

# How and when did *Arabidopsis thaliana* become highly self-fertilising

Deborah Charlesworth<sup>1\*</sup> and Xavier Vekemans<sup>2</sup>

## Summary

Changes in breeding system are a regular evolutionary change in plants, as self-fertilisation is often advantageous, particularly for weedy and colonising species. The adoption of *Arabidopsis thaliana* as a plant model species has led to interest in how self-incompatibility was lost so that this species became highly inbreeding. Molecular evolutionary approaches have recently focused on investigating two loci involved in the incompatibility recognition process in related *Arabidopsis* species; non-functional copies of these genes still exist in *A. thaliana*. New work studying polymorphism at these loci found strikingly low diversity at one of them, suggesting that spread of a mutation in this gene might have caused self-compatibility in an ancestor of *A. thaliana*.<sup>(1)</sup> However, it is difficult to be sure of the time when the selfing habit evolved in the lineage that led to *A. thaliana*. *BioEssays* 27:472–476, 2005. © 2005 Wiley Periodicals, Inc.

## Introduction

Although *Arabidopsis thaliana* is a very-well-studied model plant, many important things about its natural biology and evolutionary history are still not well known. In the greenhouse, it is highly self-fertilising, with no trace of self-incompatibility<sup>(2)</sup> though many members of its family, the Brassicaceae, are self-incompatible,<sup>(3)</sup> including closely related *Arabidopsis* species, such as *A. lyrata* (also called *A. petraea* and *Cardaminopsis petraea*)<sup>(4–9)</sup> and *A. halleri*. In nature, *A. thaliana*'s selfing rate is known only from studies of northern European populations, in these its estimated selfing rate is well above 95%.<sup>(10–12)</sup>

Compared with its outcrossing relatives, *A. thaliana* exhibits several changes commonly observed in weedy inbreeding species. Changed flower morphology like that in *A. thaliana* (small, unscented flowers with anthers and stigmas close

together)<sup>(13)</sup> is common in inbreeding plants,<sup>(14)</sup> and can evolve rapidly,<sup>(15)</sup> but there are also pervasive genome changes, implying considerable evolutionary time, unless these changes are much quicker than we think. *A. thaliana* has a reduced chromosome number compared with related species, due to chromosome fusions,<sup>(16)</sup> a common change in inbreeders.<sup>(17)</sup> Its genome size is small<sup>(18)</sup> (perhaps because rapid growth and early flowering are advantageous for a weed),<sup>(19)</sup> with reduced transposable element abundance,<sup>(20)</sup> shorter introns in many genes.<sup>(21)</sup>

It is therefore of great interest to know when *A. thaliana* became an inbreeder. The common ancestor of *A. thaliana* and *A. lyrata* (an estimated ~5 million years ago) was presumably outcrossing and self-incompatible, but determining at what point the *A. thaliana* lineage lost self-incompatibility is a difficult question. One approach that could work if the time is very recent is to study the genes involved in self-incompatibility.

## Studying genes involved in self-incompatibility: functional and molecular evolutionary approaches

Self-pollen recognition in the self-incompatible species of Brassicaceae involves a pollen ligand and a membrane-anchored pistil receptor, encoded by two linked “S-genes”. Molecular interaction between these proteins triggers a cascade of reactions leading to inhibition of pollen germination on stigmas.<sup>(22,23)</sup> If self-compatibility in *A. thaliana* evolved by a loss of function of either the pistil receptor or the pollen ligand, the new (non-functional) mutant allele should spread through the species as the lineage of self-compatible plants increases in frequency, leading to a single sequence at that locus—a “selective sweep”—resulting in zero or low diversity throughout the locus. Detecting such events by molecular evolutionary approaches requires information from functional studies (including testing which component is changed in mutants or species that have lost functional self-incompatibility). Indeed, transgenic experiments in which functional incompatibility genes from the related outcrosser, *A. lyrata*, are expressed in *A. thaliana*,<sup>(2,22)</sup> has yielded some evidence that *A. thaliana* evolved self-compatibility by a mutation in one of the incompatibility genes, rather than in a gene involved in the downstream functions needed for rejection of incompatible pollen.

<sup>1</sup>Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh, UK.

