Research review

Plant self-incompatibility systems: a molecular evolutionary perspective

Author for correspondence:
Deborah Charlesworth
Tel: +44 (0) 131 650 5751
Fax: +44 (0) 131 650 6564
Email: deborah.charlesworth@ed.ac.uk

Deborah Charlesworth[1], Xavier Vekemans[2], Vincent Castric[2] and
Sylvain Glémin[3]

[1]Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Ashworth Laboratory, King's Buildings, West Mains Road, Edinburgh EH9 3JT, UK; [2]Laboratoire de Génétique et Evolution des Populations Végétales, UMR CNRS 8016 – FR CNRS 1818, Bâtiment SN2, Université de Lille 1, F-59655 Villeneuve d'Ascq Cedex, France; [3]UMR5171, 'Génome, Populations, Interactions, Adaptation', Université Montpellier II, CC63 Place Eugène Bataillon, 34095 Montpellier cedex 05, France

Summary

Incompatibility recognition systems preventing self-fertilization have evolved several times in independent lineages of Angiosperm plants, and three main model systems are well characterized at the molecular level [the gametophytic self-incompatibility (SI) systems of Solanaceae, Rosaceae and *Anthirrhinum*, the very different system of poppy, and the system in Brassicaceae with sporophytic control of pollen SI reactions]. In two of these systems, the genes encoding both components of pollen-pistil recognition are now known, showing clearly that these two proteins are distinct, that is, SI is a lock-and-key mechanism. Here, we review recent findings in the three well-studied systems in the light of these results and analyse their implications for understanding polymorphism and coevolution of the two SI genes, in the context of a tightly linked genome region.

Introduction

The evolution of functionally complex systems is an important question. Self-incompatibility (SI) pollen-pistil recognition systems in flowering plants are particularly fascinating, but understanding their evolution is more difficult than for other recognition systems because the component genes are highly polymorphic within species. It is easy to imagine the evolution of recognition systems involved in developmental signalling: changes in one component may not immediately abolish the old function; an allele encoding a protein with an additional useful developmental function, even if imperfect, can thus replace the species' pre-existing gene, and the new function can later improve by fixation of alleles with further

changes in either component (Löhr *et al.*, 2001). In the evolution of S-loci, however, gaining a new specificity requires losing the old one (to preserve the unique specificities of alleles). Moreover, changes in specificity of just the pollen or the pistil should lead to full or partial self-compatibility; thus, both recognition components must change, and both changes must occur in the same haplotype.

We are, however, gaining a better understanding of SI evolution as a result of the combination of molecular genetic and molecular evolutionary work. SI systems have evolved at least three times independently, providing opportunities for identifying common features in the different systems and thus for testing evolutionary hypotheses. Studies of SI should lead to a general understanding of how variation in these genes

is maintained and how selection, including coadaptation between the pollen and pistil components, affects the evolution of genome regions around the incompatibility loci. In turn, an understanding of such genomic effects in well-characterized model systems is necessary to rigorously evaluate evidence on where, in genomes, selection is maintaining variability.

Knowing the components of polymorphic recognition systems does not necessarily indicate what maintains the variants. There is still uncertainty for vertebrate major histocompatibility complex (MHC) systems, and the maintenance of variability in incompatibility systems in fungi (e.g. Casselton, 1998) and some sessile animals (e.g. Grosberg, 1988) is also not yet understood. Plant SI systems are excellent model systems for using to understand such variability because we know that the chemical recognition of pollen by pistils prevents self-fertilization, creating an advantage for any allele that is rare in a population. This frequency-dependent selection (a form of balancing selection) leads to the long-term maintenance of many alleles with different incompatibility types (Vekemans & Slatkin, 1994), and alleles thus become widely dispersed throughout a species' different populations (Muirhead, 2001).

General features of different plant SI systems

Although plant SI systems all prevent self-fertilization through recognition and rejection of pollen by pistils expressing 'cognate' allelic specificity (described later), very different cellular mechanisms are involved in the three systems where the genes have been characterized. The genetics and mechanisms of these systems have recently been reviewed in detail (Nasrallah, 2002; Kao & Tsukamoto, 2004). Molecular data are so far available only for systems in which a single polymorphic '*S*-locus region' controls the different incompatibility specificities. Pistil SI proteins are often soluble and can be studied in pistil extracts. Thus, the first molecular information, from *Brassica* and *Nicotiana*, came through the electrophoresis of major pistil proteins, combined with cDNA analysis. Other self-incompatible species in these two plant families were subsequently shown to have homologous *S* loci. *Petunia* species (and other Solanaceae) and Rosaceae have stylar *S*-RNases, and the important genus for molecular research, *Arabidopsis*, has a system similar to that of *Brassica*. *A. lyrata* is self-incompatible, with a system like that in *Brassica*, while its close relative, *A. thaliana*, is autogamous and highly self-fertilizing in nature (Abbott & Gomes, 1988; Nasrallah *et al.*, 2004) with nonfunctional *S*-locus homologues (Kusaba *et al.*, 2001).

It has long been thought that the S-locus region contains separate pistil and pollen protein genes (i.e. that recognition in SI is a 'lock-and-key' system, with components as different as in other ligand–receptor systems), and this has now been confirmed. Evidence for a distinct pollen determinant was, until recently, indirect, based mainly on mutagenesis studies, in which pollen and pistil incompatibility can be affected

independently (Lewis, 1947; Kao & Tsukamoto, 2004). Potential 'pollen-*S*' genes have now been identified in laborious map-based searches of large genomic regions around the pistil *S* genes of Brassicaceae, Solanaceae, *Antirrhinum* and Rosaceae, and tested in transgenic experiments (described later). For these species (but not yet in *Papaver*, with a system quite different from either the Brassicaceae or the *S*-RNase system), the lock-and-key model is thus at last confirmed, and the alternative (recognition through expression of the same proteins in pistil and pollen) is excluded.

