Example Data Analyses and Principal Results from *Species Selection Maintains Self-incompatibility in Solanaceae*[†]

Boris Igić & Emma E. Goldberg

840 W. Taylor St. (M/C067)
University of Illinois at Chicago
Chicago, IL 60607, U.S.A.

Contents

1 Introduction **2**

2 run-mcmc.R **3**

3 plot-mcmc.R **5**

4 Analyses of Character Evolution **7**
4.1 Character state encoding 7
4.2 Bayesian analysis 8
4.3 Results with regain of SI allowed 8

5 Phylogenetic Tree **8**
5.1 Alignment and tree construction pipeline 8
5.2 Relaxed molecular clock implementation 9

6 PDF Reprint of Wright and Barrett (2010) **13**

7 PDF Reprint of Goldberg et al. (2010) **15**

[†]Goldberg E.E., J.R. Kohn, R. Lande, K.A. Robertson, S.A. Smith, and B. Igić. 2010. Species selection maintains self-incompatibility. Science **330**:493–495. [DOI: 10.1126/science.1194513.]

1 Introduction

This document contains sufficient resources to independently obtain the main analyses and figures performed by Goldberg et al. (2010), give or take a random seed, a node-fixing procedure, and version changes to `diversitree`. Specifically, one can estimate rates of speciation, extinction, and transition rates associated with alternate breeding systems in the plant family Solanaceae. The alternate breeding systems that we examined are self-incompatibility (a form of obligate outrossing) and self-compatibility (potential for self-fertilization). The precision and accuracy of estimates are approximately as good as the ability of underlying models to approximate biological reality. In other words unclear, but we do believe reasonable[1].

The main result of these analyses implies that self-incompatible lineages can be maintained over evolutionary time because of their relatively higher net diversification rate, despite a very high rate of transitions to self-compatibility (fixation of self-compatible mutations within lineages). Within lineages (i.e. species), it is likely that self-compatible mutants, which commonly arise in all examined otherwise self-incompatible species, have a fluctuating but persistent advantage, and they are frequently fixed (the whole lineage makes a transition to self-compatibility). In the long term, however, something about self-compatibility fails to result in a positive net diversification rate—self-compatible lineages appear to be, on average, a dead end.

We assume that you have a passing familiarity with `R` programming language, and a working installation of both `diversitree` and the packages it requires. On the later point, it's fairly easy; all you have to do is type: `install.packages(diversitree)` in your R window.

The example analyses require two data files, which contain the ML tree and character state data, two R code files, and optionally, the provided original results file, which contains 10,000 Markov-chain Monte Carlo samples. You can download the entire packet from Igic lab website. The data files are separately available on DRYAD:

▶ `NicSolPLML.tre`, a newick formatted ultrametric phylogenetic tree with 356 species, used in our analyses –
http://dev.datadryad.org/bitstream/handle/10255/dryad.1890/NicSolPLML.tre?sequence=1.

▶ `NicSolstatus.csv`, a character state file in Comma-separated Values (.csv) format –
http://dev.datadryad.org/bitstream/handle/10255/dryad.1889/NicSolstatus.csv?sequence=1.

▶ `run-mcmc.R`, a text file containing simplified step-by-step commented R code necessary to obtain maximum likelihood and posterior estimates of character evolution and diversification parameters. (You need the tree and data files, above, to perform these analyses.)

▶ `plot-mcmc.R`, a text file containing R code to plot the obtained posterior estimates.

▶ `mcmc-results-orig-10k.dat`, a text file with the original data.

Finally, the last sections of this file include Supplementary Information and the original paper PDF (please do not disseminate). Methodological nerds and purists may not be happy with these analyses, but they appear powerful when properly employed (FitzJohn et al. 2009).

