI lineages, but this apparent advantage is erased by an SC extinction rate that exceeds that of SI lineages and even the SC speciation rate [$\mu_I < \lambda_I < \lambda_C < \mu_C$, with posterior probability 0.99 (Fig. 1A)]. SI lineages consequently show a substantially higher rate of net diversification than do SC lineages [$r_I > r_C$, with posterior probability 1.0 (Fig. 1B)]. Furthermore, this diversification difference is large enough to counter the irreversible loss of SI [$r_I > r_C + q_{IC}$, with posterior probability 1.0 (Fig. 2A)]. The equilibrium frequency of SI is therefore positive [$p_{SI} > 0.25$, with posterior probability 0.99 (Fig. 2B)], demonstrating that the macroevolutionary advantage of SI is enough for it to persist indefinitely, despite its loss through transitions to SC. These results are robust to phylogenetic uncertainty and against the assumption that SI is not regained (*19*).

The higher rate of diversification associated with SI means that selection is acting among species, which can produce trends that are not predictable from studies of selection among individuals within species. Natural selection can act on many levels of evolutionary hierarchy (including genes, individuals, populations, and species), with the potential for cascading effects of selection from any level to those above and below (*18*, *23*–*26*). Species, as units, may vary in heritable traits associated with different rates of multiplication. A higher form of selection may thus be manifested as a trait-dependent net diversification rate. There is wide acceptance of the existence of higher-order selection (*18*, *25*, *26*) but relatively sparse data on its importance (*25*, *26*).

Our results point to a strong role for species selection in determining the evolutionary dynamics and distribution of mating systems among species. Both theoretical and empirical studies on the shape of outcrossing rate distributions have overlooked the possibility that species selection may produce an arbitrary proportion of selfing, mixed-mating, and outcrossing species (*8*, *27*).

It remains unknown why strict avoidance of self-fertilization yields such strong long-term evolutionary advantages. Outcrossing plants, which generally maintain high effective population sizes and recombination rates, may have increased polymorphism, lower linkage disequilibrium, more efficient response to purifying selection, more extensive geographic distribution of genetic diversity, and lower rates of extinction (*28*–*31*). Various mechanisms other than SI, such as the physical and temporal separation of sexes within hermaphrodite flowers, or single-sex individuals, can effect outcrossing in the absence of SI (*3*). Moreover, although SI may be a relatively efficient mechanism for obligate outcrossing (*32*), delayed selfing ability and plastic sex expression afforded by other breeding systems may additionally provide reproductive assurance (*7*).

At short time scales, the expression of inbreeding depression can oppose the spread of mutations that break down SI, but many ecological factors favor their rapid and frequent fixation (*7*). This is problematic because natural selection cannot maintain features slated for future use (*18*). We find that SC is associated with a macroevolutionary disadvantage from high extinction rates, despite its high speciation rates relative to SI. Both theoretical and empirical evidence link inbreeding with an increased probability of extinction (*27*, *29*, *31*). Self-fertilization and increased inbreeding are also closely tied to higher among-population differentiation (*30*), which may in turn lead to escalating speciation rates.

We conclude that species selection opposes the short-term advantages provided by the ability of flowering plants to self-fertilize, and that volatile dynamics underlie the scattered phylogenetic distribution of SC plant species. A strong relationship between traits, including SI, and speciation and extinction may result in the uneven patterns of diversification rates observed among clades of flowering plants (*2*, *4*), especially when coupled with time heterogeneity and historical contingency. These results provide evidence that the evolutionary origin and maintenance of SI may play an important role in shaping the immense extant diversity of flowering plants (*18*, *33*).

