

# Genomic consequences of selection on self-incompatibility genes

## Mikkel Heide Schierup<sup>1</sup> and Xavier Vekemans<sup>2</sup>

Frequency-dependent selection at plant self-incompatibility systems is inherent and well understood theoretically. A self-incompatibility locus leads to a strong peak of diversity in the genome, to a unique distribution of diversity across the species and possibly to increased introgression between closely related species. We review recent empirical studies demonstrating these features and relate the empirical findings to theoretical predictions. We show how these features are being exploited in searches for other genes under multi-allelic balancing selection and for inference on recent breakdown of self-incompatibility.

### Addresses

<sup>1</sup> Bioinformatics Research Center and Department of Biology, University of Aarhus, Hoegh Guldbergs Gade 10, 8000 Aarhus C, Denmark

<sup>2</sup> Université des Sciences et Technologies de Lille 1, Laboratoire de génétique et évolution des populations végétales, CNRS UMR 8016, Villeneuve d'Ascq, France

Corresponding author: Schierup, Mikkel Heide ([mheide@birc.au.dk](mailto:mheide@birc.au.dk)) and Vekemans, Xavier ([xavier.vekemans@univ-lille1.fr](mailto:xavier.vekemans@univ-lille1.fr))

Current Opinion in Plant Biology 2008, 11:116–122

This review comes from a themed issue on  
Genome Studies and Molecular Genetics  
Edited by Juliette de Meaux and Maarten Koornneef

Available online 7th March 2008

1369-5266/\$ – see front matter  
Published by Elsevier Ltd.

DOI 10.1016/j.pbi.2008.01.003

### Introduction

Self-incompatibility systems are widespread genetic systems promoting outbreeding in hermaphrodite plants. They are based on several genetic and biochemical mechanisms for recognition and rejection of self-pollen that have arisen several times independently [1•]. Common to most systems is the presence of a single Mendelian locus (the S-locus) consisting of several genes, with at least two responsible for the male and female specificities, respectively. In gametophytic self-incompatibility (SI) systems, the gene expressing male specificity is expressed gametophytically in the pollen tube, whereas in sporophytic SI systems it is expressed sporophytically in the anther tissues. In spite of these differences, SI systems share similar population genetics and evolutionary properties marked by strikingly high variation at the S-locus, and dramatic effects on diversity and recombination on the S-locus genomic neighbourhood. These properties derive from a particular form of natural selection inherent

to these systems, strong negative frequency-dependent selection and its interaction with population structure and introgression processes.

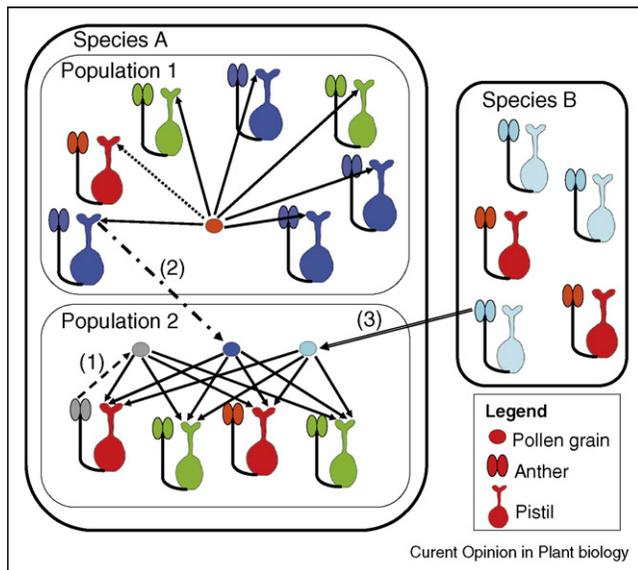
Here we review current knowledge of the forces maintaining genetic variation at the S-locus and how they interact, resulting in a genomic signature of elevated polymorphism. This is of importance in understanding the generation of new specificities at these systems and the effects of a breakdown of a functional SI system. We discuss how theory predicts that diversity might be maintained by introgression during speciation processes. Our examples are taken from the system studied in most detail, the sporophytic SI system of Brassicaceae, but the conclusions are expected to apply equally to any other SI system. Furthermore, there are presumably many less understood candidates for genes in plants under similar multi-allelic balancing selection regimes, for example the highly polymorphic resistance genes, which may be expected to behave in a similar way. We thus compare expectations from S-locus systems with genomic patterns observed in recent surveys of variation in *Arabidopsis thaliana* and rice.