<sup>2</sup>Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8016, Université de LILLE, Villeneuve d'Ascq, France.

\*Correspondence to: Deborah Charlesworth, Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Ashworth Lab. King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK. E-mail: deborah.charlesworth@ed.ac.uk

DOI 10.1002/bies.20231

Published online in Wiley InterScience (www.interscience.wiley.com).

Plant *S*-loci are highly polymorphic genes, which makes the molecular evolutionary approach particularly promising for this case (if variability is high in the ancestor, the loss of diversity due to the selective sweep will be easy to detect). Outcrossing *Arabidopsis* species have numerous, highly diverged alleles of the stigma cell surface receptor gene, *SRK*,<sup>(5,7)</sup> and the recently identified pollen ligand gene *SCR*<sup>(5)</sup> is also very polymorphic (the best estimates for *SCR* in Brassica, show huge nucleotide diversity, at least 30% for synonymous sites, i.e. on average, allele pairs differ at roughly one in three nucleotides; diversity is even higher for non-synonymous sites).<sup>(24)</sup> The high polymorphism in the *S*-locus genome region (see Fig. 1) may be associated with low recombination. This would preserve linkage between alleles at the two loci—recombination will create haplotypes with the pollen allele from one incompatibility type and the pistil allele of another, leading to self-compatibility,<sup>(25)</sup> and allow changes in gene arrangement, making recombination more unlikely. Gene order, number and orientation are indeed polymorphic in the two *A. lyrata* haplotypes that have been physically mapped,<sup>(5)</sup> and there is some evidence for suppressed recombination in the *SRK* gene.<sup>(9)</sup>

Identification of the incompatibility loci in *A. lyrata* opened the way for searching for the homologous loci in *A. thaliana*.<sup>(5)</sup> Even though *A. thaliana* has no trace of incompatibility, the *S*-loci in the Col-0 strain that has been completely sequenced are not deleted; both pollen and pistil genes are, however, non-functional.<sup>(5)</sup> In *SRK*, a single nucleotide substitution introduces a stop codon near the beginning of exon 4 (leading to an incomplete protein), while three *SCR* copies, all incomplete, are present. The fact that the *S*-loci are not deleted suggests a fairly recent loss of incompatibility. However, transgenic experiments introducing *A. lyrata* *SRK* and *SCR* genes into *A. thaliana* fail to produce incompatibility reactions in many strains, showing that genes for the downstream pathway functions needed for rejection of incompatible pollen are also non-functional, which might suggest that self-compatibility evolved quite long ago. Against this, however, strain C24 gains incompatibility in the transgenic tests, and therefore appears to have a complete functional downstream pathway.<sup>(2)</sup> It is thus uncertain whether an *S*-gene or a downstream gene mutation caused the evolution of *A. thaliana*'s self-compatibility, or how long ago.

The results from strain C24 are encouraging that mutations at one or other of the *S*-loci may be the cause of *A. thaliana*'s self-compatibility. In gel blots, *SCR* sequences were not detected in C24,<sup>(2)</sup> while its *SRK* gives a band pattern seen only in one other strain, Mt-0, and neither strain yields a product with Col-0 *SRK* PCR primers,<sup>(1)</sup> so they may both have missing or non-functional *SRK* and *SCR*. Detailed molecular evolutionary study of these genes in *A. thaliana* might thus be illuminating. Such a study has now been done, and the results are indeed interesting.<sup>(1)</sup>