S-gene polymorphism

We have explained in the previous section that *S* genes should have higher polymorphism than other genes. To test this, diversity must be quantified in samples of alleles from natural populations. Genomic gel blotting (restriction fragment length polymorphism analysis) can provide some data, but shows merely that variants exist, without adequate quantification. Allele numbers depend on recombination rates and so are an unsatisfactory measure of diversity (they can be enormous for sequences with several polymorphic sites that recombine, such as MHC genes). Diversity should thus be quantified as 'nucleotide diversity', using measures such as the mean fraction of sites that differ between two randomly sampled alleles at a locus, π. Amino acid and nonsynonymous site diversity (π_n) will be particularly high in codons directly affected by balancing selection, and regions with high π_n or $\pi_n > \pi_s$ (synonymous site diversity) can often help to identify these (Table 1; Richman *et al.*, 1996). As a basis for comparison, π_s in maize has been shown to range up to 3.6% across a sample of 21 loci (Tenaillon *et al.*, 2002).

It is now understood that variability will also often be high at unselected sites close to sites where frequency-dependent selection maintains many alleles for long time-periods (Schierup *et al.*, 2000; Takebayashi *et al.*, 2003). If functionally different *S* alleles rarely recombine, their sequences will evolve like alleles in isolated populations with low 'effective population sizes', and hence diversity among different instances of the same functional allele will be low. Between different alleles, however, synonymous and even nonsynonymous differences can accumulate in the absence of recombination (Vekemans & Slatkin, 1994; Charlesworth *et al.*, 2003). Given estimated plant recombination values of approx. 1 Mb per centiMorgan, the predicted diversity peaks should span only a few hundred nucleotides, which would nevertheless greatly hinder precise identification of the sites within *S* proteins that are involved in recognition (Charlesworth *et al.*, 2003), a chief reason for obtaining sequences. In the MHC systems, recombination occurs and diversity is indeed largely highest at sites near known peptide-binding amino acids (Table 1).

In SI systems, maintaining coadapted combinations of pollen and pistil alleles should result in the evolution of low-recombination frequencies between the *S* loci. The

Table 1 Diversity per nucleotide of different regions of genes of several recognition systems

SI system and gene	Gene region	Diversity (%)			References
		Nonsynonymous (π_n value[a])	Synonymous (π_s value)	All nucleotides	
Brassica SRK	S-domain	12.0	20.8	–	Hinata *et al.* (1995)
2 alleles	Kinase domain	4.8	21.7	–	
28 alleles	HV	29.9	21.9	–	Sato *et al.* (2002)
	Non-HV	7.1	17.3	–	
SRK	Positively selected sites	–	–	38.3	Takebayashi *et al.* (2003)
	Conservatively evolving sites	–	–	7.4	
SCR(SP11)	Positively selected sites	–	–	> 100	
	Conservatively evolving sites	–	–	8.2	
6 alleles	All	50.5	32.1	–	Sato *et al.* (2002)
SLG (34 alleles)	All	7.1	17.3	–	Sato *et al.* (2002)
Arabidopsis lyrata SRK	S-domain	62.8	28.3	–	Charlesworth *et al.* (2003)
15 alleles	Kinase domain	58.1	17.4	–	
S-RNase	Positively selected sites			> 100	Takebayashi *et al.* (2003)
	Conservatively evolving sites			39.6	
MHC	Antigen-binding regions	14.1	4.7	–	Hughes & Nei (1988)
	Nonantigen-binding regions in exons 2 and 3	2.4	5.1	–	

[a]Values shown are π or similar estimates. The different methods used to correct for saturation are not specified here, as all values are approximate when diversity is high. (Note that values > 100% are possible after such correction.)
The table shows the high diversity in regions likely to be under balancing selection, and also in other regions of the *S*-gene sequences [comparing either the *S*-domain and kinase domains of *SRK*, or positively selected vs conservatively evolving sites, or sites within hypervariable (HV) regions vs other regions]. Comparable data for a major histocompatibility complex (MHC) gene (*HLA-DRB*1) are shown to illustrate the lower synonymous site diversity in regions not under balancing selection, in contrast to *S* loci.

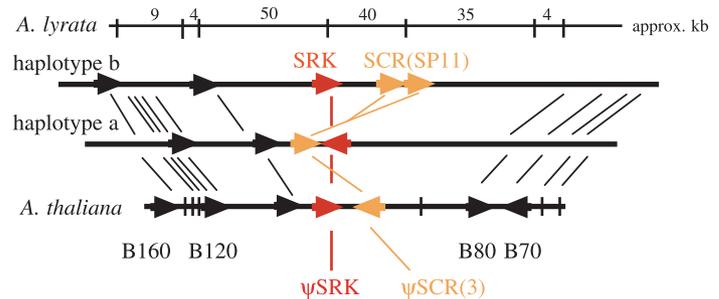**(a)** SRK–SCR(SP11) system in Arabidopsis

Fig. 1 Different genome arrangements near plant incompatibility loci. (a) The *SRK–SCR* region and flanking colinear genes, showing rearrangements between *Arabidopsis lyrata* haplotypes Sa and Sb (Kusaba *et al.*, 2001). (b) The *S-RNase–SFB* region of *Prunus mume*, showing large size differences between two haplotypes (Entani *et al.*, 2003).