### Forces maintaining variation at S-loci

In a seminal paper, Wright [2] noticed that SI not only leads to the rejection of self-pollen but also of outcrossed pollen expressing identical specificity. He also pointed out that pollen grains expressing a rare specificity would benefit from a higher access to compatible partners (Figure 1). This mechanism leads to a strong negative frequency-dependent selection on the S-locus. This selection promotes high number of different specificities (also termed S-alleles) at the S-locus, since it leads to similar population frequencies of all alleles and also ensures protection against loss of rare alleles through genetic drift. Negative frequency-dependent selection belongs to a broad category of selective processes with similar evolutionary properties referred to as balancing selection [3], that is the selective maintenance of several alleles at a locus.

Frequency-dependent selection is also expected to cause the coexistence of highly divergent allelic specificities at the S-locus within species [4,5]. This is due to the fact that within allelic genealogies, coalescence among allelic lineages only occurs when a given allele generates a new selected allele through mutational events [3]. In contrast to neutral gene genealogies, the lower the rate of origination of new allelic specificities, the higher would be the timescale of the S-locus allelic genealogy (see Figure 2). The expected origination rate of new S-alleles is very low

Figure 1

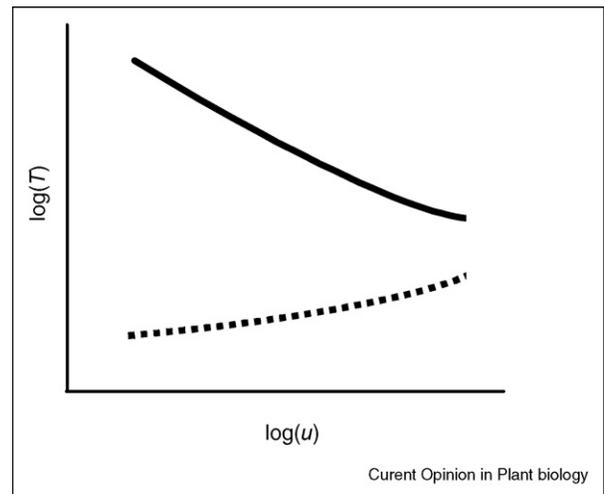


Evolutionary processes acting on the S-locus. The fundamental process is negative frequency-dependent selection: within population 1 of species A, a pollen grain expressing a rare red specificity will benefit from the ability to fertilize almost all individuals (solid arrows) with the exception of the rare individuals expressing its own specificity (dotted arrow). Several mechanisms cause the introduction of a new specificity within a population (here population 2 of species 1) that would be compatible with all individuals and would increase in frequency. (1) *Mutation*: an individual carrying a red specificity produces through mutation in the anther a pollen grain with a new grey specificity (dashed arrow). (2) *Migration*: a pollen grain with a blue specificity from population 1 migrated into population 2 where it was previously absent (interrupted arrow). (3) *Introgression*: through hybridization, a pollen grain from species B is introduced in species A (double arrow) where it will benefit from the advantage of the rare once recombined away from the genetic background of species B.

because (1) the evolution of a new specificity necessitates matching changes in the amino acid sequences of both male and female specificity genes within the same haplotype [6]; (2) intermediate stages in the evolution of a new specificity may be selectively disadvantageous [7]; (3) a new specificity may suffer from the expression of highly deleterious recessive mutations linked to the S-locus when forming heterozygous genotypes with its direct progenitor allele [8], for experimental evidence, see references [9<sup>\*</sup>,10]. In concordance with these predictions, S-alleles are extremely divergent from each other, with typical amino acid divergence between random pairs of alleles in the range of 20–70% [11–13]. This may easily correspond to 10–50 million years of divergence [11,14], which is strikingly longer than divergence times of polymorphisms at neutral loci that are typically restricted to a few million years.

Another unique feature of balancing selection acting at the S-locus is the way diversity is distributed among local populations. Indeed, selection interacts with migration to

Figure 2



Effect of the rate of mutation to new alleles ( $u$ ) on coalescence times of alleles ( $T$ ) in allelic genealogies of a neutral locus (dotted line) and a locus under strong balancing selection (solid line). For a neutral locus, average coalescence times in a gene genealogy are independent of mutation rate ( $T = 4Nu$ ), but in an allelic genealogy, average  $T$  decreases with  $u$  because of an increase in the proportion of genealogies with zero mutations occurring (and thus the population is fixed for a single allele leading to  $T = 0$ ). For a locus under strong balancing selection, where a single allele can never go to fixation, coalescence events in both gene and allelic genealogies are associated to mutation events so that average  $T$  increases when  $u$  is decreasing. Adapted from Vekemans and Slatkin [5].

maintain high allelic diversity within local populations [15]. This is because migration may introduce locally rare alleles, which then will be favoured by the negative frequency-dependent selection acting on the S-locus [16,17]. This effect has been shown to be substantial even under rather low levels of migration. Another consequence is that genetic differentiation among populations at the S-locus is expected to be much lower than for neutral genes. Such phenomenon can also be viewed as considering that selection is increasing substantially the effective migration at the S-locus [17]. Empirical studies confirm the limited population structure at the S-locus ([18,19], Schierup *et al.*, in press) as well as for a plant resistance gene suspected to be under a similar selection regime [20<sup>\*</sup>].

