Molecular Ecology (2004) **13**, 2873–2889 doi: 10.1111/j.1365-294X.2004.02267.x

INVITED REVIEW

Plant self-incompatibility in natural populations: a critical assessment of recent theoretical and empirical advances

VINCENT CASTRIC and XAVIER VEKEMANS

Laboratoire de génétique et évolution des populations végétales, UMR CNRS 8016, Cité Scientifique, Bâtiment SN2, 59655 Villeneuve d'Ascq Cedex, France

Abstract

Self-incompatibility systems in plants are genetic systems that prevent self-fertilization in hermaphrodites through recognition and rejection of pollen expressing the same allelic specificity as that expressed in the pistils. The evolutionary properties of these self-recognition systems have been revealed through a fascinating interplay between empirical advances and theoretical developments. In 1939, Wright suggested that the main evolutionary force driving the genetic and molecular properties of these systems was strong negative frequency-dependent selection acting on pollination success. The empirical observation of high allelic diversity at the self-incompatibility locus in several species, followed by the discovery of very high molecular divergence among alleles in all plant families where the locus has been identified, supported Wright's initial theoretical predictions as well as many of its later developments. In the last decade, however, advances in the molecular characterization of the incompatibility reaction and in the analysis of allelic frequencies and allelic divergence from natural populations have stimulated new theoretical investigations that challenged some important assumptions of Wright's model of gametophytic self-incompatibility. We here review some of these recent empirical and theoretical advances that investigated: (i) the hypothesis that *S*-alleles are selectively equivalent, and the evolutionary consequences of genetic interactions between alleles; (ii) the occurrence of frequency-dependent selection in female fertility; (iii) the evolutionary genetics of self-incompatibility systems in subdivided populations; (iv) the evolutionary implications of the self-incompatibility locus's genetic architecture; and (v) of its interactions with the genomic environment.

Keywords: mating system, molecular evolution, natural selection, population genetics theory, self-incompatibility, self-recognition

Received 22 March 2004; revision received 28 May 2004; accepted 28 May 2004

Introduction

The ability to prevent self-fertilization is an essential feature of many plants' mating systems, that probably evolved as a means to avoid the deleterious effects of inbreeding. However, this trait is very variable both within and among species (Richards 1986), and interacts with a number of ecological and evolutionary processes (Jain 1976). A key to understanding this variation is to investigate the evolutionary properties of the genes controlling plant mating systems. Plant self-incompatibility (SI) systems are genetic systems that prevent self-fertilization in hermaphrodites through recognition and rejection of pollen expressing the same allelic specificity as that expressed in the pistils (de Nettancourt 2001). Two classes of SI are known, namely heteromorphic and homomorphic systems. Heteromorphic systems are characterized by morphological differences between the two or three genotype classes associated with different incompatibility specificities, particularly differences in the length or shape of the styles (heterostyly). In contrast, in homomorphic systems, the incompatibility genotypes cannot be distinguished morphologically, and the incompatibility response relies entirely on physiological mechanisms. For clarity, and because they have been characterized at the molecular level, this review will focus

Correspondence: Vincent Castric. Fax: (+ 33) 3 20 43 69 79; E-mail: Vincent.Castric@Univ-lille1.fr

on homomorphic SI systems only (see Barrett 1992 for a comprehensive coverage of heteromorphic systems). The evolutionary properties of homomorphic SI systems have been revealed through a fascinating interplay between empirical advances and theoretical developments, dating back to the year 1939 when Emerson published the first estimate of allelic diversity at the genetic locus responsible for gametophytic SI (*S*-locus) in *Oenothera organensis* (Emerson 1939), and Wright presented a model describing the evolutionary dynamics of alleles at the *S*-locus (Wright 1939).

Molecular characterization of the *S*-locus in several plant families has revealed that they evolved independently at least three times, with sharply different genetic, molecular and biochemical properties (Franklin-Tong & Franklin 2003). Genetically, homomorphic SI systems occur in two types, namely the gametophytic and the sporophytic systems, which differ with respect to the genetic determination of pollen specificity. In gametophytic SI, pollen specificity factors are expressed during pollen tube growth, and are encoded by the haploid genotype of the pollen tube. In sporophytic SI, both the diploid cells of the anther tapetum of the paternal plant and the pollen itself express pollen specificity proteins that are transferred to the pollen grain coat (Kusaba *et al*. 2002). In both systems, allelic specificities are generally expressed co-dominantly in the pistils, whereas complex dominance relationships among allelic specificities expressed in the pollen are generally observed in sporophytic SI. The biochemical and molecular details of the incompatibility process have been partly unravelled in two independently derived gametophytic SI systems, one originally described in the Solanaceae plant family, and the other known only in the Papaveraceae, as well as in one family with sporophytic SI, the Brassicaceae (see recent reviews in Franklin-Tong & Franklin 2003; Hiscock & McInnis 2003). In these three systems, distinct but tightly linked genes encode allelic specificities in the pistil and pollen, such that the genomic region controlling SI specificity and genetically defined as 'the *S*-locus' should actually be referred to as 'the *S*-genes complex'. For convenience, we will stick in the present review to the more usual '*S*-locus' terminology. At present, the pistil specificity gene is known in each of these systems (Anderson *et al*. 1986; Stein *et al*. 1991; Sassa *et al*. 1992; Foote *et al*. 1994; Xue *et al*. 1996; Takasaki *et al*. 2000), but the gene encoding pollen specificity has as yet been identified in the Brassicaceae only (Schopfer *et al*. 1999; Takayama *et al*. 2000), although candidates are being studied in other systems (McCubbin *et al*. 2000; Entani *et al*. 2003; Ikeda *et al*. 2004). Altogether, the biochemical mechanisms of SI in these different systems are so contrasted in terms of genetic determination, gene nature and gene expression pattern that it has become clear that SI evolved several times independently in the flowering plants (Steinbachs & Holsinger 2002 and references therein). In spite of this, homomorphic SI systems share several outstanding ecological and evolutionary features that have fascinated evolutionary biologists for over a century.

Common features shared by SI systems

First, the number of coexisting alleles at the *S*-locus (*S*-alleles) is usually very high, with between six and 36 *S*-alleles observed in single species (Table 1), corresponding to overall estimates of allelic richness ranging from 14 to 193 (Lawrence 2000). This stimulated Wright's theoretical analysis of the stationary frequency distribution of alleles at a gametophytic SI locus in a finite population (Wright 1939). Wright suggested that the main evolutionary force driving allelic frequencies at the *S*-locus was strong negative frequency-dependent selection acting on pollination success: rare pollen specificities are rejected by pistils at lower rates than those with common specificities. Wright showed quantitatively that this selection causes a substantial proportion of rare alleles arising from *de novo* mutations to escape loss through drift because of selection, and thus it generates high allelic diversity within populations. An essential feature of Wright's model is that all allelic specificities are selectively equivalent. Under this condition, Nagylaki (1975), Boucher (1993) and Steiner & Gregorius (1994) demonstrated that the deterministic equilibrium corresponds to strictly identical allelic frequencies. In finite populations, because of the strength of frequency-dependent selection, Wright assumed that allelic frequencies should lie very close to the deterministic equilibrium, an assumption that allowed mathematical treatment. Although Wright's analytical derivations have been criticised on several grounds (Fisher 1958; Moran 1962; Ewens 1964), early studies using numerical simulations confirmed his approximations (Ewens & Ewens 1966; Kimura 1965; Mayo 1966). In turn, Wright's predictions stimulated empirical studies that characterized the distribution of allelic frequencies in samples from natural populations, either through diallel crossing schemes (Atwood 1944; Campbell & Lawrence 1981; Levin 1993; Brennan *et al*. 2002) or through molecular typing of putative *S*-alleles (Richman *et al*. 1995, 1996b; Raspé & Kohn 2002). Although the number of studies has remained limited, most available evidence supports Wright's prediction of high allelic richness at the *S*-locus (Lawrence 2000; Table 1). In a review of empirical studies, Lawrence (2000) reported no departure from the identical allelic frequencies hypothesis in 16 out of 19 sampled populations from 12 species with gametophytic SI. Thus, although these tests have low statistical power because of low sample sizes, most species with gametophytic SI support the equal allelic frequencies hypothesis (Table 1). Interestingly, the three populations with strongly unequal allelic frequencies all belonged to the same species, *Papaver rhoeas*. A recent study

Table 1 Summary of empirical studies in natural populations of plants species with gametophytic or sporophytic SI