(b) S-RNase–F-box system in *Prunus mume*
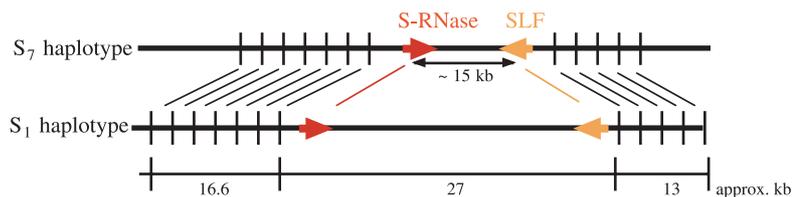
S-locus genome region is physically too small (Fig. 1) to test this by genetic mapping with feasible family sizes, but population genetic approaches may be able to test for low recombination through detecting haplotypes of only certain pistil and pollen *S*-allele combinations (linkage disequilibrium). If recombination is rare, diversity may be elevated over large genomic tracts, possibly including other genes and sequences located within the *S*-locus region.

High diversity is a necessary condition for accepting a gene as an *S* locus, and candidate genes can sometimes be eliminated because of low diversity (Casselman *et al.*, 2000; Takebayashi *et al.*, 2003). If the *S* loci lie within a nonrecombining region, however, high diversity may hinder even identifying the functional *S* genes. Moreover, diversity data from SI plants are scarce for non-*S* loci, so one often cannot test whether candidate *S* loci have significantly elevated diversity. Transgenic experiments are thus also essential to show that candidate genes change the incompatibility reaction appropriately.

Brassicaceae

In the Brassicaceae, the pistil protein is a receptor kinase, SRK. Initial work in *Brassica* identified an S-domain protein, the highly polymorphic *SLG* (reviewed in Sato *et al.*, 2002), but the evidence from polymorphism was misleading: *SLG* has no major role in recognition, but seems merely to enhance the SI response – its polymorphism probably results from close linkage to the true pistil-recognition gene, *SRK* (Nasrallah, 2002). SRK is anchored to the stigma membrane, presenting its extracellular *S* domain to ligand molecules on the pollen surface (Nasrallah, 2002). Both domains are extraordinarily polymorphic (Table 1), although the *S* domain is presumably the sole region subject to balancing selection. This suggests that the *S* locus may lie within a region of unusually low recombination frequency. As shown in Fig. 1, physical mapping shows that allelic *S*-locus regions are organized into haplotypes with the genes in different orders and orientations (Boyes *et al.*, 1997). Moreover, high synonymous and also nonsynonymous site diversity in the *A. lyrata SRK* kinase domain (Table 1) are difficult to explain unless recombination with the *S* domain is infrequent. These findings are all consistent with the *S*-locus region containing at least two coadapted genes, the most plausible evolutionary reason for a low recombination rate. The alternative, that the region is located in a genome region, such as a centromeric region, that rarely recombines, seems not to apply in *A. lyrata* (Kusaba *et al.*, 2001) whose gene locations within chromosome arms are probably similar to those in *A. thaliana* (Kuittinen *et al.* 2004).

A breakthrough in establishing the existence of separate pollen *S* genes was the identification of the *Brassica* and *Arabidopsis* pollen SI determinant. The candidate gene for the ligand in Brassicaceae is a small, cysteine-rich pollen-coat protein, SCR (also called SP11; Schopfer *et al.*, 1999; Takayama *et al.*, 2000). Its sequence diversity in *Brassica* is high (Watanabe *et al.*, 2000; Shiba *et al.*, 2002); however, a full picture of variability is not yet available because alleles in dominant haplotypes differ so greatly that not all genetically detectable alleles can be sequenced. In *A. lyrata*, with only two alleles compared to date, many differences were found in this short protein sequence (Kusaba *et al.*, 2001).

Direct evidence from transgenic experiments shows that *SCR*(*SP11*) is indeed the pollen *S* gene. Transfer of *SRK*

and *SCR* alleles of a particular *A. lyrata* haplotype causes SI in *A. thaliana* buds, though only transiently in many strains (Nasrallah *et al.*, 2004), and transferring Raphanus *SCR* alleles into *B. campestris* (*rapa*) caused the recipients' pollen to be rejected by *Brassica* plants with a particular incompatibility type (Sato *et al.*, 2004), showing that *SCR* determines the pollen incompatibility type and also that certain alleles in the two species have the same incompatibility type (despite several sequence differences).

Even with this detailed knowledge from Brassicaceae, unexpected properties of *S* genes continue to emerge, notably expression differences of SCR(SP11) alleles. In heterozygotes with alleles of differing dominance, only the more dominant allele is expressed (Kusaba *et al.*, 2002; Shiba *et al.*, 2002). This mono-allelic expression is like that in other chemoreceptor systems, including mammalian odorant receptors (Goldmit & Bergman, 2004), and ensures that each pollen grain has a single SI type. Unlike other receptor systems, expression choice is deterministic: dominant alleles are expressed in both the anther tapetum and, gametophytically, in pollen, whereas recessive alleles are not expressed in pollen. Expression is presumably controlled by different upstream sequences, which suggests that the haplotype structure of *S*-locus regions may embrace those regions also. New alleles may thus be constrained to have the same dominance as their progenitors. This has not yet been incorporated into evolutionary models, but it may explain the clustering of *Brassica* allele sequences according to their dominance.