### Very low diversity at the pollen ligand gene

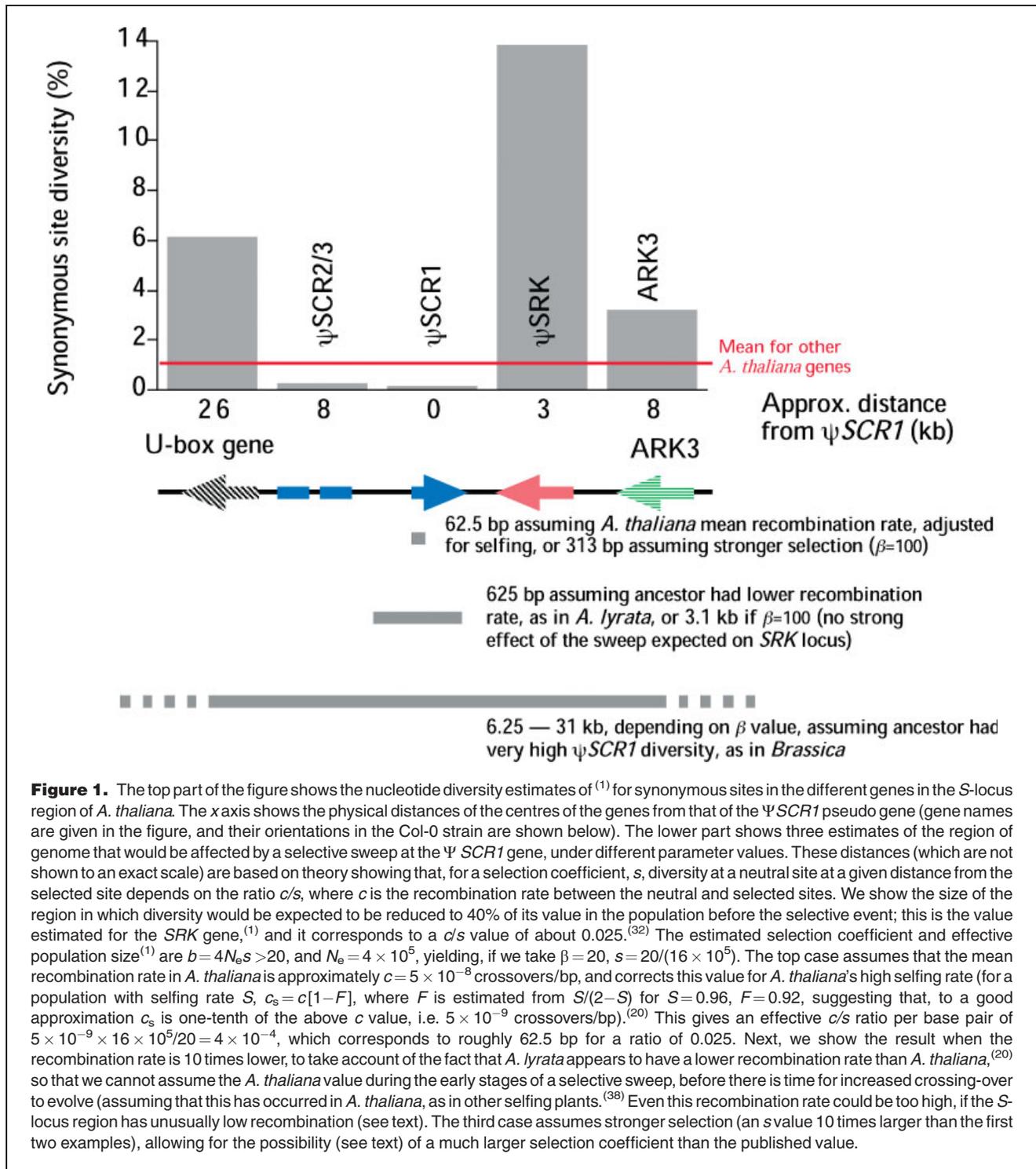
The selective sweep event hypothesised above will result in a self-compatible population with, initially, a single allele at the mutated locus. After some time, mutations will arise at the locus, restoring some diversity, but will only slowly obscure the imprint of the event, which remains detectable for some period of evolutionary time. Now polymorphism in the *A. thaliana* *S*-gene region has been studied, to try to assess whether a mutation in one of the component *S*-loci is indeed likely to have been the cause of loss of SI.<sup>(1)</sup> Strikingly low diversity was found at *SCR1*, the least incomplete of the three *SCR* loci (nucleotide diversity per silent site is estimated to be 0.12%, compared with 0.87% based on published data for 18 *A. thaliana* genes). Detailed modelling can test whether low observed diversity could be explained without any selection (in which case one would conclude that some other diversity-reducing event must be responsible, such as a population bottleneck). Otherwise, selection can be inferred.<sup>(26)</sup> This analysis suggests very recent, but very weak, selection at the *SCR1* locus, and estimates that this happened as recently as a few hundred thousand years ago.