[1]The text in this vignette was compiled by Boris Igić, and the included statements do not necessarily reflect the opinion of other co-authors

2 run-mcmc.R

```
################################
###  run-mcmc-bodega.R       ###
### MCMC analysis on ML tree ###
### (Goldberg et al. 2010)   ###
################################

### DATA PREPARATION ###

library(diversitree)

# Import Tree
# Maximum likelihood tree, smoothed with penalized likelihood.
# Be sure to place the NicSolPLML.tre file into the right directory or paste the directory path
 immediately preceeding the file name.

tree <- read.tree(file="NicSolPLML.tre")
i <- which(tree$node.label == "")
tree$node.label[i] = paste("node", i, sep="")

# Import Tip States (see README.txt)
# 0 = SI, 1 = SC
# states 2:5 are more difficult to encode, but 2,3,& 5 are all mostly SI, and 4 is mostly SC.

dat <- read.csv("NicSolstatus.csv")
states <- dat$Status
names(states) <- dat$Species
states[states %in% c(2, 3, 5)] <- 0
states[states == 4] <- 1
table(states)

# Adjustment for Incomplete Sampling
# The proportion of species sampled for each state (SI,SC); SI species were relatively more likely to
 be found on the tree, although both states were sampled around 15% of the total estimated to occur
 in the family (from the character state data that includes species without phylogenetic data).

sampling.f <- c(0.1628325, 0.1500414)

### ANALYSES ###

# Initial Parameter Values
# lambda.SI = speciation rate for SI
# lambda.SC = speciation rate for SC
# mu.SI = extinction rate for SI
# mu.SC = extinction rate for SC
# q.IC = transition rate from SI to SC
# q.CI, the reversal rate, is not estimated here; extra credit exercise opportunities abound.

par.start <- c(3.1, 5.2, 2.6, 5.4, 0.5)
names(par.start) <- c("lambda.SI", "lambda.SC", "mu.SI", "mu.SC", "q.IC")
```

```
# The Likelihood Function

lnL.full <- make.bisse(tree, states, sampling.f=sampling.f)
lnL.constr <- constrain(lnL.full, q10 ~ 0)

# Find ML
# It is much better to avoid single point estimation,
# when the process has so much uncertainty,
# but here is the ML for two particularly interesting models,
# (full 6-parameter, and constrained, no-reversals 5-parameter)
# NOTE: we fixed some nodes to SI in original analyses (see paper)

find.mle(lnL.full,p) # lnL = -941.9432
find.mle(lnL.constr,p) # lnL = -949.2878

# Run the MCMC
# This takes a long time, so it's generally safer to run it in batches, writing out the results along the
 way.
# Below, we write results in user-defined chunks = ntotal / nstep, to the same mcmc-results.dat file.

outfile <- "mcmc-results.dat"
cat("i", names(par.start), "p\n", file=outfile)

ntotal <- 100000
nstep  <- 1000
nbatch <- ntotal/nstep

for (i in seq(nbatch))
{
   ans <- mcmc(lnL.constr, par.start, nsteps=nstep,
            w=c(1.75, 1.75, 1.75, 1.75, 0.75),
            upper=c(10, 10, 10, 10, 3), lower=0,
            prior=make.prior.exponential(0.3))
   write.table(ans, file=outfile, quote=F, row.names=F, append=F, col.names=T)
   par.start <- as.numeric(ans[nstep,2:6])
}
```