Solanaceae, Anthirrhinum and Rosaceae

Unlike the SI of Brassicaceae, where inhibition occurs on the stigma surface, in these species incompatible pollen tubes are inhibited in the stylar transmitting tract. Consistent with this, the pistil S-RNase protein is expressed in style tissue. It is taken up by growing pollen tubes, degrading RNAs within incompatible pollen tubes and causing subsequent arrest of growth in the style (Kao & Tsukamoto, 2004). Like *SRK*, *S*-RNases are highly polymorphic (Table 1), although most sequence data come from the reverse transcription–polymerase chain reaction amplification of pistil cDNA. Synonymous and nonsynonymous diversity are both high (e.g. Richman *et al.*, 1996). Only rarely are segregation tests carried out to verify that sequences are allelic and that different incompatibility classes in families correspond with different sequences. Natural population studies generally simply assume that different sequences represent different alleles (presumably treating cases with few differences as sequencing errors).

The pollen determinant of SI has now been identified in species of all three families with S-RNase systems (Solanaceae, Rosaceae and *Antirrhinum*), and the findings suggest a mechanism for these systems, and may explain some of their odd features. Genes encoding F-box proteins have been candidates for the gene controlling pollen recognition since the discovery
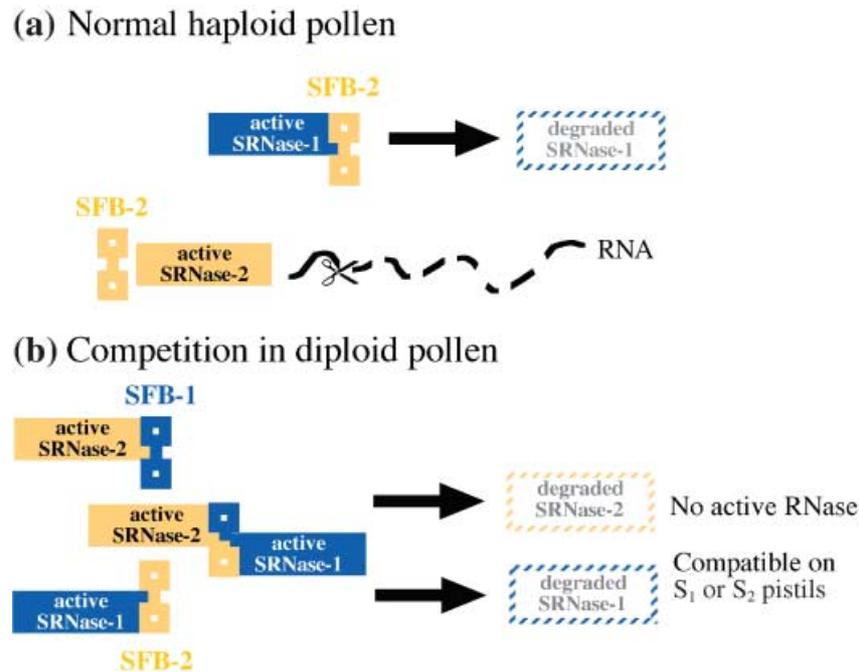
(a) Normal haploid pollen

(b) Competition in diploid pollen

Fig. 2 Diagram showing a possible mechanism of self-incompatibility in *S*-RNase systems (a), and the competitive interaction phenomenon (b). In this model, SI is caused by failure of each pollen *S*-linked F-box (SFB) allele product to destroy its cognate pistil S-RNase. (a) A pollen tube from a haploid pollen grain carrying the S_2 *SFB* allele is shown, in the self-pollination of a plant with genotype S_1S_2. Pistil *S*-RNases encoded by both alleles enter the tube, and that from the S_1 haplotype is degraded while the S_2 *S*-RNase is not. RNA in the pollen tube is then digested by the S_2 *S*-RNase (i.e. the pollen is incompatible). (b) In pollen containing both S_1 and S_2 *SFB* alleles, both *S*-RNases are degraded, leading to compatibility. Other interactions between alleles in diploid pollen (Lewis, 1947, 1960) can also be explained on this model. Weakening of SI suggests that the pollen SFB proteins must interact, probably forming dimers (or a higher-order complex) that fail to eliminate all RNase, even though each homodimer still leads to destruction of some of its noncognate RNase, as with 'competing' *SFB* alleles; as heterodimers should represent half of the SFB protein present, sufficient active RNAse might be present to give weak incompatibility. If heterodimers fail to cause destruction of either of their cognate RNases (even though homodimers lead to degradation), some RNase activity of both allelic types might remain (coexpressed alleles). Finally, if heterodimers affect only one RNase (say, the S_2 product), some of the other – S_1 – RNase might remain active and digest RNA; if even a small amount of active RNase causes incompatibility, the pollen S_1 allele would behave as dominant.

of *S*-linked F-box (*SFB*) genes in *A. hispanicum* (Lai *et al.* 2002) and *Prunus mume* (Rosaceae, Entani *et al.*, 2003). However, formal proof was needed that an *SFB* gene determines pollen specificity, as this is a very large gene family (694 estimated genes in *A. thaliana*; Gagne *et al.*, 2002), and several such genes have been found in the *S*-locus region of some species (Entani *et al.*, 2003; Kao & Tsukamoto, 2004). Transformation experiments now confirm F-box genes as the pollen *S* of both *P. inflata* (Sijacic *et al.*, 2004) and *A. hispanicum* (Qiao *et al.* 2004).