A similar mechanism may promote exchanges of S-alleles among closely related species, as long as hybridization between the two species does occur. If an allele is absent from one species, a hybrid individual carrying this allele will have a strong mating advantage that may offset its decreased hybrid fitness, and this process may lead to selection for introgression of the S-allele. Such adaptive introgression process may play an important role in rescuing diversity at the S-locus after speciation events, when these are accompanied by genetic bottlenecks, and thus contribute to the maintenance of highly diverged allelic

lineages through strikingly long evolutionary times (Figure 1). Pairs of S-alleles shared trans-specifically between *Arabidopsis lyrata* and *Arabidopsis halleri* are indeed more similar to each other in nucleotide sequence than expected given the estimated speciation time, suggesting that introgression has aided transfer of alleles between these two species that can be artificially crossed in the greenhouse [21,22,23\*].

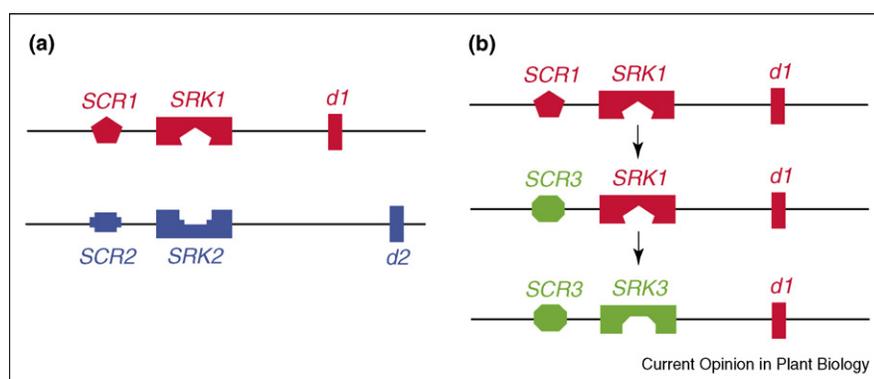
One may even take this idea one step further and speculate that incomplete reproductive barriers between closely related species may facilitate the evolution of new specificities. As stated above, for a new specificity to evolve, changes need to occur in both the male and female component genes (Figure 3), probably without a self-compatible intermediate allele since this will be selectively disadvantageous owing to inbreeding depression [6,7]. One possible model [24] suggests that the structural flexibility of the pollen ligand protein in the sporophytic system of Brassicaceae can lead to segregation of slightly different pollen specificities that could eventually differentiate into fully distinct specificities, once corresponding polymorphism within the female component gene originates. However, very little polymorphism is expected within an allelic type because each allele is maintained over time at low frequency [5], unless there is very strong population subdivision [17]. In this context, two species with incomplete reproductive barriers can indeed be considered as a common gene pool for S-alleles with very strong population subdivision that can promote the origination of new allelic specificities. Furthermore, there is the possibility that trans-specific pairs of alleles may slowly drift to different specificities in the two species and later be transferred between species through introgression.

Introgression may possibly also rescue variation at other genes under balancing selection. Clark *et al.* [25] and Bakker *et al.* [26] reported very high diversity associated with disease resistance genes in *Arabidopsis thaliana* (see also reference [27] for a review of immune system genes under putative balancing selection). The genomes of *Capsella rubella* and *A. lyrata*, two close relatives of *A. thaliana*, are being sequenced and will probably reveal the extent to which gene introgression influences patterns of diversity of genes under balancing selection.

### The genomic signature of balancing selection

Balancing selection is not only expected to affect patterns of diversity at the self-incompatibility locus itself, but also at positions that are sufficiently closely linked to the S-locus that they indirectly experience balancing selection [28,29]. Several empirical studies have shown evidence for genetic hitch-hiking effects in the S-locus genomic region that resulted in high nucleotide diversity [30–33] and low population structure (see reference [34\*] in *A. lyrata* and reference [65] in *A. halleri*) or that generated trans-specific polymorphisms [35]. These studies showed that the physical region influenced by the S-locus is quite large ( $\approx 100$  kb) [33,34\*,65], see Figure 4. Earlier theoretical and simulation work showed that the region of elevated variation would be proportional to a region of size  $4Nr = n$ , where  $n$  is the number of specificities [29,36–38]. Exemplified by realistic values of parameters from *A. lyrata*, if  $N = 100\,000$ ,  $r = 5 \times 10^{-8}$  [39] and  $n = 50$ , the expected region of increased polymorphism would be  $\sim 2.5$  kb, much smaller than observed in SI systems of Brassicaceae and Solanaceae [34\*,40,41]. The elongation of the genealogy near the balanced polymorphism leaves ample time for recombination events to occur [37], lead-