3 plot-mcmc.R

```
################################
###  plot-mcmc-bodega.R      ###
### view MCMC results          ###
### (Goldberg et al. 2010)     ###
################################

library(diversitree)   # for profiles.plot()

# Reads the results from the saved file
# In case of errors, check file location, or use this code to browse for it:
# dat <- read.table(file=file.choose(), header=T)

dat <- read.table("mcmc-results-orig-10k.dat", header=T)

cex <- 1.4
nbreaks <- 200

pdf("mcmc-hists.pdf", width=10, height=10)
par(mfrow=c(2,2))

### all params ###

legend.text1 <- c(expression(q[IC]), "", expression(lambda[I]),
                  expression(lambda[C]), expression(mu[I]), expression(mu[C]))
col1 <- c("black", "white", "#ff0000", "#cccc00", "#0000ff", "#00aaff")
col1fill <- diversitree:::add.alpha(col1, alpha=0.55)
profiles.plot(dat[c(6,2,3,4,5)], col.line=col1[-2], col.fill=col1fill[-2],
              xlab="", ylab="", n.br=nbreaks, cex.axis=cex)
title(main="parameter estimates", xlab="rate", ylab="density", cex.lab=cex,
      cex.main=cex, font.main=3)
legend("topright", legend.text1, text.col=col1, col=col1, pch=22,
       pt.bg=col1fill, bty="n", cex=cex, ncol=3)

### net diversification ###

dat.r <- data.frame(rSI=dat$lambda.SI - dat$mu.SI,
                    rSC=dat$lambda.SC - dat$mu.SC)
legend.text2 <- c(expression(r[I] == lambda[I] - mu[I]),
                  expression(r[C] == lambda[C] - mu[C]))
col2 <- c("#7d26cd", "#00cdcd")
col2fill <- col2
profiles.plot(dat.r, col.line=col2, col.fill=col2fill, xlab="", ylab="",
              n.br=nbreaks, cex.axis=cex)
legend("topright", legend.text2, text.col=col2, col=col2, pch=22,
       pt.bg=col2fill, bty="n", cex=cex)
title(main="net diversification", xlab="rate", ylab="density", cex.lab=cex,
```

```
    cex.main=cex, font.main=3)

### fate of SI ###

dat.fate <- data.frame(fate=dat.r$rSI - dat.r$rSC - dat$q.IC)
legend.text3 <- expression(r[I] - r[C] - q[IC])
col3 <- "black"
col3fill <- diversitree:::add.alpha(col3, alpha=0.3)
profiles.plot(dat.fate, col.line=col3, col.fill=col3fill, xlab="", ylab="",
            n.br=nbreaks, cex.axis=cex, xlim=c(0,max(dat.fate)))
title(main="fate of SI", xlab="rate", ylab="density", cex.lab=cex, cex.main=cex,
    font.main=3)
legend("topright", legend.text3, text.col=col3, col=col3, pch=22,
       pt.bg=col3fill, bty="n", cex=cex)

### equilibrium frequency of SI ###

dat.eqfreq <- data.frame(eqfreq = 1 - dat$q.IC / (dat.r$rSI - dat.r$rSC))
legend.text4 <- expression(p[SI])
profiles.plot(dat.eqfreq, col.line=col3, col.fill=col3fill, xlab="", ylab="",
            n.br=nbreaks, cex.axis=cex, xlim=c(0,max(dat.eqfreq)))
title(main="equilibrium frequency of SI", xlab="frequency", ylab="density",
    cex.lab=cex, cex.main=cex, font.main=3)
legend("topright", legend.text4, text.col=col3, col=col3, pch=22,
       pt.bg=col3fill, bty="n", cex=cex)

dev.off()
```

4 Analyses of Character Evolution

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 (Mitter et al. 1988; Sanderson and Donoghue 1996; Ricklefs 2006). These methods consider only clades in which most tips are in the same state and assume that ancestors within the clade had that state. We instead employ the recently-developed BiSSE method for estimating from a phylogeny speciation and extinction rates that depend on character states (Maddison et al. 2007). We include the correction for incomplete sampling (FitzJohn et al. 2009); the probability of being sampled in the tree is 0.162 for self-incompatible (SI) species and 0.150 for self-compatible (SC) species.

Given that best-fitting model accounts for different speciation, extinction, and transition rates, explicitly modeling these parameters allows us to more accurately quantify the effects of state-dependent diversification and state transitions.

4.1 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 (Goodwillie et al. 2005). We use the same general principles as described elsewhere (Igić et al. 2006). 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 (Igić et al. 2006)), 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, named for their uniform haploid number of chromosomes. 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.

For transparency and user convenience, the character state file `NicSolstatus.csv` contains original encoding (0-5), which breaks down as follows:

0: SI (n=116)
1: SC (n=196)
2: SI & SC (n=17; does not distinguish whether any pops are fixed or not)
3: SI & SC & Dioecy (n=1; *Lycium californicum*)
4: Dioecy (n=25; individuals not hermaphrodites; express separate sexes)
5: SI & Dioecy (n=1; *Solanum wendlandii*)

The analyses presented were performed on the following combining scheme, essentially separating any self-incompatibility from none:

0: any SI (combines 0,2,3,5 above)
1: no SI (combines 1,4 above)

4.2 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 (FitzJohn et al. 2009): 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* (FitzJohn 2012). 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.