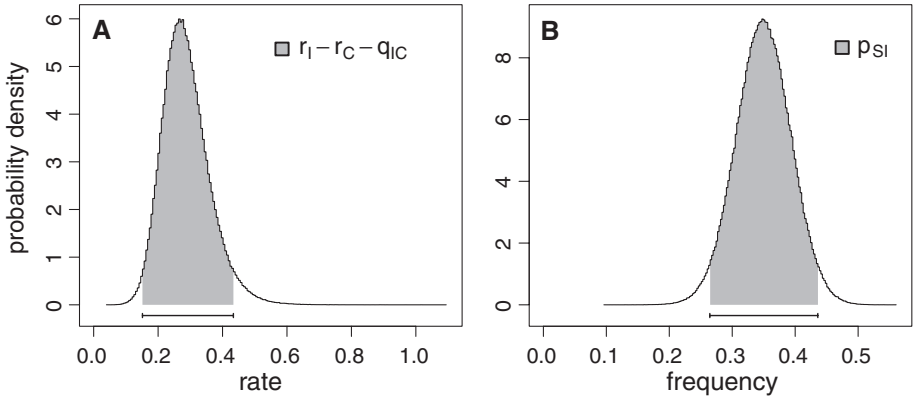

Fig. 2. Further analyses of the posterior rate distributions in Fig. 1. (**A**) The diversification rate of SI is sufficiently higher than that of SC for it to persist deterministically (shown by positive values here) despite frequent and irreversible loss. (**B**) The equilibrium frequency of SI is consequently always positive.

References and Notes

1. C. Darwin, F. Darwin, A. C. Seward, *More Letters of Charles Darwin: A Record of His Work in a Series of Hitherto Unpublished Letters* (J. Murray, London, 1903).
2. S. Magallón, M. J. Sanderson, *Evolution* **55**, 1762 (2001).
3. S. C. H. Barrett, *Philos. Trans. R. Soc. London Ser. B* **358**, 991 (2003).
4. T. J. Davies *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 1904 (2004).
5. W. P. Maddison, *Evolution* **60**, 1743 (2006).
6. E. E. Goldberg, B. Igić, *Evolution* **62**, 2727 (2008).
7. C. Goodwillie, S. Kalisz, C. G. Eckert, *Annu. Rev. Ecol. Evol. Syst.* **36**, 47 (2005).
8. B. Igić, J. R. Kohn, *Evolution* **60**, 1098 (2006).
9. V. Franklin-Tong, *Self-Incompatibility in Flowering Plants: Evolution, Diversity, and Mechanisms* (Springer, Berlin, 2008).
10. E. Porcher, R. Lande, *Evolution* **59**, 46 (2005).
11. J. W. Busch, D. J. Schoen, *Trends Plant Sci.* **13**, 128 (2008).
12. B. Igić, R. Lande, J. R. Kohn, *Int. J. Plant Sci.* **169**, 93 (2008).
13. B. Igić, L. Bohs, J. R. Kohn, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1359 (2006).
14. G. L. Stebbins, *Flowering Plants* (Belknap Press, Cambridge, MA, 1974).
15. J. E. Steinbachs, K. E. Holsinger, *Mol. Biol. Evol.* **19**, 825 (2002).
16. T. R. Ioerger, A. G. Clark, T. H. Kao, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9732 (1990).
17. B. Igić, L. Bohs, J. R. Kohn, *New Phytol.* **161**, 97 (2004).
18. G. C. Williams, *Natural Selection: Domains, Levels, and Challenges* (Oxford Univ. Press, New York, 1992).
19. Materials, methods, and further results are available as supporting material on *Science* Online.
20. R. G. Olmstead, L. Bohs, *Acta Hortic.* **745**, 255 (2007).
21. R. G. FitzJohn, W. P. Maddison, S. P. Otto, *Syst. Biol.* **58**, 595 (2009).
22. W. P. Maddison, P. E. Midford, S. P. Otto, *Syst. Biol.* **56**, 701 (2007).
23. R. C. Lewontin, *Annu. Rev. Ecol. Syst.* **1**, 1 (1970).
24. S. J. Gould, *Science* **216**, 380 (1982).
25. D. Jablonski, *Annu. Rev. Ecol. Evol. Syst.* **39**, 501 (2008).
26. D. L. Rabosky, A. R. McCune, *Trends Ecol. Evol.* **25**, 68 (2010).
27. D. J. Schoen, J. W. Busch, *Int. J. Plant Sci.* **169**, 119 (2008).
28. G. L. Stebbins, *Am. Nat.* **91**, 337 (1957).
29. D. J. Schoen, A. H. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4494 (1991).
30. J. L. Hamrick, M. J. W. Godt, *Philos. Trans. R. Soc. London Ser. B* **351**, 1291 (1996).
31. S. Glémin, E. Bazin, D. Charlesworth, *Proc. R. Soc. London Ser. B* **273**, 3011 (2006).
32. K. Karoly, *Am. Nat.* **144**, 677 (1994).
33. M. S. Zavada, T. N. Taylor, *Am. Nat.* **128**, 538 (1986).