			Test of equal allelic frequencies		Nucleotide polymorphism		
Species	N_{ind}	$n_{alleles}$	χ^2 statistic	Significance	π	R_{SD}	References
Gametophytic SI system							
Crataegus monogyna	13	17	11.3	NS*	0.280	4.94†	Raspé & Kohn (2002)
Lycium andersonii	16	22	10.5	NS	0.463	4.74	Richman (2000)
						$P < 0.01$	Raspé & Kohn (2002)
Oenothera organensis	67	34	33.46	NS	—	—	Emerson (1939)
							Campbell & Lawrence (1981)
Papaver rhoeas	51	31	64.26	$P < 0.001$	—	—	Campbell & Lawrence (1981)
Phlox drummondii	24	30		NS	—	—	Levin (1993)
Physalis cinerascens	14	13	14.7	NS	0.208	2.59	Richman & Kohn (1999)
						NS	Richman & Kohn (2000)
Physalis crassifolia	22	28	17.8	NS*	0.387	2.59	Richman *et al*. (1996b)
						NS	Richman (2000)
							Raspé & Kohn (2002)
Prunus lannesiana	67	21	98.8	$P < 0.001$	—	—	Kato & Mukai (2004)
Solanum carolinense	24	12	3.3	NS	—	5.801	Richman *et al*. (1995)
						$P < 0.01$	Richman (2000)
Sorbus aucuparia	20	20	15.8	NS	0.251	7.39	Raspé & Kohn (2002)
Trifolium repens	25	36	13.84‡	NS‡	—	—	Atwood (1944)
Witheringia maculata	12	10	—	—	—	1.48	Richman & Kohn (2000)
						NS	
Sporophytic SI system							
Arabidopsis halleri	20	17	—	—	0.301	8.88	Castric and Vekemans, unpublished
Arabidopsis lyrata	20	11	26.10	$P < 0.001$	0.257	7.59	Mable *et al*. (2003)
							Charlesworth *et al*. (2003)
							Schierup *et al*. (2001b)
Brassica campestris	17	18	28.93‡	$P < 0.05$‡	0.130	8.87	Nou *et al*. (1993)
							Schierup *et al*. (2001b)
Ipomoea trifida	41	16	1144.12‡	$P < 0.0001$‡	—	—	Kowyama *et al*. (1994)
Raphanus raphanistrum	26	13	8.00‡	NS‡	—	—	Sampson (1964, 1967)
Raphanus sativus	29	22	—	—	—	—	Karron *et al*. (1990)
Senecio squalidus	25	6	13.16	$P < 0.05$	—	—	Brennan *et al*. (2003)
Sinapis arvensis	35	35	27.82‡	NS‡	—	—	Stevens & Kay (1989)

The results highlight that: (i) high allelic diversity has been found in most species; (ii) allelic frequencies are generally equal in gametophytic SI but unequal in sporophytic SI; (iii) nucleotide diversity at the *S*-locus is extremely high for intraspecific comparisons; (iv) *S*-allele genealogies are generally characterized by long terminal branches ($R_{SD} > 1$).

N_{ind} is the number of individuals sampled in the population. $n_{alleles}$ is the number of distinct alleles found in the sample. The χ^2 statistic and significance are the results of a χ^2 test of the allelic frequency distribution against expectation of equal frequencies among all alleles. π is the nucleotide diversity, i.e. the number of nucleotide differences per site between two randomly chosen sequences, at the *S*-locus pistil-expressed gene (computed for the *S*-domain of *SRK* in *Brassica* and *Arabidopsis*). R_{SD} is a statistic measuring the length of terminal branches in the allelic genealogy at the pistil-expressed gene, scaled by the size of the tree (expected $R_{SD} \approx 1$ under Wright's model of gametophytic SI according to Uyenoyama 1997). *P*-values below R_{SD} indicate the significance of departure from theoretical expectation, whenever this test was performed. When several natural populations were analysed in a given species, allelic diversity data were reported for the population with the largest sample size N_{ind} only. A dash indicates that the data were not available.

*For table-wide homogeneity purposes, we report the significance of the χ^2 statistic as computed from the theoretical χ^2 distribution instead of the value reported by the authors who computed an *ad-hoc* distribution of the χ^2 statistic by permutation; NS, not significant.

†R_{SD} statistic was computed based on published data according to Schierup *et al*. (2001b).

‡χ^2 statistics and tests were computed according to Campbell & Lawrence (1981) based on published data.

on two natural populations of the flowering cherry, *Prunus lannesiana*, also rejected the identical allelic frequency hypothesis in both study populations (Kato & Mukai 2004). Such consistent observations suggest that additional selective processes may occur in these species that are ignored by Wright's model.

A second common striking evolutionary feature consistently found in the different SI systems is unusually high

amino acid sequence divergence among *S*-alleles sampled within species (Ioerger *et al*. 1990; Kusaba *et al*. 1997), with for instance as little as 58% amino acid sequence identity between random pairs of *S*-alleles in the specificity-determining region of the pistil gene responsible for SI (*Aly13*) in *Arabidopsis lyrata* (Schierup *et al*. 2001a). This diversity is generally accompanied by trans-specific, or even transgeneric polymorphisms (Fig. 1, Dwyer *et al*. 1991; Lu 2001). To interpret data on *S*-allele molecular polymorphism, Wright's model of SI needed extending, since its original derivation dealt only with allele frequencies. Theoretical extensions based on Wright's model of gametophytic SI confirmed that extremely long average coalescence times are indeed expected for allelic genealogies at the selected locus (Clark 1993; Vekemans & Slatkin 1994). The rationale behind this property of the model is that fixation of a given *S*-allele in a population is very unlikely to occur under strong negative frequency-dependent selection, as opposed to the situation for a neutral allele. As a consequence, coalescence between *S*-allele lineages can only occur through the process of mutation, when a new mutant *S*-allele is generated from its parental copy, thus splitting the parental allele lineage. Because mutation rates are low, coalescence times are also very long on average. Takahata (1990) showed that the *shape* of the allelic genealogy under frequency-dependent selection was expected to be similar to that for a neutral gene, but that the *depth* of the genealogy (the time to the most recent common ancestor of all extant alleles) is expected to be expanded by several orders of magnitude. The estimated ages of the putative transgeneric polymorphisms observed in surveys of *S*-allele sequences in the Solanaceae were found to be roughly compatible with the expected time scale of allelic genealogies under gametophytic SI (Vekemans & Slatkin 1994; Uyenoyama 1997). This seems to be a general feature of SI systems, as shown by the high level of nucleotide diversity (π) among *S*-alleles observed in every species investigated so far [average $\pi = 0.284$ in nine species (Table 1), while an average $\pi = 0.029$ was reported for synonymous differences only for five genes unlinked to the *S*-locus in *Arabidopsis lyrata* (Wright *et al*. 2003)]. In contrast, several studies revealed that the observed shapes of the *S*-allele genealogies in the Solanaceae (Uyenoyama 1997) or the Brassicaceae (Schierup *et al*. 1998, 2001b) depart consistently from the expected topology under gametophytic SI or sporophytic SI, respectively. In all cases, terminal branches are considerably longer than expected (Table 1, but see Richman & Kohn 2000). Again, this observation suggests that Wright's model fails to capture important features of the *S*-alleles' evolutionary dynamics.

Third, although genealogies of *S*-alleles conferring different specificities have long coalescence times, theoretical studies predict that samples from different gene copies of the same allelic specificity should show fairly low average divergence (Vekemans & Slatkin 1994). This is because gene copies of the same specificity are expected to evolve neutrally as if in a population-equivalent, whose average size is approximated by the deterministic equilibrium frequency of allelic specificities ($\approx 2N/n$, where N is the population size and n is the actual number of distinct allelic specificities). Low nucleotide diversity among copies of the same allelic specificity has indeed been found in the very few studies of variation among allelic replicates (Richman *et al*. 1995; Miege *et al*. 2001; Charlesworth *et al*. 2003; see also *May et al*. 1999 in fungi). On this particular issue, empirical data lag behind the potential provided by theoretical developments.

Extending and challenging Wright's model of self-incompatibility

Overall, Wright's seminal model of frequency-dependent selection has been a rich source of inspiration for both empirical and theoretical research, and has remained mostly unchallenged until the mid-1990s. However, recent advances in elucidating the biochemical mechanism of SI, together with the accumulation of data on molecular polymorphism in natural populations of a growing number of species, have triggered new theoretical developments challenging different aspects of Wright's model. In this paper, we review these recent theoretical investigations, focusing on studies modelling the effects of (i) selective equivalence among *S*-alleles, (ii) frequency-dependent selection on female fecundity, (iii) population subdivision, (iv) the bipartite nature of the *S*-locus, and (v) evolution in genomic tracts linked to the *S*-locus. All these features represent successive steps towards an increasingly realistic description of SI in natural populations. Whenever appropriate, we therefore illustrate the continually fecund interplay that has occurred between theoretical and empirical advances since Wright's initial paper. From these studies, a more complex and challenging view of evolutionary processes in self-incompatibility systems is emerging.