4.3 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 incorporating 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 4) are similar to those with $q_{CI} = 0$ (Table 1). 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 (Maddison et al. 2007)). 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.

5 Phylogenetic Tree

5.1 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 (Smith and Dunn 2008;

Smith et al. 2009). We implemented this pipeline in *Python* (v2.5) with the *BioPython* (v1.48) module (Cock et al. 2009) and using the *BioSQL* (v1.0.1) database schema (Lapp et al. 2010). 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 (Katoh et al. 2005), which resulted in a 1084-taxon, 16,405bp (eight-gene) matrix. Gene-partitioned phylogenetic reconstructions were performed using RAxML (v7.0.4) (Stamatakis 2006). 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.

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. **??**).

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.

5.2 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 (Sanderson 2002). 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) (Sanderson 2002). 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 (Paape et al. 2008). This dating procedure relies on the use of two key constraints derived from dating and identification of *Solanum*- and *Physalis*-like seeds (Benton 1993). Those two dates were used to anchor a normally distributed age prior ($\mu = 10.0$ Ma, $\sigma = 4.0$ Ma; 95% CI = [3.4, 16.6]), along with the date for stem group age of Solanaceae (divergence from Convolvulaceae; $\mu = 52$ Ma, $\sigma = 5.2$ Ma; 95% CI = [43, 60]). The divergence dates we obtained are comparable to others recovered with different strategies [e.g.,(Magallón and Sanderson 2001; Wang et al. 2008)].

	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 1: 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 2: 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.

Literature Cited

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

Cock, P. J. A., T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski, and M. J. L. de Hoon, 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422–1423.

FitzJohn, R. G., 2012. Diversitree: Comparative phylogenetic analyses of diversification in R. Methods in Ecology and Evolution 3:1084–1092.

FitzJohn, R. G., W. P. Maddison, and S. P. Otto, 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. Systematic Biology 58:595–611.

Goldberg, E. E., J. R. Kohn, R. Lande, K. A. Robertson, S. A. Smith, and B. Igić, 2010. Species selection maintains self-incompatibility. Science 320:493–495.

Goodwillie, C., S. Kalisz, and C. G. Eckert, 2005. The evolutionary enigma of mixed mating

	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 3: 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 4: Summary of MCMC results from the ML tree when regain of SI is not prohibited.

systems in plants: Occurrence, theoretical explanations, and empirical evidence. Ann Rev Ecol Evol Syst 36:47–79.

Igić, B., L. Bohs, and J. R. Kohn, 2006. Ancient polymorphism reveals unidirectional breeding system shifts. Proceedings of the National Academy of Sciences USA 103:1359–1363.

Katoh, K., K.-i. Kuma, H. Toh, and T. Miyata, 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. Nucleic Acids Res 33:511–8.

Lapp, H., A. Mackey, E. Stupka, and E. Birney, 2010. *http://www.biosql.org.*

Maddison, W. P., P. E. Midford, and S. P. Otto, 2007. Estimating a binary character's effect on speciation and extinction. Systematic Biology 56:701–710.

Magallón, S. and M. J. Sanderson, 2001. Absolute diversification rates in angiosperm clades. Evolution 55:1762–1780.

Mitter, C., B. Farrell, and W. B. J, 1988. The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? The American Naturalist 132:107–128.

Paape, T., B. Igic, S. D. Smith, R. Olmstead, L. Bohs, and J. R. Kohn, 2008. A 15-Myr-old genetic bottleneck. Molecular Biology and Evolution 25:655–63.

Ricklefs, R. E., 2006. Global variation in the diversification rate of passerine birds. Ecology 87:2468–2478.

Sanderson, M. J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Molecular Biology and Evolution 19:101–109.

Sanderson, M. J. and M. J. Donoghue, 1996. Reconstructing shifts in diversification rates on phylogenetic trees. Trends in Ecology and Evolution 11:15–20.

Smith, S. A., J. M. Beaulieu, and M. J. Donoghue, 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC Evol Biol 9:37–48.

Smith, S. A. and C. W. Dunn, 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics 24:715–716.

Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

Wang, Y., A. Diehl, F. Wu, J. Vrebalov, J. Giovannoni, A. Siepel, and S. D. Tanksley, 2008. Sequencing and comparative analysis of a conserved syntenic segment in the solanaceae. Genetics 180:391–408.

Mezić *et al*. aim to find not just barriers but also regions that are particularly active in mixing the fluid. Such regions can lead, for instance, to a more rapid breakdown of oil into less harmful microdroplets. Their models rely on the concept of hyperbolicity of a flow: Hyperbolic flows are good at stretching fluid parcels. As the flow in the Gulf of Mexico is not periodic, and we have limited data, they introduce mesohyperbolicity: hyperbolicity on average. This reflects the fact that a typical fluid parcel may meet regions that are both favorable and adverse to mixing over a finite stretch of time. Mesohyperbolicity is a generalization to finite time of the concept of a Floquet analysis for periodic flows, where we can deduce the long-time behavior by examining a single period.

Mezić *et al*. were successful in describing how the oil was dispersed after the Deepwater Horizon oil leak. The oil spill occurred just as interest in these dynamical systems ideas is the subject of renewed interest. But hopefully we will be better prepared when another oil leak occurs, and applied mathematicians can work in conjunction with cleanup crews to predict where efforts are best focused.

References

1. T. J. Crone, M. Tolstoy, *Science*; published online 23 September 2010 (10.1126/1195840).
2. I. Mezić *et al*., *Science* **330**, 486 (2010); published online 2 September 2010 (10.1126/science.1194607).
3. V. I. Arnold, *C. R. Acad. Sci. Paris* **261**, 17 (1965).
4. M. Henon, *C. R. Acad. Sci. Paris* **262**, 312 (1966).
5. H. Aref, *J. Fluid Mech.* **143**, 1 (1984).
6. E. N. Lorenz, *J. Atmos. Sci.* **20**, 130 (1963).
7. G. Haller, A. C. Poje, *Physica D* **119**, 352 (1998).
8. G. Haller, *Physica D* **149**, 248 (2001).
9. G. Haller, *Phys. Fluids* **14**, 1851 (2002).

10.1126/science.1197554

EVOLUTION

The Long-Term Benefits of Self-Rejection

A trait that prevents self-fertilization in plants appears to promote evolutionary diversification.

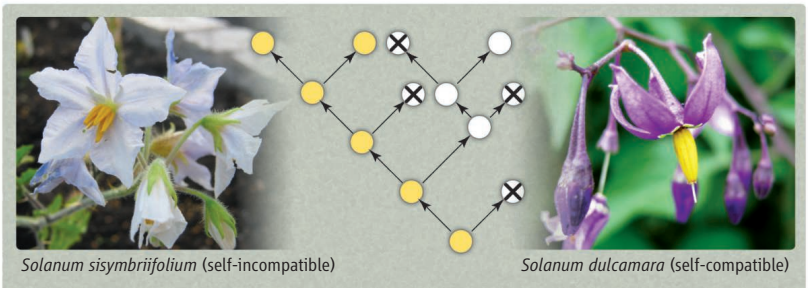
Stephen I. Wright and Spencer C. H. Barrett

Charles Darwin provided the key explanation for the striking adaptive fit of organisms to their environments; traits that confer a greater chance of survival and reproduction spread through populations by the process of evolution by natural selection. But natural selection has no foresight; only immediate advantages matter. So the same trait that confers higher fitness in the short term could also increase the chance that a species eventually becomes extinct, or decrease the chance that it will form new species. Conversely, another trait may enhance a species' ability to diversify, leading to more species with that trait.

Many evolutionary biologists refer to such trait-specific effects on species diversification as "species selection," an evolutionary process that occurs above the species level (*1*). Biologists mostly accept the concept, but its general importance has not been clear. On page 493 of this issue, however, Goldberg *et al*. (*2*) provide a compelling case for the importance of species selection. In a study of hermaphroditic plants, in which individuals function as both males and females, they show that one trait—the ability of an individual to prevent self-fertilization by recognizing and rejecting its own pollen—appears to have strongly affected species diversification in the nightshade family, Solanaceae.