34. We thank R. FitzJohn, A. Mooers, R. Olmstead, T. Paape, L. Popović, D. L. Shade, and R. Ree for help and discussions. Data are deposited in the Dryad database with accession no. 1888. Funded by NSF grant DEB-0919089 to B.I.

Supporting Online Material

www.sciencemag.org/cgi/content/full/330/6003/493/DC1
Materials and Methods
SOM Text
Figs. S1 and S2
Tables S1 to S4
References

1 July 2010; accepted 25 August 2010
10.1126/science.1194513

Downloaded from www.sciencemag.org on October 21, 2010

www.sciencemag.org/cgi/content/full/330/6003/493/DC1

Supporting Online Material for

Species Selection Maintains Self-Incompatibility

E. E. Goldberg, J. R. Kohn, R. Lande, K. A. Robertson, S. A. Smith, B. Igic*

*To whom correspondence should be addressed. E-mail: boris@uic.edu

Published 22 October 2010, *Science* **330**, 493 (2010)
DOI: 10.1126/science.1194513

This PDF file includes:

Materials and Methods
SOM Text
Figs. S1 and S2
Tables S1 to S4
References

Materials and methods

Phylogeny construction

Alignment and tree construction pipeline

The procedure for reconstructing phylogenetic relationships among a subset of Solanaceae closely followed the one developed and reported in detail by Smith and colleagues (*S1, S2*). We implemented this pipeline in *Python* (v2.5) with the *BioPython* (v1.48) module (*S3*) and using the *BioSQL* (v1.0.1) database schema (*S4*). Here, we present the most important aspects of this procedure.

We found eight loci within Solanaceae with more than 200 taxa represented in GenBank. Six of these genes are encoded in the chloroplast genome: atpB (ca. 300 entries), matK (520), ndhF (500), rbcL (220), trnK (210), trnL-trnF (700), and two in the nuclear genome: GBSSI (790), ITS (910). We removed taxa with identical (duplicated) names and produced a first-pass multiple alignment using MAFFT (*S5*), which resulted in a 1084-taxon, 16,405bp (eight-gene) matrix. Gene-partitioned phylogenetic reconstructions were performed using RAxML (v7.0.4) (*S6*). Subsequent analyses were performed on the maximum likelihood (ML) tree found in a heuristic search and on 100 bootstrap replicates. We manually curated the tree obtained after the initial pass. The resulting matrix contained 995 ingroup taxa, plus five outgroup species of Convolvulaceae (Fig. S1).

The resulting trees were then pruned to include only those species contained within the "x=12" clade, the best-sampled monophyletic group in the family. Cultivars and known hybrid varieties were removed from the analyses, as were those species for which the self-incompatibility status is unknown. The final dataset contained 356 species (Fig. S2).

In addition, we present below abbreviated results for the dataset consisting of all 397 species in the family on the tree and with known character states.

Relaxed molecular clock implementation

In order to ultrametricize our phylogenetic trees and obtain node and tip dates proportional to true time (rather than substitutions per site), we used the semiparametric method developed by Sanderson (*S7*). This method accommodates rate variation among lineages by combining a parametric model with different nucleotide substitution rates on every branch with a nonparametric roughness penalty. The λ parameter for penalized likelihood calculations determines the contribution of this penalty. It was estimated using a cross-validation procedure implemented in *r8s* (v1.71) (*S7*). We applied 13 values of λ, spaced evenly on a logarithmic scale from 1 to 1000, after randomly pruning the ML tree to 71 taxa (for computational feasibility) 10 times. In each pruned set, the best estimate was $\log_{10} \lambda = 2.25$, which we used for penalized likelihood rate-smoothing on the ML and bootstrap trees.