The *Petunia* study made elegant use of the 'competition interaction' phenomenon, which creates difficulties for transgenic experiments with pollen *S* genes, but provides a characteristic specific for the male SI determinant. In many known gametophytic systems, pollen grains heterozygous for S alleles may be compatible with plants carrying both alleles (Lewis, 1947). Thus, self-compatible plants in species with S-RNase systems commonly arise by tetraploidy or duplication of the *S* locus, making some of the pollen effectively heterozygous for the pollen *S* gene. This does not occur in the Brassicaceae:

tetraploids maintain normal incompatibility reactions (Mable *et al.*, 2004). The pollen SFB protein belongs to a protein family involved in ubiquitin-mediated protein degradation. F-box proteins often play crucial roles, delivering appropriate targets to the ubiquitin–protein ligase complex (Gagne *et al.*, 2002). The 'inhibitor model' for SI proposes that interactions of pollen-recognition proteins with nonself *S*-RNases do not allow RNase activity, whereas activity does occur after interaction with the cognate *S*-RNase (Entani *et al.*, 2003; Kao & Tsukamoto, 2004). Unexpectedly, the new results offer a simpler model and can explain the competition phenomenon. Competition suggests active destruction of RNase activity by any pollen *S*-protein not recognized as the cognate one, for example compatibility resulting from degradation of all nonself S-RNase proteins (Fig. 2).

Surprisingly, expression of the *P. inflata PiSLF* gene peaks in immature bicellular pollen in the anthers, not during pollen tube growth. Diversity data and duplications in self-compatible strains (Tsukamoto *et al.*, 2005), however, support the conclusion that the true pollen *S* has been found, although

further data are needed. An approximate quantification for *PiSLF*, using a few alleles, suggests extremely high amino acid diversity (Sijacic *et al.*, 2004), consistent with the extremely high *S*-RNase π_s and π_n values compared with other plant loci (Table 1). However, in some *Prunus* and *Petunia* species with *S*-RNase systems, the *S* loci may lie close to centromeric regions (with low recombination). Until linkage has been quantified, and diversity estimated for nearby loci, evidence from diversity of candidate genes is incomplete.

Papaveraceae

In Papaveraceae, only the pistil *S* protein has been characterized to date. In *Papaver rhoeas*, it is the pistil S protein that is a small extracellular signal molecule (see Thomas & Franklin-Tong, 2004), presumably interacting with a receptor on pollen tube surfaces. Despite the pollen determinant still being hypothetical, the cellular mechanism of SI is known in considerable detail. Self-stigma S proteins elicit a rapid increase in Ca^{2+} within pollen tubes growing *in vitro*, suggesting the involvement of programmed cell death (PCD) processes in the SI response. Indeed, inhibiting caspase-3-protease, a key PCD enzyme, abolishes endonuclease activity and prevents DNA fragmentation and pollen-tube growth inhibition that occur in normal incompatibility (Thomas & Franklin-Tong, 2004).

Downstream biochemical pathways involved in SI, and evolution of self-compatibility

Downstream cellular pathways leading to self-pollen rejection are also becoming better known for the *Brassica* and *S*-RNase systems (Vanoosthuyse *et al.*, 2003; O'Brien *et al.*, 2004). This should illuminate how secondary self-compatibility arises, a common evolutionary change in plants (Goodwillie, 1999; Nasrallah *et al.*, 2004). SI can be lost by mutations in either the *S* locus or the downstream pathway. Many components of the *Brassica* SI process are now known, including the ARC1 protein that interacts with SRK (Stone *et al.*, 2003), and the *MLPK* gene, recently characterized in a self-compatible *B. rapa* cultivar, which is unlinked to the *S* locus but encodes a SRK-like membrane-anchored cytoplasmic protein kinase (Murase *et al.*, 2004).

Incompatibility breakdown in cases of natural loss of SI is less well understood. In *A. thaliana* there is evidence for several potential mutations, and the timescale of SI loss is uncertain. In the Columbia strain, the *SRK* and *SCR* orthologues are pseudogenes (Kusaba *et al.*, 2001), suggesting loss of SI through mutations in an *S* locus. Transgenic plants expressing *SRK* and *SCR* from *A. lyrata* indeed show some incompatibility, and strain C24 can be made strongly incompatible in such experiments, suggesting that this strain's downstream pathway is functional (Nasrallah *et al.*, 2004). However, other strains expressing the same transgenes remain

compatible, indicating variability for mutations in downstream pathway genes. SRK sequences in *A. thaliana* are also surprisingly polymorphic (Shimizu *et al.*, 2004; Nasrallah *et al.*, 2004). If compatibility evolved by selection of a mutant *S* gene, a 'selective sweep' should have occurred, and this region's diversity would be lower (or at least no higher) than that of other genes, unless recombination is extremely frequent so that the selective sweep could affect just the mutated *S* gene, which seems improbable. Low diversity was indeed found for one *SCR(SP11)* pseudogene (Shimizu *et al.*, 2004). The high diversity of *SRK* and other flanking loci (Nasrallah *et al.*, 2004; Shimizu *et al.*, 2004) suggests that there has been no rapid fixation in this species of any particular *S* haplotype.

How do new specificities evolve?

The origin of new specificities is a major outstanding puzzle. Both well-understood SI systems are two-component (lock-and-key) systems. The origin of new haplotypes thus requires at least two mutations, yet a change in the specificity of one component will lead to self-compatibility. One possibility is thus that such a mutation causing self-compatibility spreads in the population, but another suitable mutation in the mutant haplotype restores incompatibility before all *S* alleles are lost from the population (Uyenoyama *et al.*, 2001). Targeted mutagenesis and domain-swapping experiments with pistil-recognition genes show that chimaeric 'dual-specificity' proteins can be formed which reject two different pollen types (Kao & Tsukamoto, 2004), suggesting another way that new specificities might evolve. Experiments in *Brassica* show that the SCR-binding affinity for SRK can be altered without losing the ability of the resulting complex to elicit incompatibility responses *in vivo* (Chookajorn *et al.*, 2003). SCR and SRK proteins of each specificity might thus form 'clouds' of sequences that may subsequently split under disruptive selection, optimizing recognition.