Hermaphroditic plants exhibit either self-compatibility (SC), which means an individual can act as both mother and father to its own offspring, or self-incompatibility (SI), which enforces outcrossing. SI is a genetically controlled physiological mechanism that is mediated by protein-protein interactions. It is analogous to a lock-and-key mechanism, in which the maternal parent recognizes pollen grains (or pollen tubes) that share alleles, initiating a biochemical pathway that rapidly blocks fertilization (*3*). The primary advantage of SI is that it avoids the harmful effects of inbreeding, particularly self-fertilization. However, SI comes at a cost because plants produce offspring that are just 50% related to them, rather than 100%, thus reducing transmission of their own genes to the next generation. Also, fertility can be severely compromised if pollinators or mates are unavailable. Because of these disadvantages, flowering plants have lost SI on numerous occasions during their evolutionary history, resulting in SC. Because of the complexity of SI, the trait is rarely, if ever, regained. This implies that SI may be on its way to extinction, yet at least half of all living flowering plant species are SI. Perhaps SI has some macroevolutionary advantage allowing it to be maintained by species selection.

Evaluating the importance of species selection is difficult, because it is tricky to estimate rates of character transition (such as a switch from SI to SC) separately from rates of speciation and extinction. Recent statistical advances, however, now allow researchers to analyze this question by using evolutionary trees (*4*, *5*). Using extensive phylogenetic information on species relationships in Solanaceae (which includes important crop plants, such as tomato, potato, and tobacco), the authors reconstructed the evolutionary history of transitions from SI to SC. They relied on well-resolved clades (groups of species



Know thyself. Species selection maintains self-incompatibility. Self-incompatible species (yellow circles) frequently experience evolutionary transitions to self-compatibility (white circles), leading to the potential loss of the trait over evolutionary time. This is balanced by a higher rate of extinction of self-compatible species (crosses), causing the stable maintenance of both types over long periods of evolutionary time.

PHOTO CREDITS: M. VALLEJO MARÍN

Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada. E-mail: stephen.wright@utoronto.ca

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

that share common ancestry). In this family, the evolutionary transition from SI to SC has occurred independently on numerous occasions. This is important because it improves statistical power and increases confidence that the transition itself is closely associated with differences in species diversification.

Goldberg *et al*. showed that SI lineages have a higher net diversification rate——a key quantity that determines the rate of increase in species numbers—than SC lineages. One key finding of their study is that this difference is not due to a higher rate of speciation for SI lineages; in fact, inferred speciation rates were higher in SC lineages. Instead, SC lineages have higher extinction rates than SI lineages. As a result, SI lineages have higher net diversification rates that apparently have been sufficient to counter-balance the repeated loss of the trait, allowing it to be maintained over evolutionary time (see the figure). The results provide a convincing macroevolutionary explanation for how SI has persisted over tens of millions of years despite its repeated breakdown to SC.

Given these results, a number of outstanding questions remain. Most importantly, why should SC plants that have the potential to self-fertilize experience higher rates of extinction? This study did not address the issue of how much "selfing" occurs in SC species, but the process is commonly associated with reduced genetic diversity and lower rates of recombination. This reduces the chance of eliminating deleterious mutations and can decrease opportunities for adaptive mutations to succeed (*6*), both of which can increase the probability of extinction. Because some of the species included in the study probably self-fertilize at high rates, it is possible that the actual driver of differential diversification is not SI per se but the rate of self-fertilization. If the researchers had been able to use actual data on rates of cross- and self-fertilization (rather than only classifying plants as SC or SI), even stronger differences in diversification may have been found. However, obtaining this information for the many species included in this study would be a Herculean task.