A recent origin of selfing is consistent with theories that suggest a short evolutionary lifespan for selfers<sup>(27–29)</sup> as for asexual species.<sup>(30)</sup> Loss of incompatibility caused by a mutation in the pollen incompatibility gene, rather than the *SRK* gene, is also consistent with theory,<sup>(25)</sup> since this increases male fitness. However, the strength of selection may be an underestimate. The starting diversity assumed for *SCR* before the selective event was the current average value for *A. thaliana* genes,<sup>(1)</sup> but the ancestor should have had much higher *SCR* diversity than this (although diversity has not yet been quantified for the *A. lyrata* *SCR*, it is evidently high).<sup>(5)</sup> *A. thaliana*'s low *SCR* diversity is thus consistent with a selective substitution at this locus, but, if so, selection must have been stronger than estimated (and must thus occurred even more recently than estimated).

### High diversity at nearby loci

Another test for recent selection, rather than some historical event affecting the population, is that selection will affect other nearby sequences, but not loci elsewhere in the genome, whereas bottlenecks affect all loci. Theoretical models of selective sweeps show effects limited to small regions,<sup>(31,32)</sup> allowing genome scans to pinpoint loci where advantageous substitutions have occurred recently,<sup>(33,34)</sup> but strong selection eliminates variability fast, so that diversity is reduced over wide regions. This is seen at selected loci in domesticated species; for example, in an ~50 kb region near the maize *tb1* gene.<sup>(35)</sup>

In the *A. thaliana* *S*-locus region, however, high diversity is seen at nearby genes.<sup>(1)</sup> This contrasts with the low *SCR1* and is inconsistent with strong selection (as suggested above) having occurred. Two loci not involved in pollen recognition, but located in the *A. thaliana* *S*-locus region (see Fig. 1) have



very high diversity, similar to values found in *A. lyrata*<sup>(36)</sup> (and unpublished data of E. Kamau and D. Charlesworth). *SRK* diversity is also very high compared with other *A. thaliana* loci. Silent site nucleotide diversity is at least 13.8% (an under-

estimate, given that some alleles with deletions were excluded from the analysis). These results are at first sight consistent with very recent weak selection at *SCR1* leading to the evolution of selfing, only mildly affecting *SRK* and other

flanking loci.<sup>(1)</sup> *SRK* is at least 800 bp away from *SCR1*, and even this small distance is probably too far for a weak selective sweep at *SCR1* to affect its diversity (see Fig. 1). However, *SRK* diversity is reduced to about 38% of the *A. lyrata* value, and only three sequence types are seen at this locus.<sup>(1)</sup> Most alleles (17 of the 21 that yielded sequences, including the Col-0 strain whose genome was sequenced) fall into a single sequence type, haplogroup A, and three of the other four alleles have another similar sequence, C. Because *S*-alleles are maintained in populations for long evolutionary times, the *A. thaliana* sequences should probably still be present in extant self-incompatible *Arabidopsis* species. Indeed, the haplogroup B sequence (which was found only in the *A. thaliana* strain Cvi) is similar to a known *A. lyrata* allele, and A sequences are very similar to an *SRK* allele recently found in *A. lyrata* and *A. halleri* (unpublished results of V. Castric, X. Vekemans, J. Bechsgaard and M.H. Schierup), while the C haplotype, from three strains, is so far unknown in *A. lyrata* (but may yet be found).

If a selective sweep is not the reason for the loss of other *S*-alleles that would have been present in the ancestral, self-incompatible, species, what might explain *SRK*'s reduced diversity? After evolving inbreeding, genetic drift should lead to loss of variability at these loci, and their diversity should slowly approach the lower levels of other *A. thaliana* loci. According to a standard population genetics formula, a population of effective size  $N_e$  should, after  $t$  generations, have diversity equal to  $\exp(-t/2N_e)$  times its initial value; for a reduction in diversity to 40% as reported for *SRK*,<sup>(1)</sup> this time is around  $1.5N_e$ . Since  $N_e$  is estimated to be about 400,000,<sup>(1)</sup> this implies many generations. Thus, the high observed *SRK* diversity shows that there has been too little time for alleles to be lost since selfing evolved, but a very recent event is also unlikely.