If several slightly different *SRK* sequences have the same specificity, a selective advantage accrues to an allele encoding a pollen protein variant that escapes rejection by some of these, providing it with more compatible mating opportunities than other variants of this specificity (Uyenoyama *et al.*, 2001). However, unless the advantage of this subspecificity diminishes as it spreads among its allele class, it would merely replace the initial allele, not add a new one (Uyenoyama & Newbigin, 2000). Occasionally, however, the pollen gene of another variant in the same allele class may acquire a second change that leads to rejection exclusively by a different pistil variant, replacing one initial incompatibility haplotype by two new functionally different ones with similar sequences (Fig. 3).

Is the necessary polymorphism present among sequences of pistil and pollen genes with the same specificities? Such sequence differences are expected to be rare because of the small effective population sizes of *S* alleles. The few data on diversity within *S*-allele types in *P. rhoeas* (O'Donnell *et al.*, 1993),
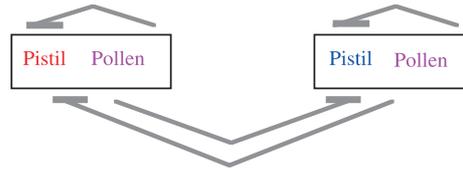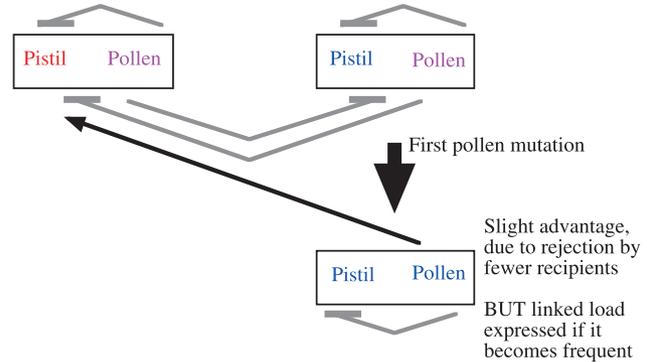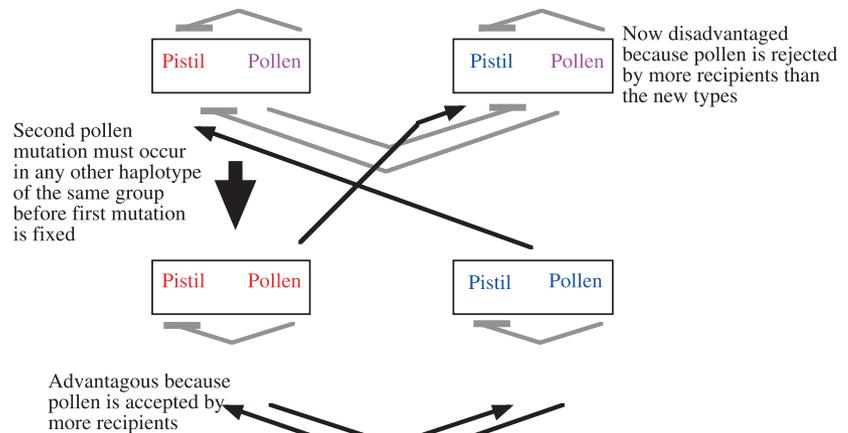
Fig. 3 Hypothesis for the evolutionary origin of new *S* haplotypes without a self-compatible intermediate state (Uyenoyama & Newbigin, 2000; Chookajorn *et al*., 2003). Compatible pollinations are shown as black lines ending in arrowheads, while incompatible pollinations are shown as grey lines ending with bars. For simplicity, of the many specificities present in the population, only the one giving rise to new alleles is shown (part 1). Part 1 of the figure shows the starting state for the evolution of new specificities – multiple pistil sequences that all function as this specificity (multiple pollen protein sequences are also possible, but if the first mutation changing the specificity affects the pistil protein, it cannot invade; see Uyenoyama & Newbigin, 2000). Part 2 shows the first pollen *S* mutation, and part 3 shows the complete new *S* haplotype. The second mutation must occur before fixation of the first (if the first mutation is fixed before the second appears, there is haplotype replacement without addition of a new specificity). Because the pistil compatibility types do not change, but continue to reject both initial pollen types, for clarity we have omitted the incompatible pollinations between the pollen proteins in the initial haplotypes and pistil proteins encoded by unchanged genes carried in new haplotypes.

Brassica (Kusaba *et al.*, 2000) and *A. lyrata* (Charlesworth *et al.*, 2003) suggest occasional differences, including some amino acid differences, between sequences of the same *S* alleles; no data are yet available for *S*-RNase systems.

A question for future study is how SI systems evolved in the first place. The evolution of these systems, with two components involved in recognition, seems difficult unless a previously existing cellular recognition system changed function to become a self-pollen recognition system. New knowledge about the detailed mechanisms of SI should illuminate the origins of SI systems. The three cases reviewed here each appear to have co-opted different cellular systems. All *S*-RNase systems may have a common origin (Steinbachs & Holsinger, 2002), while the *Brassica* and poppy systems are different (though ubiquitination may also be involved in the *Brassica* signalling process, Stone *et al.*, 2003). With plant genome sequences, relationships between different SI systems may become evident, and suggestions of similarities to functions such as pathogen-recognition systems (Thomas & Franklin-Tong, 2004) will become testable.