Accurate absolute divergence dates are not crucial for our analyses. Nevertheless, we obtained approximate values in order to make the units of the results of our character evolution analyses meaningful for comparative study. We constrained the age of the focal crown group ("x=12" clade) to 36 Ma based on the node date obtained by Paape and colleagues (*S8*). This dating procedure relies on the use of two key constraints derived from dating and identification of *Solanum*- and *Physalis*-like seeds (*S9*). Those two dates were used to anchor a normally distributed age prior (μ = 10.0 Ma, σ = 4.0 Ma; 95% CI = [3.4, 16.6]), along with the date for stem group age of Solanaceae (divergence from Convolvulaceae; μ = 52 Ma, σ = 5.2 Ma; 95% CI = [43, 60]). The divergence dates we obtained are comparable to others recovered with different strategies [e.g., (*S10, S11*)].

Character evolution analyses

Testing for a correlation between character state and net diversification rate has commonly relied on associations of clade age and number of species, either through the use of sister clade

or regression analyses (*S12–S14*). These methods consider only clades in which most tips are in the same state and assume that ancestors within the clade had that state. Because the transition rate from SI to SC is high relative to the family's age (*S15*), few non-polymorphic clades are available, and such tests will have low statistical power (*S13, S16*). We therefore employ the recently-developed BiSSE method for estimating from a phylogeny speciation and extinction rates that depend on character states (*S17*). We include the correction for incomplete sampling (*S18*): the probability of being sampled in the tree is 0.162 for SI species and 0.150 for SC species. By explicitly modeling speciation and extinction associated with the two states of a binary character, BiSSE allows us to quantify the effects of state-dependent diversification and state transitions.

Character state encoding

For macroevolutionary models of binary character evolution to be applied to breeding system evolution, each species must be classified as either SI or SC, and this can be difficult (*S19*). We use the same general principles as described elsewhere (*S20*). The best available data, however sparse, indicate that the breakdown of SI is so common that no SI species studied in any detail fails to reveal segregating SC mutants, SC populations, or sister species. Our intention is to measure the average rate at which an entire lineage, ancestrally SI (see main text and (*S20*)), transitions to SC. Therefore, we encode as SI all "polymorphic" taxa that have not entirely transitioned to SI. This encoding should yield a conservative estimate of the advantage of SI lineages. The breeding system data were available for 356 species represented on the phylogeny of "x=12" Solanaceae. Of those, 196 were described in the literature and encoded by us as SC. Of the remaining species encoded as SI, 116 were unambiguously described as SI, 17 as both SI and SC, 25 as dioecious, and one each as SI/SC/dioecious and SI/dioecious.

Bayesian analysis

We performed a Bayesian analysis of BiSSE on the Solanaceae ML phylogeny. Compared with maximum likelihood parameter estimation, this approach better handles the potentially-rough likelihood surfaces and provides a natural means of testing how meaningful parameter differences are. Although a Bayesian analysis across a posterior set of trees would be more philosophically consistent, obtaining such a set was not computational feasible due to the large number of species and sites in the phylogenetic analysis.

We obtained 1,000,000 post-burn-in samples of slice-sampling Markov chain Monte Carlo (MCMC) on the maximum likelihood tree and 10,000 samples on each of the 100 bootstrap trees. Priors on each parameter were exponentially distributed and broad, with rate 0.3. We used the conditional likelihood root state assumption (*S18*): because the rate of gain of SI was fixed to zero, this effectively fixes as SI the root and hence the other internal nodes leading to SI tips. Analyses were conducted with the *R* package *diversitree* (*S21, S22*). Once posterior distributions of the five model parameters were obtained (Fig. 1A), further analyses (Fig. 1B, Fig. 2) simply involved computing additional quantities from each MCMC sample. Reported posterior probabilities are the proportion of MCMC samples for which a statement was true.