S-alleles may be selectively non-equivalent

Wright's model of gametophytic SI assumes that all alleles at the *S*-locus are selectively equivalent, the 'equivalence' assumption. However, several types of selective non-equivalence among allelic specificities have been considered in recent theoretical and empirical investigations, with the following two main causes: the occurrence of dominance/recessive relationships between alleles; and the association of *S*-alleles with different sets of linked deleterious alleles.

From genotype to phenotype via dominance relationships

In species with sporophytic SI systems, such as species from the plant families Brassicaceae and Asteraceae,

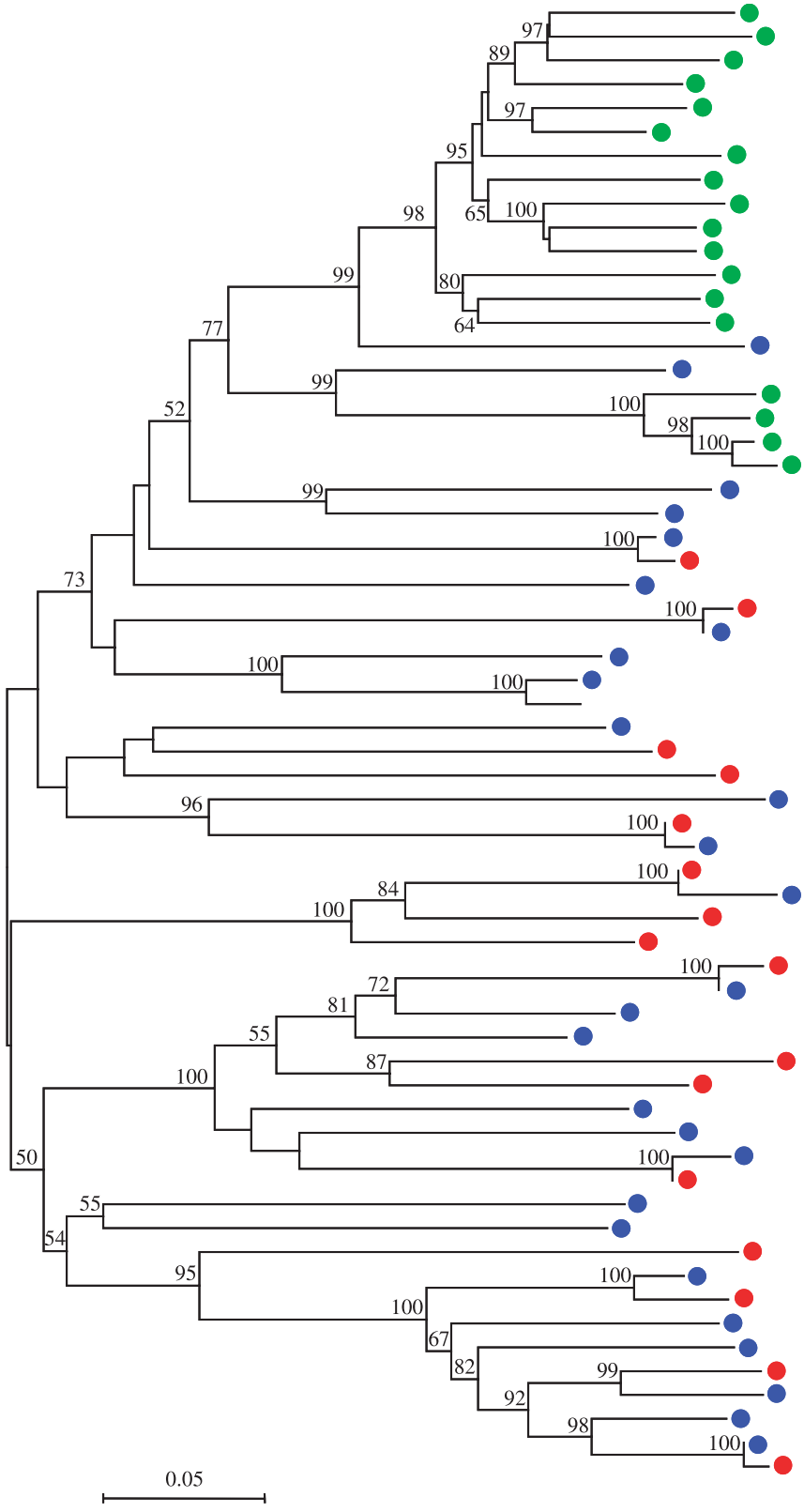

Fig. 1 Transgeneric and trans-specific polymorphism among *S*-alleles in the Brassicaceae. Two *Arabidopsis lyrata* alleles cluster within the group of *Brassica oleracea* alleles, suggesting maintenance over extended evolutionary times. The two closely related *A. halleri* and *A. lyrata* show a high proportion of shared alleles. Note the higher divergence among *Arabidopsis* alleles as compared to the *Brassica* alleles. Note also that all three species show very long terminal branches (see Table 1). Green dots indicate *B. oleracea* (GenBank accession numbers in Schierup *et al*. 2001a), blue dots indicate *A. lyrata* (Charlesworth *et al*. 2003), and red dots indicate *A. halleri* (Castric and Vekemans, unpublished data). Sequences ~600 base pairs long of the *S*-domain (primer pair 13Seq2F-SLGR, see Schierup *et al*. 2001a) were amino acid translated and aligned using CLUSTALW (Thompson *et al*. 1994) in MEGA3 (Kumar *et al*. 2004). The phylogeny was reconstructed using the Jukes & Cantor distance measure and the neighbour-joining procedure. Numbers above branch nodes represent bootstrap values > 50% and are based on 1000 replicates.

dominance interactions among certain alleles in the expression of the pollen and pistil phenotypes are known (Bateman 1952; Kowyama *et al*. 1994; Hatakeyama *et al*. 1998; Kusaba *et al*. 2002). In models that incorporate dominance, recessive alleles are expected to occur at higher equilibrium frequency than dominant alleles, the so-called 'recessive' effect (Cope 1962; Imrie *et al*. 1972). This arises from a general property of frequency-dependent selection models: at equilibrium all SI phenotypes are expected to have equal frequencies ('isoplethy', Finney 1952). Under serial dominance of allele specificities, a given SI phenotype comprises fewer different SI genotypes as recessivity increases, until finally the most recessive phenotypic class is represented by a single allele in a homozygous state. Therefore, isoplethy implies that allelic frequencies should increase with recessivity (see Uyenoyama 2000a for a complete coverage of this issue). Although this phenomenon was described long ago (e.g. Bateman 1952), evolutionary models of sporophytic SI systems in finite populations have only recently been explored in detail (Schierup *et al*. 1997; Uyenoyama 2000c). In a sporophytic system with a linear dominance hierarchy in both the pistil and pollen phenotypes, dominance/recessivity is expected to be a major determinant of an allele's evolutionary dynamics. On the one hand, the probability that new alleles that have arisen by mutation will become incorporated in a population increases with their dominance. On the other hand, the higher equilibrium frequency of recessive alleles reduces the probability of loss by genetic drift (Schierup *et al*. 1997). Altogether, dominant alleles are simultaneously more likely to become incorporated and also more likely to become lost from the population. Dominant alleles are thus expected to have a substantially higher turnover rate (shorter lifespan) than recessive alleles. This conclusion, however, does not hold under all possible dominance relationships. Under a model with full co-dominance in the pistil and a linear dominance hierarchy in the pollen, for instance, very different evolutionary dynamics arise (Schierup *et al*. 1997), whereby recessive alleles are more easily lost by drift than are dominant alleles. This situation may lead to a process of continuous evolution towards ever increasing dominance, as new mutant alleles will spread most readily if they are more dominant than the present alleles, such that all previous alleles will shift towards lower relative dominance levels. As alleles fall down this 'relative dominance ladder' during their lifespan, alleles of increasing age and recessivity tend to be preferentially lost from the population when they finally become the most recessive ones. Uyenoyama (2000c) studied a model with strict co-dominance in the pistil and two levels of dominance in pollen, with co-dominance between different alleles at the same level. This model was intended to mimic the situation described in self-incompatible *Brassica* species. Such a model does not allow evolution towards ever increasing dominance. She showed that several alleles from the dominant class were then expected to co-occur, all experiencing equivalent and regular turnover, while a single allele was expected to occur in the recessive class in most situations. This allele would impede the spread of other recessive alleles, so it is expected to have an extremely long lifespan. Hence, in this model, the equivalence assumption may hold only among alleles from the dominant class, not overall.