Many other self-incompatible plants are known, but molecular details are so far lacking. Another important task will be to learn about these incompatibility systems, including the important self-incompatible monocotyledon (see Hackauf & Wehling, 2005) and legume species, as well as the classic SI species, *Oenothera organensis*. The approaches and understanding

developed in the three currently well-studied systems should be helpful for studying the others. Other systems may reveal common features, giving mechanistic and evolutionary understanding, or new systems, exposing new plant cellular-recognition processes.

Acknowledgements

References

Abbott RJ, Gomes MF. 1988. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. *Heredity* **62**: 411–418.

Boyes DC, Nasrallah ME, Vrebalov J, Nasrallah JB. 1997. The self-incompatibility (S) haplotypes of *Brassica* contain highly divergent and rearranged sequences of ancient origin. *Plant Cell* **9**: 237–247.

Casselman AL, Nasrallah ME, Nasrallah JB. 2000. Using S-locus deletions to evaluate self-incompatibility candidate genes: an example for a novel anther-expressed gene. *Sexual Plant Reproduction* **12**: 227–231.

Casselton LA. 1998. Molecular genetics of mating recognition in Basidiomycete fungi. *Microbiology and Molecular Biology Reviews* **62**: 55–70.

Charlesworth D, Bartolomé C, Schierup MH, Mable BK. 2003. Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. *Molecular Biology and Evolution* **20**: 1741–1753.

Chookajorn T, Kachroo A, Ripoll DR, Clark AG, Nasrallah JB. 2003. Specificity determinants and diversification of the *Brassica* self-incompatibility pollen ligand. *Proceedings of the National Academy of Sciences, USA* **101**: 911–917.

Entani T, Iwano M, Shiba H, Che F-S, Isogai A, Takayama S. 2003. Comparative analysis of the self-incompatibility (S-) locus region of *Prunus mume*: identification of a pollen-expressed F-box gene with allelic diversity. *Genes to Cells* **8**: 203–213.

Gagne JM, Downes BO, Shiu SHAM, Durski & Vierstra RD. 2002. The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. *Proceedings of the National Academy of Sciences, USA* **99**: 11524.

Goldmit M, Bergman Y. 2004. Monoallelic gene expression: a repertoire of recurrent themes. *Immunological Reviews* **200**: 197–214.

Goodwillie C. 1999. Multiple origins of self-compatibility in *Linanthus* section Leptosiphon (Polemoniaceae): phylogenetic evidence from internal-transcribed-spacer sequence data. *Evolution* **53**: 1387–1395.

Grosberg RK. 1988. The evolution of allorecognition specificity in clonal invertebrates. *Quarterly Reviews of Biology* **63**: 377–412.

Hackauf B, Wehling P. 2005. Approaching the self-incompatibility locus *Z* in rye (*Secale cereale* L.) via comparative genetics. *Theoretical and Applied Genetics* **110**: 832–845.

Hinata K, Watanabe M, Yamakawa S, Satta Y, Isogai A. 1995. Evolutionary aspects of the *S*-related genes of the *Brassica* self-incompatibility system: synonymous and nonsynonymous base substitutions. *Genetics* **140**: 1099–1104.

Hughes A, Nei M. 1988. Pattern of nucleotide substitution at MHC class I loci reveals overdominant selection. *Nature* **335**: 167–170.

Kao TH, Tsukamoto T. 2004. The molecular and genetic bases of S-RNase-based self-incompatibility. *Plant Cell* **16**: S72–S83.

Kuittinen H, Haan AA, d Vogl C, Oikarinen S, Leppälä J, Koch M, Mitchell-Olds T, Langley C, Savolainen O. 2004. Comparing the linkage maps of the close relatives *Arabidopsis lyrata* and *Arabidopsis thaliana*. *Genetics* **168**: 1575–1584.

Kusaba M, Dwyer K, Hendershot J, Vrebalov J, Nasrallah JB, Nasrallah ME. 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.

Kusaba M, Matsushita M, Okazaki K, Satta Y, Nishio T. 2000. Sequence and structural diversity of the S-locus genes from different lines with the same self-recognition specificities in *Brassica oleracea*. *Genetics* **154**: 413–420.

Kusaba M, Tung CW, Nasrallah ME, Nasrallah JB. 2002. Monoallelic expression and dominance interactions in anthers of self-incompatible *Arabidopsis lyrata*. *Plant Physiology* **128**: 17–20.

Lai Z, Ma M, Han B, Liang L, Zhang Y, Hong G, Xue Y. 2002. An F-box gene linked to the self-incompatibility (S) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. *Plant Molecular Biology* **50**: 29–42.

Lewis D. 1947. Competition and dominance of incompatibility alleles in diploid pollen. *Heredity* **1**: 85–108.

Lewis D. 1960. Genetic control of specificity and activity of the S antigen in plants. *Proceedings of the Royal Society of London B* **151**: 468–477.

Löhr U, Yussa M, Pick L. 2001. Drosophila *fushi tarazu*: a gene on the border of homeotic function. *Current Biology* **11**: 1403–1412.

Mable BK, Beland J, Berardo CD. 2004. Inheritance and dominance of self-incompatibility alleles in polyploid *Arabidopsis lyrata*. *Heredity* **93**: 476–486.

Muirhead CA. 2001. Consequences of population structure on genes under balancing selection. *Evolution* **55**: 1532–1541.

Murase K, Shiba H, Iwano M, Che F-S, Watanabe M, Isogai A, Takayama S. 2004. A membrane-anchored protein kinase involved in *Brassica* self-incompatibility signaling. *Science* **303**: 1516–1519.