Supporting text

Additional character evolution results

Results from maximum likelihood tree

Mean values, standard deviations, and the standard errors of the mean (determined from time series analysis as implemented in the *R* package *coda* (*S23*)) are reported in Table S1. The results for $\lambda_C - \lambda_I$ and $\mu_C - \mu_I$ provide further quantification of our finding that SC lineages have a higher rate of speciation and an even higher rate of extinction than do self-incompatible lineages. The results for r_I, r_C and their difference show not only a higher rate of net diversifi-

cation for SI lineages, but in fact negative net diversification for SC lineages. The equilibrium frequency of SI is only slightly less than the observed frequency and is certainly positive, showing that SI is not bound for deterministic extinction due to its loss through transitions to SC.

Correlations between the five directly estimated parameters are reported in Table S2. The speciation and extinction rates for each state are very strongly correlated, consistent with the much tighter estimates of net diversification rates than of speciation and extinction rates separately (Fig. 1AB, Table S1). The moderate correlations between the transition rate and the rest of the rates reflects the balance between state change and diversification in explaining the observed state frequencies.

Results from bootstrap trees

The results reported above include uncertainty from the parameter estimation process but not uncertainty in the tree topology or branch lengths. Variation in individual parameter estimates across the bootstrap trees is comparable to uncertainties in estimation on the ML tree (Table S3). For net diversification rates, however, variation across bootstrap trees is much less (Table S3), likely because diversification rate is determined largely by the frequency of tip states and the root age, both of which are constant across the bootstrap set.

Our main results are robust to phylogenetic uncertainty. On 92 out of the 100 bootstrap trees, $\mu_I < \lambda_I < \lambda_C < \mu_C$ is true with a posterior probability of at least 0.90. On all 100 bootstrap trees, the following statements are true with posterior probability 0.90: $r_I > 0.35$, $r_C < -0.01$, $r_I - r_C - q_{IC} > 0.12$, $p_{SI} > 0.27$.

Results with regain of SI allowed

Although we feel we are justified in fixing the rate of SI gain, q_{CI}, to zero in our analyses, we understand that results without that restriction may be of interest. We therefore also ran the BiSSE MCMC analysis on the ML tree without the general $q_{CI} = 0$ constraint, but incorporat-

ing information on specific lineages in which reversals were shown not to occur. For the seventeen species on our tree in which trans-specific polymorphism has been documented (*Solanum chilense, S. peruvianum, S. carolinense, S. chacoense, Lycium californicum, L. andersonii, L. parishii, L. ferocissimum, L. cestroides, Physalis cinerascens, P. crassifolia, P. longifolia, Witheringia solanacea, Nicotiana alata, N. glauca, Iochroma australe, Eriolarynx lorentzii*), every node that is an ancestor of that species was fixed to the SI state by setting its conditional likelihood of SC to zero in all tree likelihood calculations.

The estimate of q_{CI} is positive but small, and the estimates for other parameters (Table S4) are similar to those with $q_{CI} = 0$ (Table S1). Our main conclusions all hold under this analysis. The ordering of the speciation and extinction rates is the same (the posterior probability of $\mu_I < \lambda_I < \lambda_C < \mu_C$ is 0.99). Diversification is greater for SI lineages ($r_I > r_C$ with posterior probability 1.0). The equilibrium frequency of SI is positive ($p_{SI} > 0.30$ with posterior probability 0.90; the p_{SI} computation is somewhat more complicated than the expression given in the main text when reversals are allowed (*S17*)). Compared with the results in which q_{CI} is fixed to zero, the diversification advantage of SI is smaller when regain of SI is allowed but the equilibrium frequency of SI is slightly larger because the rate of loss of SI is smaller.

Results from whole-family analysis

On the full Solanaceae tree, pruned only to contain species with known SI status, overall sampling intensity is somewhat less than that of the "x=12" clade used in the main analyses: the sampling proportions are 0.123 for SI and 0.168 for SC. Inclusion in the tree is much less uniform outside of the "x=12" clade (e.g., *Petunia* is very well represented but the other deep genera are not), so we did not use the whole-family in the main analyses.