Figure 3



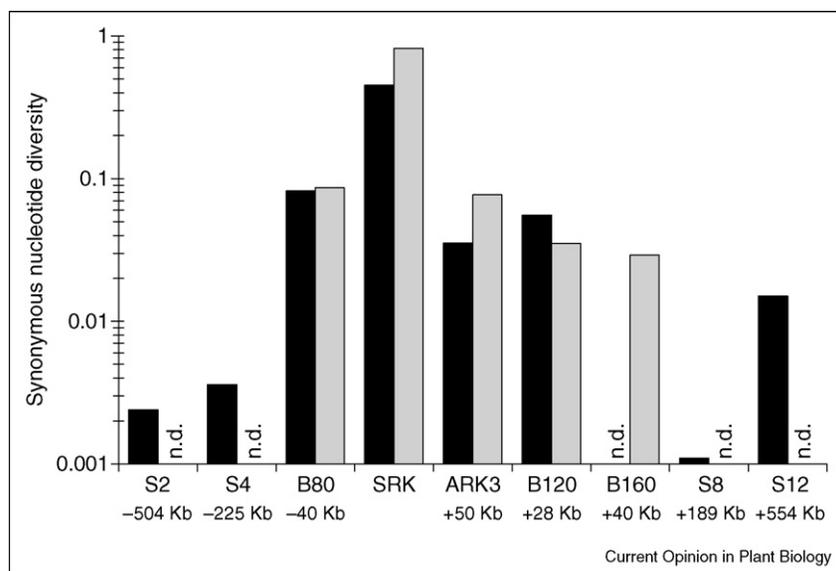
The bipartite structure of the S-locus and origination of new S-alleles. The S-locus in Brassicaceae contains two closely linked genes determining the pollen (SCR) and pistil (SRK) phenotypes. (A) The S-locus region for two distinct S-alleles (red and blue) with cartoons showing that SRK and SCR co-evolved to produce proteins forming a specific receptor–ligand interaction. Also shown is the accumulation of S-allele-specific recessive deleterious mutations at linked loci (*d1* and *d2*) as a consequence of high heterozygosity and low recombination rate in the S-locus region. (B) Causes of the low rate of origination of new S-alleles: (1) evolution of S-allele 3 (green) from S-allele 1 (red) involves the occurrence of matching mutations in SRK and SCR in the same haplotype; (2) intermediate haplotypes (e.g. SCR3/SRK1) are self-compatible as the two proteins do not form a receptor–ligand interaction, and will be counter-selected by expression of genomic inbreeding depression; (3) the parental (red) and descendant (green) S-alleles share a common recessive deleterious mutation (*d1*) that will be expressed in S1/S3 heterozygotes.

ing to an expected rapid decay of diversity away from the specificity determining genes.

The obvious explanation for the discrepancy between observed and expected regions of increased diversity is a suppression of recombination in the S-locus region. Evolution of decreased recombination rate between the male and female component genes is expected from population genetics theory because a recombination event between the two genes would produce self-compatible genotypes that would be counter-selected owing to extant strong inbreeding depression in natural populations. Uyenoyama [40] pointed out that selection for reduced recombination might lead to chromosomal restructuring of the S-locus region, as a consequence of accumulation of transposable elements, which in turn reduces recombination rate until it is practically zero around the S-locus. In Brassica, phylogenetic trees of S-alleles for the male (SCR/SP11) and female component genes are concordant [42], suggesting the absence of recombination between the two genes for many million years. Large scale sequencing of the S-locus genomic region for several S-haplotypes in the sporophytic SI system of Arabidopsis and Brassica, as well as in the gametophytic SI system of Petunia, Anthirrhinum and Prunus, showed that highly rearranged genomic structure and the accumulation of retrotransposons (in Brassica [43]; *A. lyrata* [44]; *A. thaliana* [45<sup>\*</sup>]; Petunia [46]; Prunus [47] and Antirrhinum [48]) constitute indeed general properties of these systems. Direct evidence from the crossing data that recombination events are more rare in the S-locus region than the genomic aver-