In spite of this lack of equivalence, expected allelic genealogies under these models are very similar to those for gametophytic SI systems (Schierup *et al*. 1998). A noteworthy difference, however, was expected for genealogies of gene copies of recessive alleles. Because recessive alleles have higher frequency and longer lifespan than dominant alleles, higher divergence among gene copies may be expected relative to dominant alleles. Again, this results in an expected difference among alleles in the level of nucleotide divergence among gene copies of the same allelic specificity.

Empirical studies aimed at determining dominance relationships between *S*-alleles in sporophytic systems have remained scarce. Evidence from several species of Brassicaceae indicates that co-dominant interactions between alleles are frequent, if not predominant, in the pollen or the stigma, that dominant/recessive interactions are expressed more often in the pollen than in the stigma, and that dominance in pollen and stigma are not necessarily correlated (Sampson 1964; Stevens & Kay 1989; Hatakeyama *et al*. 1998; Mable *et al*. 2003). Strict co-dominance in the stigma and either co-dominance or dominance in the pollen expression of incompatibility was reported for the species *Corylus avellana* (Betulaceae; Mehlenbacher 1997). In contrast, in species from Asteraceae (Samaha & Boyle 1989; Brennan *et al*. 2002) and from Convolvulaceae (Kowyama *et al*. 1994) dominance/recessive interactions were found between most alleles and were mostly consistent between pollen and stigma. Evidence for a higher frequency of the more recessive alleles has been found in several studies (Sampson 1964; Stevens & Kay 1989; Kowyama *et al*. 1994; Mable *et al*. 2003). In species of the genus *Brassica*, recessive and dominant alleles were found to constitute two distinct monophyletic groups of ancient origin based on their nucleotide sequence (Uyenoyama 1995), whereas no such evidence has been found in *Arabidopsis lyrata* (Mable *et al*. 2003).

As far as the limited number of empirical studies allows one to conclude, they roughly confirm theoretical predictions about the effect of dominance. The technical difficulty of precisely estimating dominance relationships seems to be the major impediment to testing the numerous theoretical predictions now available.

Lineage-specific loads of sheltered deleterious alleles

Because of the high heterozygosity observed at the *S*-locus in most systems, deleterious recessive mutations that

accumulate in the region surrounding it will normally not be exposed to purifying selection (Glémin *et al*. 2001; Stone 2004; see section below). Uyenoyama (1997) suggested that different sets of deleterious recessive mutations could therefore become associated with particular *S*-alleles as a result of the extended divergence time among *S*-allele lineages. For a given diploid genotype at the *S*-locus, mutations arising after the divergence of the two constituent haplotypes would occur in heterozygous form, while those that arose prior to divergence would be homozygous. Inbreeding depression is thus expected to be greater between more closely related *S*-alleles, which would thus share a higher proportion of identical deleterious alleles if there is a non-recombining region of some extent around the *S*-locus. In contrast, genes flanking anciently diverged *S*-alleles may have had time to recombine extensively. They may thus be as unrelated as randomly chosen pairs of alleles in the population, and no reduction in fitness is expected when they occur within a given zygote. This phenomenon should thus cause an additional source of non-equivalence among *S*-alleles. More strikingly, Uyenoyama (2003) showed that a strong negative interaction between a descendent new specificity (at low frequency) and its parental copy would systematically occur under this model, and this would decrease the probability of the new specificity successfully invading a population. As the population would then resist more strongly the invasion of new specificities, the rate of diversification of *S*-alleles would slow down over evolutionary time, so that terminal branch lengths in the phylogeny of *S*-alleles should be increased beyond the sole effect of balancing selection. This interaction is also expected to affect the process of allelic turnover. Descendent and parental specificities would tend to exclude each other, so that lineage replacement should become more frequent than lineage bifurcation.

Empirical estimates of the sheltered genetic load require experimental designs that can disentangle the genetic load from linked loci from that of the genome-wide effect of inbreeding depression at unlinked loci, and these estimates remain scarce. Stone (2004) used bud pollination to enforce crossing between identical genotypes in *Solanum carolinense*, a species with gametophytic SI, and quantified the load of deleterious mutations associated with seven different *S*-alleles (as estimated by seed abortion and reduced germination). Two out of the seven alleles showed a severe load of deleterious mutations. Interestingly, one of these two alleles also occurred on a longer-than-expected terminal branch in the genealogical tree reported by Richman *et al*. 1996a).

Evidence for uneven segregation of alleles at the S*-locus*

As shown by Lawrence & Franklin-Tong (1994) and Bechsgaard *et al*. (2004), *S*-alleles may show some level of uneven segregation, i.e. the two alleles of a diploid individual may not transmit equally to the progeny. Rather, they found that some alleles enjoyed up to a fourfold transmission advantage over others. Moreover, they showed that the strength of an allele's transmission advantage depended on the *S*-allele combination: the two alleles seemed to compete for transmission in a pairwise-specific manner. The theoretical implications of this additional source of selective non-equivalence among alleles remain unclear but may be part of the explanation for the observed cases of departure from the expected equilibrium allele frequencies of *S*-alleles in natural populations (Table 1). The observation by Bechsgaard *et al*. (2004) that the largest transmission advantage occurred in the most recessive *S*-allele in *Arabidopsis lyrata*, suggests that the sheltered load of deleterious alleles may be involved in this process. Indeed, recessive *S*-alleles in sporophytic SI systems may occur in the homozygous state (though arguably at a lower rate than neutral unlinked alleles), such that associated recessive deleterious alleles will be subject to purifying selection. Thus, a weaker load may be predicted to be sheltered by recessive *S*-alleles than by dominant ones.

Frequency-dependent selection acting on female fertility

Recent models have studied frequency-dependent selection at the *S*-locus acting on female as well as male fertility, and it was found to affect the evolutionary dynamics of *S*-alleles in populations with low proportions of compatible matings ('mate availability' Vekemans *et al*. 1998). The different systems of homomorphic SI have been found to differ with respect to the average mate availability in a population (Bateman 1952; Vekemans *et al*. 1998). Gametophytic SI should exhibit the highest mate availability because each heterozygous individual produces two kinds of pollen, each expressing one of the two distinct specificities. In sporophytic SI systems in contrast, pollen is of a single type. In such systems, mate availability increases with the proportion of dominant/recessive interactions between alleles, whereas co-dominance tends to decrease mate availability (Vekemans *et al*. 1998).

Mate availability also depends critically on population size, as the low number of alleles maintained in a small population may severely reduce the proportion of compatible matings.

This has raised a concern for the conservation of endangered SI species because the low mate availability expected in small populations may severely reduce seed-set (Byers & Meagher 1992). Low mate availability has indeed been held responsible for reduced seed-set in several endangered species with SI systems (Les *et al*. 1991; Godt & Hamrick 1995; Young *et al*. 2000) or even for complete

failure of sexual reproduction in clonal species as a result of geitonogamous pollination within genetically uniform patches of individuals (Aspinwall & Christian 1992; Thien *et al*. 1983; DeMauro 1993). For similar reasons, Baker (1955, 1967; 'Baker's law') and Anderson & Stebbins (1984) suggested that SI systems may prevent the establishment of populations after long-distance dispersal. In the case of the threatened endemic species *Centaurea corymbosa*, the lack of successful colonization of potentially suitable sites has been interpreted as a consequence of its sporophytic SI system (Colas *et al*. 1997).

In small populations, it has also been suggested that variance in seed-set among individuals could arise as a result of the variance in mate availability among SI genotypes (Byers & Meagher 1992). Because this process generates an association between seed-set (as a surrogate for fitness) and genotype at the *S*-locus, with the highest seed-set expected in genotypes with the rarest *S*-alleles, it leads to an additional component of frequency-dependent selection, acting on maternal success. Vekemans *et al*. (1998) named this component of selection 'fecundity selection' and investigated its effect on allelic diversity and allelic dynamics at the *S*-locus in finite populations under different SI models. The strength of fecundity selection depends strongly on the type of SI, with stronger selection in sporophytic than gametophytic SI. Within sporophytic SI, increasing strength of selection is associated with increasing proportions of co-dominant interactions among *S*-alleles. Interestingly, the efficiency of pollen dispersal also affects fecundity selection, which highlights the fact that the diversity of pollen genotypes deposited on a given stigma is a key parameter determining the strength of this selective process. Overall, fecundity selection increases the number of alleles maintained in finite populations, and thus increases mate availability in small populations. Paradoxically therefore although the *process* of fecundity selection derives from low mate availability, its evolutionary *consequences* partially compensate for this by increasing mate availability (Vekemans *et al*. 1998).