Nasrallah JB. 2002. Recognition and rejection of self in plant reproduction. *Science* **296**: 305–308.

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, USA* **101**: 16070–16074.

O'Brien M, Major G, Chantha S-C, Matton DP. 2004. Isolation of S-RNase binding proteins from *Solanum chacoense*: identification of an SBP1 (RING finger protein) orthologue. *Sexual Plant Reproduction* **17**: 81–87.

O'Donnell S, Lane MD, Lawrence MJ. 1993. The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. VI. Estimation of the overlap between the allelic complements of a pair of populations. *Heredity* **71**: 591–595.

Qiao H, Wang F, Zhao L, Zhou J, Lai Z, Zhang Y, Robbins TP, Xue Y. 2004. The F-box protein AhSLF-S2 controls the pollen function of S-RNase-based self-incompatibility. *Plant Cell* **16**: 2307–2322.

Richman AD, Uyenoyama MK, Kohn JR. 1996. Allelic diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. *Science* **273**: 1212–1216.

Sato T, Nishio T, Kimura R, Kusaba M, Suzuki G, Hatakeyama K, Ockendon D, Satta Y. 2002. Coevolution of the S-locus genes SRK, SLG and SP11/SCR in *Brassica oleracea* and *B. rapa*. *Genetics* **162**: 931–940.

Sato Y, Okamoto S, Nishio T. 2004. Diversification and alteration of recognition specificity of the pollen ligand SP11/SCR in self-incompatibility of *Brassica* and *Raphanus*. *Plant Cell* **16**: 3230–3241.

Schierup MH, Vekemans X, Charlesworth D. 2000. The effect of hitch-hiking on genes linked to a balanced polymorphism in a subdivided population. *Genetical Research (Cambridge)* **76**: 63–73.

Schopfer CR, Nasrallah ME, Nasrallah JB. 1999. The male determinant of self-incompatibility in *Brassica*. *Science* **286**: 1697–1700.

Shiba H, Iwano M, Entani T, Ishimoto K, Che F-S, Satta Y, Ito A, Takada Y, Watanabe M, Isogai A, Takayama S. 2002. The dominance of alleles controlling self-incompatibility in *Brassica* pollen is regulated at the RNA level. *Plant Cell* **14**: 491–504.

Shimizu KK, Cork JM, Caicedo AL, Mays CA, Moore RC, Olsen KM, Ruzsa S, Coop G, Bustamante CD, Awadalla P, Purugganan MD. 2004. Darwinian selection on a selfing locus. *Science* **306**: 2081–2084.

Sijacic P, Wang X, Skirpan AL, Wang Y, Dowd PE, McCubbin AGS, Huang & Kao TH. 2004. Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature* **429**: 302–305.

Steinbachs JE, Holsinger KE. 2002. S-RNase-mediated gametophytic self-incompatibility is ancestral in Eudicots. *Molecular Biology and Evolution* **19**: 825–829.

Stone SL, Anderson EM, Mullen RT, Goring DR. 2003. ARC1 is an E3 ubiquitin ligase and promotes the ubiquitination of proteins during the rejection of self-incompatible *Brassica* pollen. *Plant Cell* **15**: 885–898.

Takayama S, Shiba H, Iwano M, Shimosato H, Che F-S, Kai N, Suzuki G, Hinata K, Isogai A. 2000. The pollen determinant of self-incompatibility in *Brassica campestris*. *Proceedings of the National Academy of Sciences, USA* **97**: 1920–1925.

Takebayashi N, Brewer PB, Newbigin E, Uyenoyama MK. 2003. Patterns of variation within self-incompatibility loci. *Molecular Biology and Evolution* **20**: 1778–1794.

Tenaillon MI, Sawkins MC, Anderson LK, Stack SM, Doebley JF, Gaut BS. 2002. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Genetics* **162**: 1401–1413.

Thomas SG, Franklin-Tong VE. 2004. Self-incompatibility triggers programmed cell death in Papaver pollen. *Nature* **429**: 305–309.

Tsukamoto T, Ando T, Watanabe H, Marchesi E, Kao T-h. 2005. Duplication of the *S*-locus F-box gene is associated with breakdown of pollen function in an *S*-haplotype identified in a natural population of self-incompatible *Petunia axillaris*. *Plant Molecular Biology* **57**: 141–153.

Uyenoyama MK, Newbigin E. 2000. Evolutionary dynamics of dual-specificity self-incompatibility alleles. *Plant Cell* **12**: 310–312.

Uyenoyama MK, Zhang Y, Newbigin E. 2001. On the origin of self-incompatibility haplotypes: transition through self-compatible intermediates. *Genetics* **157**: 1805–1817.

Vanoosthuyse V, Tichtinsky G, Dumas C, Gaude T, Cock JM. 2003. Interaction of calmodulin, a sorting nexin and kinase-associated protein phosphatase with the *Brassica oleracea S* locus receptor kinase. *Plant Physiology* **133**: 919–929.

Vekemans X, Slatkin M. 1994. Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics* **137**: 1157–1165.

Watanabe M, Ito A, Takada Y, Ninomiya C, Kakizaki T, Takahata Y, Hatakeyama K, Hinata K, Suzuki G, Takasaki T, Satta Y, Shiba H, Takayama S, Isogai A. 2000. Highly divergent sequences of the pollen self-incompatibility (*S*) gene in class-I haplotypes of *Brassica campestris* (syn. *rapa*) L. *FEBS Letters* **473**: 139–144.