Nevertheless, it is worthwhile to ask if our results for the clade spanning *Solanum* and *Nicotiana* also hold for what is known of the entire family. We find that the ordering of the

speciation and extinction rates is not as clear (on the ML tree, the posterior probability of $\mu_I < \lambda_I < \lambda_C < \mu_C$ is only 0.61), but the transition rate q_{IC} is always much less than these rates, and all of our other conclusions hold. Specifically, with posterior probabilities of at least 0.99, $r_C < -0.09$, $r_I > 0.26$, $r_I > r_C + q_{IC}$, and $p_{SI} > 0.32$.

Figure S1: Phylogenetic relationships among 995 species of Solanaceae, rooted with five species of Convolvulaceae.

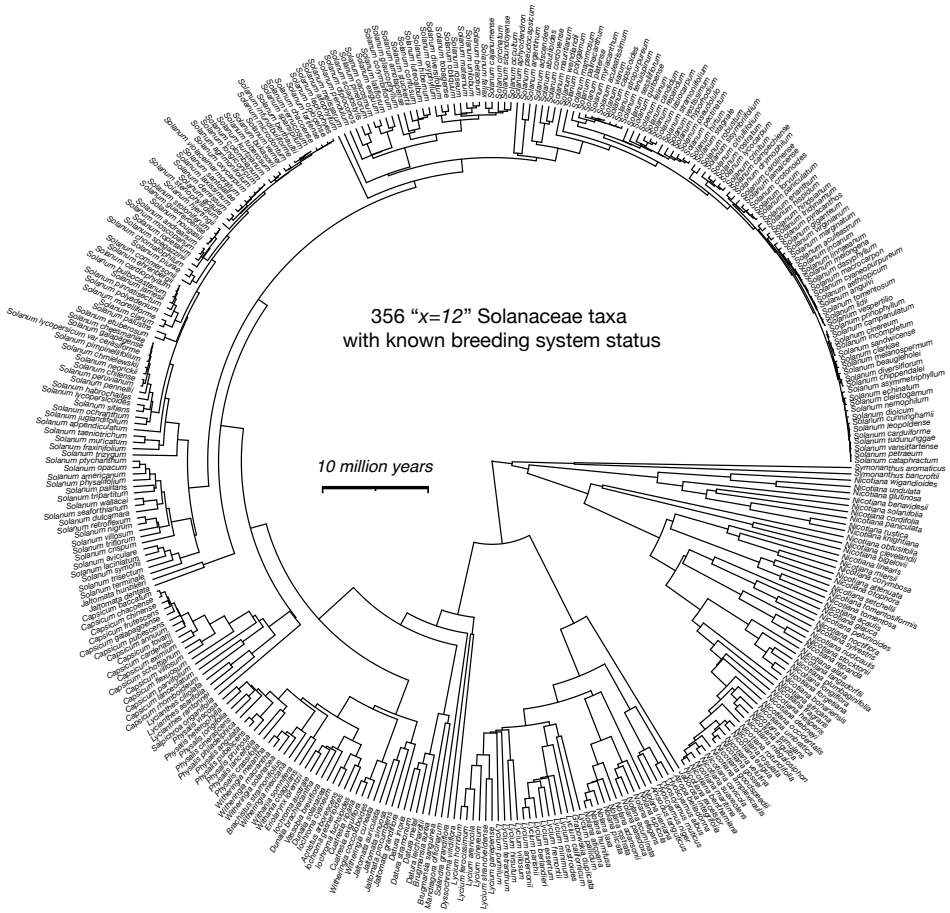

Figure S2: Phylogenetic relationships among 356 species of Solanaceae used in the analyses of character evolution. Branch lengths were obtained with penalized likelihood smoothing, implemented in *r8s* (*S7*), with the crown group age fixed at 36 Ma.

	mean	std dev	std err
λ_I	2.933	0.467	0.049
λ_C	5.199	0.738	0.063
μ_I	2.319	0.556	0.059
μ_C	5.433	0.757	0.065
q_{IC}	0.555	0.138	0.010
$\lambda_C - \lambda_I$	2.266	0.899	0.083
$\mu_C - \mu_I$	3.114	1.011	0.095
r_I	0.613	0.141	0.010
r_C	-0.234	0.075	0.002
$r_I - r_C$	0.847	0.200	0.012
$r_I - r_C - q_{IC}$	0.292	0.075	0.002
p_{SI}	0.345	0.043	0.002

Table S1: Summary of MCMC results from the ML tree, showing the mean, standard deviation, and standard error from time-series analysis. The first five quantities are the parameters directly estimated by the BiSSE analysis.