Another surprise is that haplogroup A sequences are not entirely uniform,<sup>(1)</sup> there are five non-synonymous variants and one synonymous one, giving a nucleotide diversity value of about 0.1%, well above values typically found within functional haplotypes in self-incompatible species;<sup>(9,37)</sup> haplotype C sequences also vary slightly. If this variability within haplotypes accumulated since *SRK* became non-functional, it would point to quite a long evolutionary time since the *A. thaliana* *SRK* gene stopped being under selection.

## Conclusions

Overall, these conflicting observations mean that we cannot yet conclude definitively that *A. thaliana* changed breeding system as recently as a few hundred thousand years ago. Other models, including selection at a gene for a downstream step in incompatibility, cannot readily explain the observed highly restricted loss of *SCR1* diversity. However, the simple selective sweep model in which the spread of a non-functional *SCR1* allele through the ancestral *A. thaliana* population

caused the high frequency of haplotype A, should be tested further. It predicts, for example, that the *A. lyrata* and *A. halleri* haplotypes corresponding to *SRK* haplogroups A or B of *A. thaliana* should carry an *SCR* sequence with recognisable similarity to those of the corresponding *A. thaliana* *SCR1*s, but without the mutation(s) abolishing *SCR1*'s function. Other possibilities involving changes in the *S*-locus region, including a crossover or unequal recombination event in the *S*-locus region, might also be able to generate non-functional *S*-gene haplotypes. Such an event might even create non-functional *S*-gene sequences, given the different organisation of different *A. lyrata* *S*-allele haplotypes.<sup>(5)</sup>

## Acknowledgments

We thank Vincent Castric, Brian Charlesworth, Akira Kawabe and Mikkel Schierup for discussions and/or providing information.

## References

1. Shimizu KK, Cork JM, Caicedo AL, Mays CA, Moore RC et al. 2004. Darwinian selection on a selfing locus. *Science* 306:2081–2084.
2. Nasrallah ME, Liu P, Sherman-Broyles S, Boggs NA, Nasrallah JB. 2004. Natural variation in expression of self-incompatibility in *Arabidopsis thaliana*: Implications for the evolution of selfing. *Proceedings of the National Academy of Sciences of the USA*. 101:16070–16074.
3. Bateman AJ. 1955. Self-incompatibility systems in angiosperms. III. Cruciferae. *Heredity* 9:52–68.
4. Schierup MH. 1998. The effect of enzyme heterozygosity on growth in a strictly outcrossing species, the self-incompatible *Arabis petraea* (Brassicaceae). *Heredity* 128:21–31.
5. Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, et al. 2001. Self-incompatibility in the genus *Arabidopsis*: characterization of the *S* locus in the outcrossing *A. lyrata* and its autogamous relative, *A. thaliana*. *Plant Cell* 13:627–643.
6. Charlesworth D, Awadalla P, Mable BK, Schierup MH. 2000. Population-level studies of multiallelic self-incompatibility loci, with particular reference to Brassicaceae. *Annals of Botany* 85:227–239.
7. Schierup MH, Mable BK, Awadalla P, Charlesworth D. 2001. Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. *Genetics* 158:387–399.
8. Mable BK, Schierup MH, Charlesworth D. 2003. Estimating the number of *S*-alleles in a natural population of *Arabidopsis lyrata* (Brassicaceae) with sporophytic control of self-incompatibility. *Heredity* 90:422–431.
9. Charlesworth D, Bartolomé C, Schierup MH, Mable BK. 2003. Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. *Mol Biol Evol* 20:1741–1753.
10. Abbott RJ, Gomes MF. 1988. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* 62:411–418.
11. Berge G, Nordal I, Hestmark G. 1998. The effect of breeding systems and pollination vectors on the genetic variation of small plant populations within an agricultural landscape. *Oikos* 81:17–29.
12. Stenoien HK, Fenster CB, Tonteri A, Savolainen O. 2005. Genetic variability in natural populations of *Arabidopsis thaliana* in northern Europe. *Mol Ecol* 14:137–148.
13. Pichersky E, Gershenzon J. 2002. The formation and function of plant volatiles: perfumes for pollinator attraction and defense. *Curr Opin Plant Biol* 5:237–243.
14. Ornduff R. 1969. Reproductive biology in relation to systematics. *Taxon* 18:121–133.
15. Macnair MR, Cumbes QJ. 1990. The genetic architecture of interspecific variation in *Mimulus*. *Genetics* 122:211–222.