We are currently unaware of any empirical study that has tested for fecundity selection in natural populations, yet several predictions could be tested empirically. First, the theory predicts that the relationship between female fertility and *S*-locus genotype should be more important in species with sporophytic than with gametophytic SI. Second, for any SI species, small isolated populations should be more affected by fecundity selection than large and interconnected populations. Thirdly, the strength of fecundity selection in species with sporophytic SI should correlate with the frequency of co-dominant interactions among *S*-alleles. We anticipate that such empirical tests will become increasingly practicable as SI systems are studied in more model species.

Extension of Wright's model to a subdivided population

Many, if not all, natural plant populations are spatially structured by local mating, limited dispersal and/or habitat patchiness. As a result, both neutral and selected genes show patterns of spatial genetic structure (reviewed in Rousset 2004). Empirical estimates as well as theoretical predictions about the structure of genetic diversity of selected genes, however, lag far behind those for neutral loci. Recently, remarkable progress has been made in models of balancing selection in subdivided populations. The advent of *S*-locus genotyping techniques will permit the use of SI as a model system.

Number of S*-alleles in a subdivided population*

Sewall Wright himself was the first to improve his original model to investigate the effect of population structure under an island model of migration on the number of alleles, in the second part of his 1939 paper. He conjectured that the total number of alleles maintained in a finite subdivided population under recurrent mutation should be higher than in a panmictic population of the same size, because different demes would have different sets of alleles, compensating for and outweighing the presence of fewer alleles within each deme. Although the conjecture that subdivision increases allelic richness was later shown to be correct for neutral alleles (Nagylaki 1985), it is only correct under gametophytic SI for unrealistically low migration rates among demes (Schierup 1998). Rather, numerical simulations showed that the total number of *S*-alleles in a subdivided population was generally smaller than under panmixia (Schierup 1998). Similar results were obtained by Neuhauser (1999) using numerical simulations of frequency-dependent selection in a spatially structured population. The key to this unexpected result lies in the observation that interactions between migration and selection lead to genetic differentiation among demes at an *S*-locus being much lower than expected at a neutral locus (Schierup *et al*. 2000a). Indeed, upon introduction into a new deme by migration or mutation, neutral alleles have a probability $1/(2N)$ of becoming fixed in the deme, irrespective of their frequency in the recipient deme, where N is the number of individuals in the deme. Migrating *S*-alleles, in contrast, experience positive frequency-dependent selection whenever they are rare in the recipient population, such that migration often results in successful invasion. Hence *S*-alleles migrate more 'efficiently' than neutral alleles. As a consequence, the local loss of *S*-alleles as a result of drift within each deme when migration decreases is not balanced, as it is for neutral alleles, by the maintenance of different sets of alleles in different demes of the population, and subdivision actually decreases the

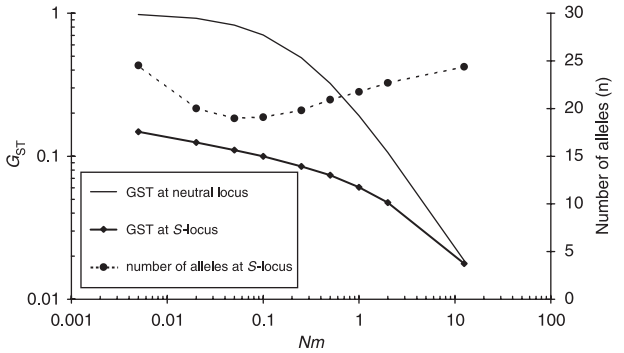

Fig. 2 Expected variation in the number of alleles maintained at the *S*-locus in a subdivided population, and in the level of among-demes differentiation (expressed as Nei's G_{ST}), as a function of migration rate among demes. Data obtained from numerical simulations of an island model of migration with 40 demes of 50 individuals each and a mutation rate to new *S*-alleles of 10^{-6} per generation (from Schierup *et al*. 2000a). The figure illustrates that the number of alleles in the total population first decreases and then increases with decreasing migration among demes, with the latter occurring only for extremely low levels of migration. For a neutral locus, the number of alleles is expected to increase monotonously with decreasing migration (not shown). The relationship between Nei's G_{ST} and migration rate also illustrates that differentiation among demes at the *S*-locus is expected to be about an order of magnitude lower than at a reference neutral locus.

number of alleles maintained (Fig. 2). Muirhead (2001) later confirmed these findings and provided an analytical solution to the problem of subdivision by modelling allelic turnover in a population as a Markov chain for the proportion of shared alleles between demes under the strong selection–weak mutation scheme. Although the number of selected alleles maintained in a population bears a more complex relationship with subdivision than neutral alleles, it is also almost insensitive to variation in the migration rate across a biologically meaningful range, assuming strong selection and infrequent migration. Hence, the expected total number of alleles maintained in a species by frequency-dependent selection should not be greatly affected by population structure, unless subdivision is exceedingly strong. Consequently, variation in the level of population subdivision among species is unlikely to influence strongly the variation in the total amount of polymorphism at *S*-loci. A rigorous test of this prediction would involve comparing numbers of alleles in subdivided vs. non-subdivided populations, which would be difficult to achieve. Yet this theoretical prediction has important implications for the sampling strategy of empirical studies aiming at estimating allelic richness in natural populations: unless populations are totally isolated from each other, panmictic units with respect to the *S*-locus are expected to extend over very wide geographical areas.

Distribution of S*-alleles among demes*

Stimulated by the possibility of access to individual genotypes at the *S*-locus provided by molecular tools (Brace *et al*. 1993; Richman *et al*. 1995), Schierup *et al*. (2000a) used numerical simulations to investigate the effect of restricted migration in an island model on genetic differentiation among demes. Differentiation at the *S*-locus is predicted to remain low even under very restricted migration, as a result of the higher 'effective' migration rate of *S*-alleles as compared to neutral alleles (Fig. 2). These results were confirmed by Muirhead (2001) who derived the distribution pattern of alleles across demes by obtaining analytically the equilibrium expectation for the percentage of alleles in a deme that are shared between k demes of a structured population (private to a single deme, found in exactly two demes, three demes, k demes …). These statistics are similar to the 'occupancy distribution' studied by Slatkin & Charlesworth (1978) and provide a more complete description of population genetic structure than F_{ST}. Remarkably, the pattern of distribution of alleles across demes at equilibrium is independent of N, the number of individuals found in a deme, and of the strength of the frequency-dependent selection, and depends only on the ratio of the migration (m) to the mutation rate (u) and the number of demes in the population.

One limit to these analytical investigations is their reliance upon the assumption of exchangeability among alleles. Although this equivalence assumption may hold true for gametophytic SI (with the restriction that different deleterious mutations may be sheltered by different *S*-alleles, see above), dominance relationships among alleles in sporophytic SI clearly violate this assumption. Dominance relationships make the model analytically intractable at the present time, so predictions for these systems all stem from simulation studies. As shown by Schierup *et al*. (2000a), dominance affects the distribution of alleles and the overall expected level of differentiation among demes (as estimated by G_{ST}), and may either increase (if dominance occurs in both pollen and pistil) or decrease (if there is co-dominance in pollen and/or pistil) compared with gametophytic SI. More importantly, dominant and recessive alleles may be affected differently by subdivision. Under some circumstances, dominant alleles migrate more effectively among demes than recessive alleles, and should thus be distributed more homogeneously across demes. Such allele-specific distribution differences further highlight the need for more detailed data on population genetic structure at the level of individual alleles, beyond summary statistics such as F_{ST}.

More realistic models of population subdivision than the island models have not been explored in detail. Brooks *et al*. (1996) and Neuhauser (1999) used computer simulations to investigate the spatial distribution of *S*-alleles

under isolation by distance across a continuous two-dimensional finite habitat. Although slight clustering of *S*-alleles was found, neither study compared the intensity of clustering with that of neutral loci, nor did they investigate the effect of variation in mating or dispersal distance. Thus, although the general conclusion fits with the expectation that *S*-alleles experience more effective migration rates than neutral alleles, more extensive simulations will be required to assess the role of restricted dispersal in generating isolation-by-distance patterns at *S*-loci.

Agreeing with theoretical expectations, most empirical studies in natural populations have found low differentiation at the *S*-locus among populations. In *Oenothera organensis*, on average 56% of *S*-alleles were shared between population pairs from three different canyons of the Organ Mountains in New Mexico, with 17–23 alleles scored within samples (Emerson 1939). In horsenettle, *Solanum carolinense*, two populations about 200 km apart had 10 *S*-alleles in common out of an overall total of 13 (Richman *et al*. 1995). In *Papaver rhoeas*, about 93% of *S*-alleles were shared between population pairs in a survey of three British populations with 25–30 alleles identified in each sample (O'Donnell *et al*. 1993). In flowering cherry, *Prunus lannesiana*, 95% of *S*-alleles were shared between two populations separated by 5 km, in spite of strong allelic frequency differences (Kato & Mukai 2004), possibly because of recent founder events associated with volcanic activity (S. Kato, personal communication).