age has been reported in *Brassica oleracea* [49], *A. lyrata* [39], and *Petunia inflata* [41]. Using an alternative approach, Kamau *et al.* [34<sup>\*</sup>] recently applied an estimator of recombination rate similar to Takahata and Satta's [38] model to show that recombination in the S-locus region in *A. lyrata* is suppressed by a factor of 10–100 compared with that of the neighbouring genomic region. Note that this is an estimate of the recombination rate that has affected the data. It is still an open question whether recombination events are prevented to occur, or whether they do occur but are selected against because they are detrimental in a highly rearranged region of the genome, or simply because they lead to self-compatible haplotypes. V Castric and X Vekemans (unpublished results) surveyed variation among allele copies corresponding to the same S-allele specificity and found evidence that recombinations do occur, suggesting that the high sequence divergence between different specificities in itself may make recombination more difficult. Also, the observation of a recent recombination event within the non-functional S-locus region in the self-compatible *A. thaliana* is compatible with the hypothesis that low recombination in this region is actively maintained in self-incompatible species by purifying selection against recombinants [50]. A recent paper by Takuno *et al.* [33] shows that flanking genes as well as the kinase domain of SRK have genealogies that are not concordant to the one observed at SRK S-domain and SCR. However, none of these results can rule out a certain degree of gene conversion among S-alleles and some evidence of this has been found in Brassica [51].

Figure 4



Synonymous nucleotide diversity at the female component gene of the S-locus (*SRK*) and at flanking genes in *A. lyrata* (black bars) [34<sup>\*</sup>,64] and *A. halleri* (grey bars) [65] illustrating the peak of diversity in the S-locus genomic region. Distances from *SRK* were computed on the basis of Kusaba *et al.* [44] and the *A. thaliana* genome annotation. n.d.: Not determined. Note the logarithmic y-axis.

Table 1

Summary of evolutionary properties of loci subject to multi-allelic balancing selection as compared with that of neutral loci

	Multi-allelic balancing selection	Neutral
Allelic diversity	High	Low
Frequency distribution	Equal frequency of alleles	Unequal frequency of alleles
Nucleotide divergence among alleles	High	Low
Ratio of non-synonymous over synonymous substitutions	High	Low
Distribution of diversity	Within populations	Between populations
Introgression rate	High	Low
Intragenic recombination	Suppressed (in self-incompatibility systems)	Normal

Genomic signature of balancing selection has been reported in other genetic systems [40]. Charlesworth [52\*\*] pointed out, however, that large differences in genomic effects are expected as different systems vary widely in terms of the timescale of maintenance of selected polymorphisms and of the local rate of recombination. For instance, in MHC [53,38] and for the sex-determining gene in honeybees [54], genomic effects are limited owing to high recombination rates, whereas strong signatures of balancing selection are found in regions with suppressed recombination such as sex chromosomes [52\*\*] or the mitochondrial genome of gynodioecious species with cytoplasmic male sterility [55]. Certain resistance genes in *A. thaliana* also show an extended region of increased polymorphism [56].

### Role of balancing selection on genome-wide patterns of nucleotide diversity

As discussed above, selection working on self-incompatibility systems is well understood, and they exert strong imprints on the amount and distribution of genetic variation in a large genomic neighbourhood of the actual genes causing the incompatibility reaction. These properties are summarized in Table 1. Thus, if we want to look for other genes under long-term balancing selection we may search genome-wide diversity data for similar effects. Genome-wide diversity surveys of plants are still scarce and are not designed to survey population differentiation as well, but two model species have recently been studied in detail, that is *Arabidopsis thaliana* [25,57] and *Oryza sativa* [58,59]. *A. thaliana* shows several regions of very high diversity, some of which coincide with resistance genes, as previously discussed. In *O. sativa*, several regions of highly elevated polymorphism [58] may also indicate balancing selection in the complex domestication history of this species. However, it is also possible that loci under multi-allelic balancing selection might be very rare. For example, in humans where detailed polymorphism data exist, Bubb *et al.* [48] failed to find convincing evidence for balancing selection outside of the MHC region using these approaches.

Genomic effects associated with balancing selection are also characterized by their long persistence times, even after extinction of the selection pressure. This is because the decay in diversity due to genetic drift after demographic perturbations [60] or changes in selective regime is rather low. This is exemplified by the observation of high diversity at genes surrounding the non-functional S-locus in *A. thaliana* [61], which was interpreted as a signature of historical hitch-hiking effects to the S-locus before the breakdown of self-incompatibility [62]. Indeed, the extant polymorphism around the remnants of the S-locus is still noticeable in the genome-wide survey of variation in *A. thaliana* [45\*]. Even though the loss of self-incompatibility in *A. thaliana* was complex and probably occurred by multiple events at the S-locus [45\*,50,63], the remnant diversity provides clues to when the breakdown occurred [21,45\*]. The loss of self-incompatibility is a common evolutionary trend and the remnants of diversity in other self-compatible species with closely related self-incompatible species is expected to provide important clues to how fast morphological and genomic changes evolve in response to a change in mating system.