	λ_I	λ_C	μ_I	μ_C	q_{IC}
λ_I	—				
λ_C	-0.067	—			
μ_I	0.977	-0.142	—		
μ_C	-0.076	0.995	-0.165	—	
q_{IC}	-0.510	0.352	-0.677	0.414	—

Table S2: Parameter correlations on the ML tree. Correlations between q_{IC} and net diversification are 0.984 and -0.719 for r_I and r_C, respectively.

	λ_I	λ_C	μ_I	μ_C	q_{IC}	r_I	r_C
ML	0.467	0.738	0.556	0.757	0.138	0.141	0.075
BS	0.548	0.812	0.547	0.869	0.124	0.059	0.033

Table S3: Standard deviations of estimated parameters, computed from MCMC on the ML tree (upper row) and from the median values on the set of bootstrap trees (lower row).

	mean	std dev	std err
λ_I	3.031	0.486	0.013
λ_C	5.108	0.744	0.018
μ_I	2.516	0.563	0.015
μ_C	5.284	0.769	0.018
q_{IC}	0.484	0.128	0.002
q_{CI}	0.021	0.011	0.000
$r_I - r_C$	0.690	0.187	0.003
p_{SI}	0.355	0.044	0.000

Table S4: Summary of MCMC results from the ML tree when regain of SI is not prohibited.

Supporting references

S1. S. A. Smith, C. W. Dunn, *Bioinformatics* **24**, 715 (2008).

S2. S. A. Smith, J. M. Beaulieu, M. J. Donoghue, *BMC Evol Biol* **9**, 37 (2009).

S3. P. J. A. Cock, *et al.*, *Bioinformatics* **25**, 1422 (2009).

S4. *http://www.biosql.org*.

S5. K. Katoh, K.-i. Kuma, H. Toh, T. Miyata, *Nucleic Acids Res* **33**, 511 (2005).

S6. A. Stamatakis, *Bioinformatics* **22**, 2688 (2006).

S7. M. J. Sanderson, *Mol Biol Evol* **19**, 101 (2002).

S8. T. Paape, *et al.*, *Mol Biol Evol* **25**, 655 (2008).

S9. M. J. Benton, *The fossil record 2* (Chapman and Hall (London), 1993).

S10. S. Magallón, M. J. Sanderson, *Evolution* **55**, 1762 (2001).

S11. Y. Wang, *et al.*, *Genetics* **180**, 391 (2008).

S12. C. Mitter, B. Farrell, W. B. J, *Am Nat* **132**, 107 (1988).

S13. M. J. Sanderson, M. J. Donoghue, *Trends Ecol Evol* **11** (1996).

S14. R. E. Ricklefs, *Ecology* **87**, 2468 (2006).

S15. B. Igić, R. Lande, J. R. Kohn, *Int J Plant Sci* **169**, 93 (2008).

S16. J. C. Heilbuth, *Am Nat* **156**, 221 (2000).

S17. W. P. Maddison, P. E. Midford, S. P. Otto, *Syst Biol* **56**, 701 (2007).

S18. R. G. FitzJohn, W. P. Maddison, S. P. Otto, *Syst Biol* **58**, 595 (2009).

S19. C. Goodwillie, S. Kalisz, C. G. Eckert, *Ann Rev Ecol Evol Syst* **36**, 47 (2005).

S20. B. Igić, L. Bohs, J. R. Kohn, *Proc Natl Acad Sci USA* **103**, 1359 (2006).

S21. R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2009). ISBN 3-900051-07-0.

S22. R. G. FitzJohn, *diversitree: diversitree: comparative phylogenetic tests of diversification* (2010). R package version 0.4-5.

S23. M. Plummer, N. Best, K. Cowles, K. Vines, *coda: Output analysis and diagnostics for MCMC* (2010). R package version 0.13-5.