Empirical studies comparing levels of population structure at the *S*-locus with that of neutral marker loci are needed to support theoretical predictions convincingly. Yet, such studies are scarce in the literature. Within a single population of *Senecio squalidus*, Brennan *et al*. (2003) compared spatial genetic structure between the *S*-locus vs. allozyme loci using spatial autocorrelation methods. Overall, no significant spatial genetic structure was found either at the *S*-locus or at allozyme loci, but this could have been because the power was too low to detect a pattern, given the very low sample size used (24 individuals).

Evolutionary consequences of population subdivision

The high effective migration rates of *S*-alleles have several evolutionary consequences. First, population isolation is generally expected to reduce maintenance of alleles within local demes. Accordingly, concern has been expressed by the plant conservation community (e.g. DeMauro 1993) about the risk incurred by self-incompatible endangered species of falling below the minimal number of *S*-alleles required for compatible crosses to occur (three for gametophytic SI, two for sporophytic SI). However, the evolutionary properties described above show that this decrease in the number of alleles is very unlikely to occur unless the population experiences extremely strong isolation. Even in such cases, the operation of fecundity selection under small population size should counteract the loss of alleles by drift (see above). Contrary to a common belief, theory thus predicts that SI species should not be particularly more threatened by habitat fragmentation than self-compatible species.

Second, population structure may allow the coexistence of a new allelic specificity arisen by mutation together with its parental copy. As shown by Uyenoyama (2003), a new allelic specificity and its parental copy are expected to exclude each other in some situations if they both occur in a panmictic population. Population structure may thus allow different descendent specificities to evolve independently in different demes of a structured population, and later coexist upon their subsequent introduction into the same deme, increasing allelic richness (Uyenoyama *et al*. 2001). This conjecture remains to be tested, especially as the weak population structure expected at the *S*-locus may not permit sufficient allele frequency differences to evolve.

Third, because hybridization among related species is not infrequent in plants, *S*-alleles can cross inter–specific barriers more easily than alleles from neutral loci, as introduced alleles increase in frequency. Thus, very low hybridization rates may be sufficient to maintain or restore similar sets of alleles at the *S*-locus among closely related species. Because of hitchhiking, alleles at closely linked loci may also undergo introgression among species together with the *S*-allele to which they are linked.

The evolution of SI systems with two linked genes

It has recently been established in both gametophytic and sporophytic SI systems that the genomic region referred to as 'the *S*-locus' contains two distinct genes, the determinants of pollen and pistil specificity (Golz *et al*. 2000; Nasrallah *et al*. 2000). In the genera *Brassica* and *Arabidopsis* (Brassicaceae), with a sporophytic SI system, both component genes have been identified (Schopfer *et al*. 1999; Takayama *et al*. 2000; Kusaba *et al*. 2001). The pollen gene (*SCR*, also coined *SP11*) is expressed in the anthers (and, for some alleles, pollen) and encodes a small protein deposited in the pollen exine layer. This protein specifically binds and activates the protein encoded by the pistil gene (*SRK*), which is expressed at the stigma surface. In all haplotypes investigated to date, the two genes are tightly linked, although their relative order and orientation vary (Boyes *et al*. 1997). In genera with gametophytic SI, the pollen specificity has been recently identified in Petunia (Sijacic *et al*. 2004). The presence of two linked genes has profound implications for the evolution of the *S*-locus itself and for the evolution of its genomic environment. The distinction between pollen and pistil genes generates strong restrictions on the level of recombination allowed in the *S*-locus

genomic region, and on the scenarios under which new allelic specificities can evolve. Moreover, the development of this bipartite model was accompanied by empirical evidence that *S*-allele specificities may overlap, thus challenging Wright's hypothesis that one allele univocally corresponds to one specificity.

How can new allelic specificities arise under the bipartite model of SI?

The bipartite structure of the *S*-locus implies that both genes must undergo concerted evolution for new allelic specificities to arise. A change in specificity as a result of mutation at only one of the two component genes will merely lead to a breakdown of SI in individuals carrying the mutant haplotype. A fully functional new specificity will only arise when a suitable mutation occurs on the same haplotype at the second component gene, such that this scenario implies that the non-functional haplotype segregates in the population until complementary mutations restore a fully functional new SI haplotype (Uyenoyama *et al*. 2001). However, a self-fertile haplotype would have low fitness because of inbreeding depression (Glémin *et al*. 2001), so it is likely to be lost from the population before the occurrence of the second mutation restoring specificity. Two alternative scenarios that do not involve self-compatible intermediates have been suggested for the evolution of new specificities. The first possibility involves mutations conferring dual specificity recognition, as obtained experimentally by Matton *et al*. (1999). This evolutionary scenario, however, has been criticised, based on theoretical arguments (Charlesworth 2000; Uyenoyama 2000b; Uyenoyama *et al*. 2001). The second scenario is based on the experimental demonstration that several mutations at the pollen gene in *Brassica* may be necessary to produce a non-overlapping new specificity (Chookajorn *et al*. 2004). The authors suggested that the occurrence of polymorphism in both component genes of a given specificity would allow gradual evolution of the cloud of variants sharing a given specificity into two distinct co-evolved specificities. Theoretical studies taking into account the bipartite model of SI are needed to investigate whether sufficient polymorphism can be maintained within specificity classes, as this is quite unlikely under Wright's model (Vekemans & Slatkin 1994).

Selection for optimal specificity acting on the two component genes

Overlap of allele specificities restricts the mate availability in a population without improving the protection against consanguineous crosses. Thus, according to Richman (2000), 'the net benefit for any particular allele may be to differ as much as possible from other competing *S*-genotypes, to avoid false rejection of nonself pollen'. Richman & Kohn (1999) compared the *S*-allele polymorphism in two species *Physalis crassifolia* and *P. cinerascens*. Because they are closely related, both species should have inherited the same set of *S*-alleles from their common ancestor. Yet, one species (*P. cinerascens*) displayed much lower allelic diversity than the other, probably because of a strong demographic bottleneck. Controlling for sampling artefacts, they showed that *S*-alleles had been lost in *P. cinerascens* since speciation in relationship to their phylogenetic proximity, such that alleles possessing a closely related allele at the time of the bottleneck were more frequently lost than alleles with no such counterpart. They suggested a selective mechanism in which the ability to discriminate non-self pollen increased as a function of the degree of sequence divergence between alleles in the pollen and the style. This hypothesis was empirically tested by Chookajorn *et al*. (2004), who recently performed experimental manipulation of pollen gene sequence using targeted modifications (domain swapping) to investigate how pollen binding affinity and specificity were affected. They identified protein domains whose alteration affected pollen binding affinity and specificity and showed that specificity of the pollen gene was robust to a large number of changes. Several alleles sharing the same pollen specificity may thus coexist within a single population, such that a specificity should be described as a cloud of alleles with variable binding affinities towards other alleles rather than as a single allelic sequence. According to the mechanism proposed in Richman (2000), selection should thus favour variants that belong to the cloud of specificities with the most restricted set of overlapping specificities.

The bipartite model of SI may select for reduced recombination within the S-locus region

The bipartite structure of the *S*-locus also implies that recombination may disrupt linkage between the two genes. Such recombination events would lead to self-compatible haplotypes because they would encode different pistil and pollen specificities. Similarly to haplotypes carrying mutations to a new specificity at only one component gene, recombinant haplotypes would be strongly selected against because selfed offspring would suffer from severe inbreeding depression. This phenomenon may indirectly select for haplotypes with low levels of recombination, because of the deleterious effects of the self-compatibility generated by recombination events between the two component genes.

S-haplotypes are highly structurally diverse in both *Brassica* (Boyes *et al*. 1997) and *Arabidopsis lyrata* (Kusaba *et al*. 2001). The observed important structural rearrangements suggest that recombination within the *S*-locus region is indeed very unlikely (but see Casselman *et al*. 2000). In

several species from the Solanaceae, the *S*-locus seems to be located close to the chromosome centromere (Brewbaker & Natarajan 1960; Golz *et al*. 1999; in *Nicotiana alata*; Bernacchi & Tanksley 1997 in *Lycopersicon*; Entani *et al*. 1999 in *Petunia hybrida*) with extensively repeated sequences (Coleman & Kao 1992) and several lines of evidence indicate that recombination is very low or even suppressed (Wang *et al*. 2003). Reduced recombination would result in strong haplotypic structure that should be evident in the structure of sequence diversity observed within natural populations. However, empirical evidence from natural populations only partially supports this prediction. In a survey of *S*-allele sequences from 21 species with gametophytic SI, evidence for rare intragenic recombination events was obtained in at least three species (Vieira *et al*. 2003), whereas in two species with sporophytic SI, evidence for recombination was found in *Brassica campestris* (Awadalla & Charlesworth 1999) but not in *Arabidopsis lyrata* (Charlesworth *et al*. 2003). Clearly, more empirical studies are needed to determine whether the *S*-locus region exhibits unusually high levels of rearrangement as compared to other genomic regions (SanMiguel *et al*. 1998; Ilic *et al*. 2003), and whether reduced recombination translates into patterns of genetic diversity around the *S*-locus. Ideally, such studies would compare recombination at the *S*-locus with that at other regions of the genome, and would provide a better assessment of the variation in recombination rates across the genome.