### Acknowledgements

MHS acknowledges support from the EU (ERA-PG). XV acknowledges support from the French National Research Agency (grant ANR-06-BLAN-0128-01).

### References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Takayama S, Isogai A: **Self-incompatibility in plants**. *Annu Rev Plant Biol* 2005, **56**:467-489.  
• An excellent review of molecular mechanisms of self-incompatibility systems.
2. Wright S: **The distribution of self-sterility alleles in populations**. *Genetics* 1939, **24**:538-552.
3. Takahata N: **A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism**. *Proc Natl Acad Sci U S A* 1990, **87**:2419-2423.
4. Schierup MH, Vekemans X, Christiansen FB: **Allelic genealogies in sporophytic self-incompatibility systems in plants**. *Genetics* 1998, **150**:1187-1198.

5. Vekemans X, Slatkin M: **Gene and allelic genealogies at a gametophytic self-incompatibility locus.** *Genetics* 1994, **137**:1157-1165.
6. Charlesworth D: **How can two-gene models of self-incompatibility generate new specificities?** *Plant Cell* 2000, **12**:309-310.
7. Uyenoyama MK, Zhang Y, Newbigin E: **On the origin of self-incompatibility haplotypes: transition through self-compatible intermediates.** *Genetics* 2001, **157**:1805-1817.
8. Uyenoyama MK: **Genealogy-dependent variation in viability among self-incompatibility genotypes.** *Theor Popul Biol* 2003, **63**:281-293.
9. Bechsgaard J, Bataillon T, Schierup MH: **Uneven segregation of sporophytic self-incompatibility alleles in *Arabidopsis lyrata*.** *J Evol Biol* 2004, **17**:554-561.  
Evidence for additional selection acting on linked variation to an SI system.
10. Stone JL: **Sheltered load associated with S-alleles in *Solanum carolinense*.** *Heredity* 2004, **92**:335-342.
11. Ioerger TR, Clark AG, Kao TH: **Polymorphism at the self-incompatibility locus in Solanaceae predates speciation.** *Proc Natl Acad Sci U S A* 1990, **87**:9732-9735.
12. Kusaba M, Nishio T, Satta Y, Hinata K, Ockendon D: **Striking sequence similarity in inter- and intra-specific comparisons of class I SLG alleles from *Brassica oleracea* and *Brassica campestris*: implications for the evolution and recognition mechanism.** *Proc Natl Acad Sci U S A* 1997, **94**:7673-7678.
13. Schierup MH, Mable BK, Awadalla P, Charlesworth D: **Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*.** *Genetics* 2001, **158**:387-399.
14. Uyenoyama MK: **A generalized least-squares estimate for the origin of sporophytic self-incompatibility.** *Genetics* 1995, **139**:975-992.
15. Schierup MH: **The number of self-incompatibility alleles in a finite, subdivided population.** *Genetics* 1998, **149**:1153-1162.
16. Muirhead CA: **Consequences of population structure on genes under balancing selection.** *Evol Int J Org Evol* 2001, **55**:1532-1541.
17. Schierup MH, Vekemans X, Charlesworth D: **The effect of subdivision on variation at multi-allelic loci under balancing selection.** *Genet Res* 2000, **76**:51-62.
18. Glemin S, Gaude T, Guillemin ML, Lourmas M, Olivieri I, Mignot A: **Balancing selection in the wild: testing population genetics theory of self-incompatibility in the rare species *Brassica insularis*.** *Genetics* 2005, **171**:279-289.
19. Kato S, Iwata H, Tsumura Y, Mukai Y: **Distribution of S-alleles in island populations of flowering cherry, *Prunus lannesiana* var. *speciosa*.** *Genes Genet Syst* 2007, **82**:65-75.
20. Ding J, Zhang W, Jing Z, Chen JQ, Tian D: **Unique pattern of R-gene variation within populations in *Arabidopsis*.** *Mol Genet Genomics* 2007, **277**:619-629.  
This study shows how resistance genes share many polymorphism features with self-incompatibility genes.
21. Bechsgaard JS, Castric V, Vekemans X, Charlesworth D, Schierup MH: **The transition to self-compatibility in *Arabidopsis thaliana* and evolution within S-haplotypes over 10 Myr.** *Mol Biol Evol* 2006, **23**:1741-1750.