Different approaches to studying the evolutionary consequences of selfing have provided conflicting results. Molecular work on protein evolution has found little evidence that selfing populations accumulate harmful mutations (*7*, *8*). In contrast, phylogenetic studies indicate that selfing species commonly produce short branches on evolutionary trees and appear to be more prone to extinction (*9*, *10*). One possible explanation is that molecular studies have focused on too coarse a level to detect the predicted differences in the efficacy of selection. New approaches that enable simultaneous estimates of the strength of positive and negative selection are likely to be more powerful (*11*, *12*). Also, with few exceptions (*8*), molecular evolutionary studies have not been done with a large number of species, making it difficult to detect repeated declines in fitness of selfing lineages. Finally, SC lineages may experience higher extinction rates for reasons unrelated to mutational decay. For example, SI species often occur in relatively large, often long-lived, populations, and these demographic properties may make them more likely to persist over longer time scales. The causes of differences in diversification rates among lineages remain a central issue in evolutionary biology, but this illuminating study indicates that we should not ignore macroevolutionary processes in trying to understand the maintenance of adaptations and biodiversity.

References

1. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer Associates, Sunderland, MA, 2004).
2. E. E. Goldberg *et al*., *Science* **330**, 493 (2010).
3. V. Franklin-Tong, Ed., *Self-Incompatibility in Flowering Plants: Evolution, Diversity, and Mechanisms* (Springer, Berlin Heidelberg, 2008).
4. W. P. Maddison *et al*., *Syst. Biol.* **56**, 701 (2007).
5. R. G. FitzJohn *et al.*, *Syst. Biol.* **58**, 595 (2009).
6. S. Glémin, *Genetics* **177**, 905 (2007).
7. S. I. Wright *et al.*, *Int. J. Plant Sci.* **169**, 105 (2008).
8. J. Escobar *et al., Evolution AOP*, published online 22 June 2010; 10.1111/j.1558-5646.2010.01045.x (2010).
9. D. J. Schoen*et al., Evolution* **51**, 1090 (1997).
10. N. Takebayashi, P. L. Morrell, *Am. J. Bot.* **88**, 1143 (2001).
11. P. D. Keightley, A. Eyre-Walker, *Genetics* **177**, 2251 (2007).
12. T. Slotte *et al., Mol. Biol. Evol.* **27**, 1813 (2010).

10.1126/science.1198063

EPIDEMIOLOGY

Environment and Disease Risks

Stephen M. Rappaport and Martyn T. Smith

A new paradigm is needed to assess how a lifetime of exposure to environmental factors affects the risk of developing chronic diseases.

Although the risks of developing chronic diseases are attributed to both genetic and environmental factors, 70 to 90% of disease risks are probably due to differences in environments (*1*–*3*). Yet, epidemiologists increasingly use genome-wide association studies (GWAS) to investigate diseases, while relying on questionnaires to characterize "environmental exposures." This is because GWAS represent the only approach for exploring the totality of any risk factor (genes, in this case) associated with disease prevalence. Moreover, the value of costly genetic information is diminished when inaccurate and imprecise environmental data lead to biased inferences regarding gene-environment interactions (*4*). A more comprehensive and quantitative view of environmental exposure is needed if epidemiologists are to discover the major causes of chronic diseases.

An obstacle to identifying the most important environmental exposures is the fragmentation of epidemiological research along lines defined by different factors. When epidemiologists investigate environmental risks, they tend to concentrate on a particular category of exposures involving air and water pollution, occupation, diet and obesity, stress and behavior, or types of infection. This slicing of the disease pie along parochial lines leads to scientific separation and confuses the definition of "environmental exposures." In fact, all of these exposure categories can contribute to chronic diseases and should be investigated collectively rather than separately.

To develop a more cohesive view of environmental exposure, it is important to recognize that toxic effects are mediated through chemicals that alter critical molecules, cells, and physiological processes inside the body. Thus, it would be reasonable to consider the "environment" as the body's internal chemical environment and "exposures" as the amounts of biologically active chemicals in this internal environment. Under this view, exposures are not restricted to chemicals (toxicants) entering the body from air, water, or food, for example, but also include chemicals produced by inflammation, oxidative stress, lipid peroxidation, infections, gut flora, and other natural processes (*5*, *6*) (see the figure). This internal chemical environment continually fluctuates during life due to changes in external and internal sources, aging, infections, life-style, stress, psychosocial factors, and preexisting diseases.