Evolution in genomic regions linked to the *S*-locus

Advances in molecular methods during the last decade have provided details of the nature, structure and diversity of the genomic context of the *S*-locus in several plant families. Remarkably, the *S*-locus seems to lie in a region of low recombination in most species, such that hitchhiking processes between variants at nearby loci and particular *S*-alleles may be important. Several theoretical models have been developed of the evolutionary implications of the two-way interaction between the *S*-locus and its genomic environment and they addressed two important questions. (i) How does selection at the *S*-locus influence diversity in linked genomic regions? (ii) Conversely, do evolutionary forces acting on the genomic context of the *S*-locus interfere with the evolutionary dynamics of the *S*-locus?

Influence of the S*-locus on polymorphism at partially linked neutral sites*

For a neutral locus, linkage to a locus under balancing selection is expected to increase nucleotide diversity (Strobeck 1983; Hudson & Kaplan 1988). This results from the non-random association of variants at the neutral locus with different allelic specificities at the selected locus, which tends to protect the neutral variants against fixation through genetic drift. Under partial linkage, however, variants at the neutral locus can still move between allelic specificities by recombination. Hence, from an empirical perspective, the occurrence and frequency of recombination are critically important parameters because they determine the genetic distances over which the molecular signature of natural selection at the selected locus may be expected to extend. Building on the foundations of Wright's model, Schierup *et al*. (2000b) recently provided predictions of the effect of selection at the *S*-locus on diversity of linked loci. They used computer simulations to investigate the consequences of linkage of a neutral locus to the *S*-locus in a subdivided population. The linked neutral locus shares the high effective migration of *S*-alleles (see above), and therefore also has decreased divergence among populations (as measured by F_{ST} or G_{ST}, Fig. 2). Further from the *S*-locus (i.e. with increased recombination rate), divergence should increase monotonically and eventually reach that expected for unlinked loci. Thus, a 'valley' of population differentiation is expected around the *S*-locus.

The effect of linkage on the number of alleles maintained in the overall population and their nucleotide diversity is less straightforward because it depends on details of the sampling scheme used to estimate the number of alleles. Since both population subdivision and linkage to the *S*-locus tend to increase diversity as compared to a neutral locus in a panmictic population, both processes may intuitively be expected to act synergistically, and a strong peak of diversity might be expected around the *S*-locus when alleles are sampled from a single deme. Population subdivision and linkage to the *S*-locus are not independent processes, because linkage results in greater effective migration among demes and therefore decreases the effect of population subdivision. The net outcome of these opposed forces is that alleles collected over the whole population should coalesce faster for intermediate levels of linkage (Schierup *et al*. 2000b). In other words, the complex interaction between linkage and population subdivision can reduce the levels of genetic diversity, even though each factor taken separately tends to increase diversity. The ability to detect balancing selection by investigating nucleotide polymorphism at linked neutral loci therefore critically depends on an adequate understanding of population structure.

Interestingly the expected 'peak' of nucleotide diversity should be markedly narrower than the expected 'valley' of population divergence around the *S*-locus (Schierup *et al*. 2000b). Thus, the assessment of population genetic structure of neutral linked loci (F_{ST} analysis, for instance) may provide a more powerful way to characterize the footprints left by balancing selection on the genomic

neighbourhood of the *S*-locus, than investigations based on either nucleotide polymorphism, or the number of alleles alone.

Population inbreeding depression and the sheltering of deleterious mutations at loci partially linked to the S-locus

Self-incompatibility is thought to have evolved and been maintained in hermaphrodite plants as a response to the deleterious effects of inbreeding (Charlesworth & Charlesworth 1979; Uyenoyama 1988, 1989). Yet paradoxically, genes around the *S*-locus are expected to exhibit a large number of sheltered deleterious alleles. This is because most deleterious mutations are partially recessive (Peters *et al*. 2003), and are selected against largely when they occur in homozygotes. Yet, homozygotes are much less frequent at the *S*-locus (and even absent under gametophytic SI) as compared to other loci. Thus, the magnitude of inbreeding depression as a result of the load of mutations sheltered by the *S*-locus depends upon the width of the genomic region with decreased homozygosity, which increases with decreasing recombination (Glémin *et al*. 2001). Glémin *et al*. (2001) showed that F_{IS} (the reduction in heterozygosity as compared to Hardy–Weinberg expectation, Wright 1951) in a finite isolated population of a species with gametophytic SI decreases with the recombination rate r between the *S*-locus and a locus accumulating deleterious mutations at a rate μ. One effect of linkage to the *S*-locus is to reverse the effect of population size on the level of inbreeding depression. Inbreeding depression generally decreases as population size declines because genetic drift allows fixation of a large number of deleterious mutations, so that the population maintains less variation and expresses lower inbreeding depression (Bataillon & Kirkpatrick 2000). Linkage to the *S*-locus offsets this effect, because selection on the *S*-locus is stronger in small populations, and the large amount of variation maintained at linked loci around the *S*-locus (see above) ensures that inbreeding depression is maintained even if the population is very small. This sheltered genetic load may play an important role in stabilizing SI systems, by counteracting the evolution of mutations at the *S*-locus conferring self-compatibility. This effect is important because self-compatibility mutations may benefit from male transmission advantages and are expected to evolve readily in the absence of counteracting forces (Uyenoyama *et al*. 2001).

To our knowledge, Stone (2004) provided the first empirical estimate of the load of deleterious mutations specifically linked to the *S*-locus. She used bud pollination to cross plants that shared *S*-alleles but that were otherwise unrelated. Such crosses produced low-fitness offspring zygotes, with rates of seed abortion nearly as high as that of selfed flowers. These results showed that a major part of the genome-wide inbreeding depression was the result of the load of deleterious mutations sheltered by the *S*-locus.

Conclusions

Despite these recent theoretical and empirical developments, many evolutionary questions remain to be elucidated. In particular, the mechanism generating new allelic specificities, in spite of the many scenarios presented, still needs both empirical and theoretical clarification. The explanation for the recurrent observation that terminal branches in allelic genealogies at the *S*-locus are longer than expected (Table 1) is also a matter of debate. Another example concerns the evolutionary circumstances leading to a breakdown of SI, which need to be investigated in the light of the recent discoveries of a potentially important load of sheltered deleterious alleles co-segregating with individual *S*-alleles.

Acknowledgements

We are most grateful to Barbara Mable (University of Guelph) for sharing the alignment of *Arabidopsis lyrata S*-alleles used in Fig. 1. Detailed comments by Sylvain Glémin, Deborah Charlesworth, Mathilde Dufaÿ and two anonymous referees were highly appreciated. V.C. is supported by a CDD-postdoctoral grant by CNRS, and X.V. is supported by a CNRS-ATIP grant from the Life Science Department, by a FEDER grant from the E.U., and by an ARCIR grant from the Région Nord-Pas de Calais.

References

Anderson GJ, Stebbins GL (1984) Dioecy versus gametophytic self-incompatibility: a test. *American Naturalist*, **144**, 423–428.

Anderson MA, Cornish EC, Mau S-L *et al*. (1986) Cloning of cDNA for a stylar glycoprotein associated with expression of self-incompatibility in *Nicotiana alata*. *Nature*, **321**, 38–44.

Aspinwall N, Christian T (1992) Pollination biology, seed production, and population structure in queen-of-the-prairie, *Filipendula rubra* (Rosaceae) at Botkin fen, Missouri. *American Journal of Botany*, **79**, 488–494.

Atwood SS (1944) Oppositional alleles in natural populations of *Trifolium repens*. *Genetics*, **29**, 428–435.

Awadalla P, Charlesworth D (1999) Recombination and selection at *Brassica* self-incompatibility loci. *Genetics*, **152**, 413–425.

Baker HG (1955) Self-compatibility and etablishment after 'long-distance' dispersal. *Evolution*, **9**, 347–348.

Baker HG (1967) Support for Baker's law as a rule. *Evolution*, **21**, 853–856.

Barrett SCH (1992) *The Evolution and Function of Heterostyly*. Springer-Verlag, Berlin.