22. Castric V, Vekemans X: **Evolution under strong balancing selection: how many codons determine specificity at the female self-incompatibility gene SRK in Brassicaceae?** *BMC Evol Biol* 2007, **7**:132.
23. Castric V, Bechsgaard JS, Schierup MH, Vekemans X: **Adaptive introgression mediated by balancing selection at the self-incompatibility locus between *Arabidopsis lyrata* and *A. halleri*.** 2007, submitted for publication.  
Evidence that self-incompatibility alleles introgress more easily than neutral genes.
24. Chookajorn T, Kachroo A, Ripoll DR, Clark AG, Nasrallah JB: **Specificity determinants and diversification of the Brassica self-incompatibility pollen ligand.** *Proc Natl Acad Sci U S A* 2004, **101**:911-917.
25. Clark RM, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA *et al.*: **Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*.** *Science* 2007, **317**:338-342.
26. Bakker EG, Toomajian C, Kreitman M, Bergelson J: **A genome-wide survey of R gene polymorphisms in *Arabidopsis*.** *Plant Cell* 2006, **18**:1803-1818.
27. Tiffin P, Moeller DA: **Molecular evolution of plant immune system genes.** *Trends Genet* 2006, **22**:662-670.
28. Charlesworth B, Nordborg M, Charlesworth D: **The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations.** *Genet Res* 1997, **70**:155-174.
29. Schierup MH, Charlesworth D, Vekemans X: **The effect of hitchhiking on genes linked to a balanced polymorphism in a subdivided population.** *Genet Res* 2000, **76**:63-73.
30. Kamau E, Charlesworth D: **Balancing selection and low recombination affect diversity near the self-incompatibility loci of the plant *Arabidopsis lyrata*.** *Curr Biol* 2005, **15**:1773-1778.
31. Takebayashi N, Brewer PB, Newbigin E, Uyenoyama MK: **Patterns of variation within self-incompatibility loci.** *Mol Biol Evol* 2003, **20**:1778-1794.
32. Entani T, Iwano M, Shiba H, Che FS, Isogai A, Takayama S: **Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity.** *Genes Cells* 2003, **8**:203-213.
33. Takuno S, Fujimoto R, Sugimura T, Sato K, Okamoto S, Zhang SL, Nishio T: **Effects of recombination on the hitchhiking diversity in Brassica self-incompatibility locus complex.** *Genetics* 2007, **177**:949-958.
34. Kamau E, Charlesworth B, Charlesworth D: **Linkage disequilibrium and recombination rate estimates in the self-incompatibility region of *Arabidopsis lyrata*.** *Genetics* 2007, **176**:2357-2369.  
Clear demonstration of reduced recombination rate around a self-incompatibility locus.
35. Charlesworth D, Kamau E, Hagenblad J, Tang CL: **Trans-specificity at loci near the self-incompatibility loci in *Arabidopsis*.** *Genetics* 2006, **172**:2699-2704.
36. Kaplan NL, Darden T, Hudson RR: **The coalescent process in models with selection.** *Genetics* 1988, **120**:819-829.
37. Schierup MH, Mikkelsen AM, Hein J: **Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes.** *Genetics* 2001, **159**:1833-1844.
38. Takahata N, Satta Y: **Footprints of intragenic recombination at HLA loci.** *Immunogenetics* 1998, **47**:430-441.
39. Kawabe A, Hansson B, Forrest A, Hagenblad J, Charlesworth D: **Comparative gene mapping in *Arabidopsis lyrata* chromosomes 6 and 7 and *A. thaliana* chromosome IV: evolutionary history, rearrangements and local recombination rates.** *Genet Res* 2006, **88**:45-56.
40. Uyenoyama MK: **Evolution under tight linkage to mating type.** *New Phytol* 2005, **165**:63-70.
41. Wang Y, Wang X, McCubbin AG, Kao TH: **Genetic mapping and molecular characterization of the self-incompatibility (S) locus in *Petunia inflata*.** *Plant Mol Biol* 2003, **53**:565-580.
42. Sato K, Nishio T, Kimura R, Kusaba M, Suzuki T, Hatakeyama K, Ockendon DJ, Satta Y: **Coevolution of the S-locus genes SRK, SLG and SP11/SCR in *Brassica oleracea* and *B. rapa*.** *Genetics* 2002, **162**:931-940.
43. Fujimoto R, Okazaki K, Fukai E, Kusaba M, Nishio T: **Comparison of the genome structure of the self-incompatibility (S) locus in interspecific pairs of S haplotypes.** *Genetics* 2006, **173**:1157-1167.

44. Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, Nasrallah ME: **Self-incompatibility in the genus *Arabidopsis*: characterization of the S locus in the outcrossing *A. lyrata* and its autogamous relative *A. thaliana***. *Plant Cell* 2001, **13**:627-643.
45. Tang CL, Toomajian C, Sherman-Broyles S, Plagnol V, Guo YL, Hu TT, Clark RM, Nasrallah JB, Weigel D, Nordborg M: **The evolution of selfing in *Arabidopsis thaliana***. *Science* 2007, **317**:1070-1072.
- Detailed elucidation of the transition to selfing in *A. thaliana* from the diversity in the remnant S-locus region.
46. Wang Y, Tsukamoto T, Yi KW, Wang X, Huang S, McCubbin AG, Kao TH: **Chromosome walking in the *Petunia inflata* self-incompatibility (S-) locus and gene identification in an 881-kb contig containing S2-RNase**. *Plant Mol Biol* 2004, **54**:727-742.
47. Ushijima K, Sassa H, Dandekar AM, Gradziel TM, Tao R, Hirano H: **Structural and transcriptional analysis of the self-incompatibility locus of almond: identification of a pollen-expressed F-box gene with haplotype-specific polymorphism**. *Plant Cell* 2003, **15**:771-781.
48. Bubba KL, Bovee D, Buckley D, Haugen E, Kibukawa M, Paddock M, Palmieri A, Subramanian S, Zhou Y, Kaul R *et al.*: **Scan of human genome reveals no new loci under ancient balancing selection**. *Genetics* 2006, **173**:2165-2177.
49. Casselman AL, Vrebalov J, Conner JA, Singhal A, Giovannoni J, Nasrallah ME, Nasrallah JB: **Determining the physical limits of the Brassica S locus by recombinational analysis**. *Plant Cell* 2000, **12**:23-33.
50. Sherman-Broyles S, Boggs N, Farkas A, Liu P, Vrebalov J, Nasrallah ME, Nasrallah JB: **S locus genes and the evolution of self-fertility in *Arabidopsis thaliana***. *Plant Cell* 2007, **19**:94-106.
51. Fujimoto R, Sugimura T, Nishio T: **Gene conversion from SLG to SRK resulting in self-compatibility in *Brassica rapa***. *FEBS Lett* 2006, **580**:425-430.
52. Charlesworth D: **Balancing selection and its effects on sequences in nearby genome regions**. *Plos Genetics* 2006, **2**:379-384.
- Easily accessible review on expected molecular signatures caused by multi-allelic balancing selection.
53. Carrington M: **Recombination within the human MHC**. *Immunol Rev* 1999, **167**:245-256.
54. Hasselmann M, Beye M: **Pronounced differences of recombination activity at the sex determination locus of the honeybee, a locus under strong balancing selection**. *Genetics* 2006, **174**:1469-1480.
55. Stadler T, Delph LF: **Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant**. *Proc Natl Acad Sci U S A* 2002, **99**:11730-11735.
56. Cork JM, Purugganan MD: **High-diversity genes in the *Arabidopsis* genome**. *Genetics* 2005, **170**:1897-1911.
57. Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M: **Recombination and linkage disequilibrium in *Arabidopsis thaliana***. *Nat Genet* 2007, **39**:1151-1155.
58. Tang T, Lu J, Huang JZ, He JH, McCouch SR, Shen Y, Kai Z, Purugganan MD, Shi SH, Wu CI: **Genomic variation in rice: genesis of highly polymorphic linkage blocks during domestication**. *Plos Genetics* 2006, **2**:1824-1833.
59. Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, Polato NR, Olsen KM, Nielsen R, McCouch SR *et al.*: **Genome-wide patterns of nucleotide polymorphism in domesticated rice**. *PLoS Genet* 2007, **3**:1745-1756.
60. Nei M: **Molecular population genetics and evolution**. *Front Biol* 1975, **40**:1-288 North-Holland Research Monographs, Amsterdam.
61. Shimizu KK, Cork JM, Caicedo AL, Mays CA, Moore RC, Olsen KM, Ruzsa S, Coop G, Bustamante CD, Awadalla P *et al.*: **Darwinian selection on a selfing locus**. *Science* 2004, **306**:2081-2084.
62. Charlesworth D, Vekemans X: **How and when did *Arabidopsis thaliana* become highly self-fertilising**. *Bioessays* 2005, **27**:472-476.
63. Shimizu KK, Shimizu-Inatsugi R, Tsuchimatsu T, Purugganan MD: **Independent origins of self-compatibility in *Arabidopsis thaliana***. *Mol Ecol* 2008, **17**:704-714.
64. Charlesworth D, Bartolome C, Schierup MH, Mable BK: **Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata***. *Mol Biol Evol* 2003, **20**:1741-1753.
65. Ruggiero MV, Jacquemin B, Castric V, Vekemans X: **Hitch-hiking to a locus under balancing selection: high sequence diversity and low population subdivision at the S-locus genomic region in *Arabidopsis halleri***. *Genetic Res* 2008, **89**, in press.