The term "exposome" refers to the totality of environmental exposures from conception onwards, and has been proposed to be a

School of Public Health, University of California, Berkeley, CA 94720–7356, USA. E-mail: srappaport@berkeley.edu

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

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

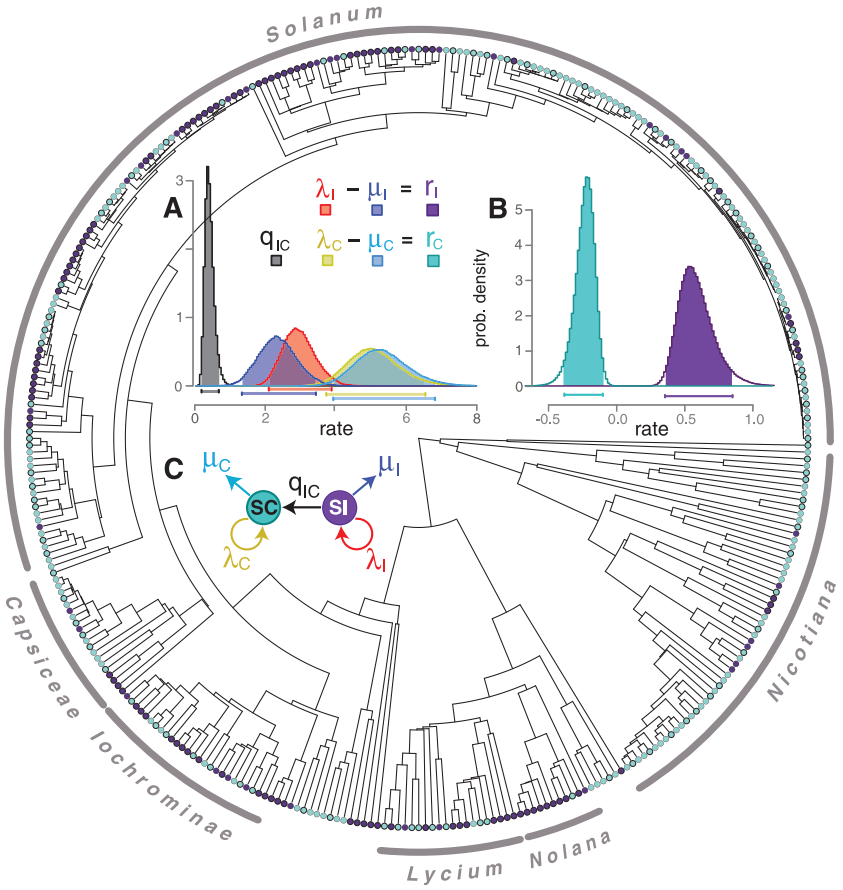


Fig. 1. Maximum likelihood tree of phylogenetic relationships among 356 species of Solanaceae. Higher ranks are indicated around the perimeter of the tree. Purple and turquoise tip colors denote SI and SC extant species, respectively. The root age is 36 million years. Inset panels display posterior probability 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*).

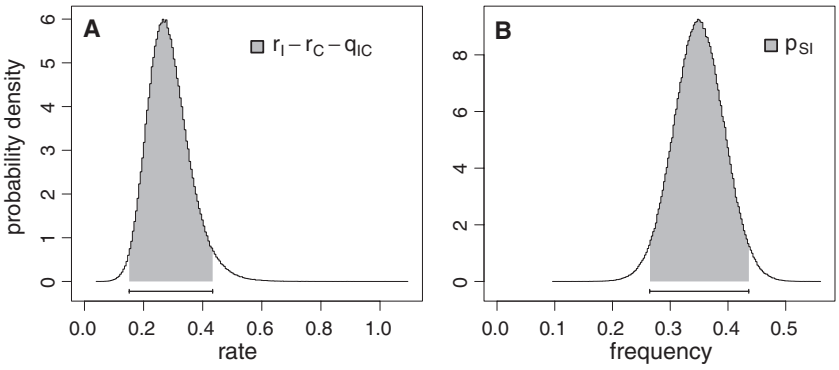


Fig. 2. Further analyses of the posterior rate distributions in Fig. 1. (**A**) The diversification rate of SI is sufficiently higher than that of SC for it to persist deterministically (shown by positive values here) despite frequent and irreversible loss. (**B**) The equilibrium frequency of SI is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