16. Kuittinen H, Haan AAd, Vogl C, Oikarinen S, Leppälä J, et al. 2004. Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Genetics* 168:575–1584.
17. Charlesworth B. 1992. Evolutionary rates in partially self-fertilizing species. *Amer Nat* 140:126–148.
18. Arumuganathan K, Earle ED. 1991. Nuclear DNA content of some important plant species. *Plant Mol Biol Reporter* 9:208–218.
19. Soltis DE, Soltis PS, Leitch IJ, Bennett MD. 2005. Evolution of DNA amounts across land plants (embryophyta). *Ann Bot (Lond)* 95:207–217.
20. Wright SI, Le QH, Schoen DJ, Bureau TE. 2001. Population dynamics of an Ac-like transposable element in self- and cross-fertilizing *Arabidopsis*. *Genetics* 158:1279–1288.
21. Wright SI, Lauga B, Charlesworth D. 2003. Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Molecular Ecology* 12:1247–1263.
22. Nasrallah ME, Liu P, Nasrallah JB. 2002. Generation of self-incompatible *Arabidopsis thaliana* by transfer of two S locus genes from *A. lyrata*. *Science* 297:247–249.
23. Nasrallah JB. 2002. Recognition and rejection of self in plant reproduction. *Science* 296:305–308.
24. Sato T, Nishio T, Kimura R, Kusaba M, Suzuki G et al. 2002. Coevolution of the S-locus genes SRK, SLG and SP11/SCR in *Brassica oleracea* and *B. rapa*. *Genetics* 162:931–940.
25. Uyenoyama MK, Zhang Y, Newbigin E. 2001. On the origin of self-incompatibility haplotypes: Transition through self-compatible intermediates. *Genetics* 157:1805–1817.
26. Coop G, Griffiths RC. 2004. Ancestral inference on gene trees under selection. *Theor Popul Biol* 66:219–232.
27. Schoen DJ, L'Heureux A-M, Marsolais J, Johnston MO. 1997. Evolutionary history of the mating system in *Amsinckia* (Boraginaceae). *Evolution* 51:1090–1099.
28. Bena G, Prosperi JM, Lejeune B, Olivieri I. 1998. Evolution of annual species of the genus *Medicago*: A molecular phylogenetic approach. *Molecular Phylogenetics and Evolution* 9:552–559.
29. Takebayashi N, Morrell PP. 2001. Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. *Am J Bot* 88:1143–1150.
30. Quattro JM, Avise JC, Vrijenhoek RC. 1991. Molecular evidence for multiple origins of hybridogenetic fish clones (Poeciliidae: Poeciliopsis). *Genetics* 127:391–398.
31. Kim Y, Stephan W. 2003. Selective sweeps in the presence of interference among partially linked loci. *Genetics* 164:389–398.
32. Charlesworth B. 1992. New genes sweep clean. *Nature* 356:475–476.
33. Harr B, Kauer M, Schlotterer C. 2002. Hitchhiking mapping: A population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 99:12949–12954.
34. Nair S, Williams JT, Brockman A, Paiphun L, Mayxay M, et al. 2003. A selective sweep driven by pyrimethamine treatment in Southeast Asian malaria parasites. *Molecular Biology and Evolution* 20:1526–1536.
35. Clark RM, Linton E, Messing J, Doebley JF. 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc Natl Acad Sci USA* 101:700–707.
36. Charlesworth D, Mable BK, Schierup MH, Bartolomé C, Awadalla P. 2003. Diversity and linkage of genes in the self-incompatibility gene family in *Arabidopsis lyrata*. *Genetics* 164:1519–1535.
37. Lawrence MJ, Lane MD, O'Donnell S, Franklin-Tong VE. 1993. The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. V. Cross-classification of the S-alleles from three natural populations. *Heredity* 71:581–590.
38. Charlesworth D, Charlesworth B, Strobeck C. 1977. Effects of selfing on selection for recombination. *Genetics* 68:213–226.