Bataillon T, Kirkpatrick M (2000) Inbreeding depression due to mildly deleterious mutations in finite populations: size does matter! *Genetical Research*, **75**, 75–81.

Bateman AJ (1952) Self-incompatibility systems in Angiosperms I. Theory. *Heredity*, **6**, 285–310.

Bechsgaard J, Bataillon T, Schierup MH (2004) Uneven segregation of sporophytic self-incompatibility alleles in *Arabidopsis lyrata*. *Journal of Evolutionary Biology*, **17**, 554–561.

Bernacchi D, Tanksley SD (1997) An interspecific backcross of *Lycopersicon esculentum* × *L. hirsutum*: linkage analysis and a QTL study of sexual compatibility factors and floral traits. *Genetics*, **147**, 861–877.

Boucher W (1993) A deterministic analysis of self-incompatibility alleles. *Journal of Mathematical Biology*, **31**, 149–155.

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.

Brace J, Ockendon D, King GJ (1993) Development of a method for the identification of *S*-alleles in *Brassica oleracea* based on digestion of PCR-amplified DNA with endonucleases. *Sexual Plant Reproduction*, **6**, 133–138.

Brennan AC, Harris SA, Tabah DA, Hiscock SJ (2002) The population genetics of sporophytic self-incompatibility in *Senecio squalidus* L. (Asteraceae). I. *S*-allele diversity in a natural population. *Heredity*, **89**, 430–438.

Brennan AC, Harris SA, Hisckock SJ (2003) Population genetics of sporophytic self-incompatibility in *Senecio squalidus* L. (Asteraceae) II: a spatial autocorrelation approach to determining mating behaviour in the presence of low *S* allele diversity. *Heredity*, **91**, 502–509.

Brewbaker JL, Natarajan AT (1960) Centric fragments and pollen part mutation of incompatibility alleles in *Petunia*. *Genetics*, **45**, 699–704.

Brooks RJ, Tobias AM, Lawrence MJ (1996) The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. XI. The effects of limited pollen and seed dispersal, overlapping generations and variation in plant size on the variance of *S*-allele frequencies in populations at equilibrium. *Heredity*, **76**, 367–376.

Byers DL, Meagher TR (1992) Mate availability in small populations of plant species with homomorphic sporophytic self-incompatibility. *Heredity*, **68**, 353–359.

Campbell JM, Lawrence MJ (1981) The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. II. The number and frequency of *S*-alleles in a natural population (R106). *Heredity*, **46**, 81–90.

Casselman AL, Vrebalov J, Conner J *et al*. (2000) Determining the physical limits of the *Brassica* S locus by recombinational analysis. *Plant Cell*, **12**, 23–34.

Charlesworth D (2000) How can two-gene models of self-incompatibility generate new specificities? *Plant Cell*, **12**, 309–310.

Charlesworth D, Charlesworth B (1979) The evolutionary genetics of sexual systems in flowering plants. *Proceedings of the Royal Society of London B*, **205**, 513–530.

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, Clark AG, Nasrallah JB (2004) Specificity determinants and diversification of the *Brassica* self-incompatibility pollen ligand. *Proceedings of the National Academy of Science USA*, **101**, 911–917.

Clark AG (1993) Evolutionary inferences from molecular characterization of self-incompatiblity alleles. In: *Mechanisms of Molecular Evolution* (eds Takahata N, Clark AG), pp. 79–108. Sinauer, Sunderland, MA.

Colas B, Olivieri I, Riba M (1997) *Centaurea corymbosa*, a cliff-dwelling species tottering on the brink of extinction: a demographic and genetic study. *Proceedings of the National Academy of Sciences of the USA*, **94**, 3471–3476.

Coleman CE, Kao T-H (1992) The flanking regions of two *Petunia inflata S* alleles are heterogeneous and contain repetitive sequences. *Plant Molecular Biology*, **18**, 725–737.

Cope FW (1962) The effects of incompatibility and compatibility on genotype proportions in populations of *Theobroma cacao* L. *Heredity*, **17**, 183–195.

DeMauro MM (1993) Relationship of breeding system to rarity in the lakeside daisy (*Hymenoxys acaulis* var. *glabra*). *Conservation Biology*, **7**, 542–550.

Dwyer KG, Balent MA, Nasrallah JB, Nasrallah ME (1991) DNA sequences of self-incompatibility genes from *Brassica campestris* and *B. oleracea*: polymorphism predating speciation. *Plant Molecular Biology*, **16**, 481–486.

Emerson S (1939) A preliminary survey of the *Oenothera organensis* population. *Genetics*, **24**, 524–537.

Entani T, Wano MI, Shiba H, Takayama S, Fukui K (1999) Centromeric localization of an *S*-RNase gene in *Petunia hybrida* Vilm. *Theoretical and Applied Genetics*, **99**, 391–397.

Ewens WJ (1964) On the problem of self-sterility alleles. *Genetics*, **50**, 1433–1438.

Ewens WJ, Ewens PM (1966) The maintenance of alleles by mutation–Monte Carlo results for normal and self-sterility populations. *Heredity*, **21**, 371–378.

Finney DJ (1952) The equilibrium of self-incompatible polymorphic species. *Genetica*, **26**, 33–64.

Fisher RA (1958) *The Genetical Theory of Natural Selection*, 2nd revised edn. Dover Publications, New York.

Foote HCC, Ride JP, Franklin-Tong VE *et al*. (1994) Cloning and expression of a distinctive class of self-incompatibility in flowering plants. *Proceedings of the National Academy of Sciences of the USA*, **91**, 2265–2269.

Franklin-Tong VE, Franklin FCH (2003) Gametophytic self-incompatibility inhibits pollen tube growth using different mechanisms. *TRENDS in Plant Science*, **8**, 598–605.

Glémin S, Bataillon T, Ronfort J, Mignot A, Olivieri I (2001) Inbreeding depression in small populations of self-incompatible plants. *Genetics*, **159**, 1217–1229.

Godt MJW, Hamrick JL (1995) The mating system of *Liatris helleri* (Asteraceae), a threatened plant species. *Heredity*, **75**, 398–404.

Golz JF, Su V, Clark AE, Newbigin E (1999) A molecular description of mutations affecting the pollen component of the *Nicotiana alata S* locus. *Genetics*, **152**, 1123–1135.

Golz JF, Clark AE, Newbigin E (2000) Mutational approaches to the study of self-incompatibility: revisiting the pollen-part mutants. *Annals of Botany*, **85** (Suppl. A), 95–103.

Hatakeyama K, Watanabe M, Takasaki T, Ojima K, Hinata K (1998) Dominance relationships between *S*-alleles in self-incompatible *Brassica campestris* L. *Heredity*, **80**, 241–247.

Hiscock SJ, McInnis SM (2003) Pollen recognition and rejection during the sporophytic self-incompatibility response: *Brassica* and beyond. *TRENDS in Plant Science*, **8**, 606–613.

Hudson RR, Kaplan NL (1988) The coalescent process in models with selection and recombination. *Genetics*, **120**, 831–840.

Ilic K, SanMiguel PJ, Bennetzen JL (2003) A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. *Proceedings of the National Academiy of Sciences of the USA*, **100**, 12265–12270.

Imrie BC, Kirkman CJ, Ross DR (1972) Computer simulations of a sporophytic self-incompatibility breeding system. *Australian Journal of Biological Science*, **25**, 343–349.

Ioerger TR, Clark AG, Kao T-H (1990) Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proceedings of the National Academy of Sciences of the USA*, **87**, 9732–9735.

Jain SK (1976) The evolution of inbreeding in plants. *Annual Review in Ecology and Systematics*, **7**, 469–495.

Karron JD, Marshall DL, Oliveras DM (1990) Numbers of sporophytic self-incompatibility alleles in populations of wild radish. *Theoretical and Applied Genetics*, **79**, 457–460.

Kato S, Mukai Y (2004) Allelic diversity of *S*-RNase at the self-incompatibility locus in a natural flowering cherry populations (*Prunus lannesiana* var. *speciosa*). *Heredity*, **92**, 249–256.

Kimura M (1965) Simulation studies on the number of self-sterility alleles maintained in a small population. *Annual Report of the National Institute of Genetics Japan*, **16**, 86–88.

Kowyama Y, Takahasi H, Muraoka K *et al.* (1994) Number, frequency and dominance relationships of *S*-alleles in diploid *Ipomoea trifida*. *Heredity*, **73**, 275–283.

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.

Kusaba M, Nishio T, Satta Y, Hinata K, Ockendon D (1997) 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. *Proceedings of the National Academy of Sciences of the USA*, **94**, 7673–7678.

Kusaba M, Dwyer K, Hendershot J *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.

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

Lawrence MJ (2000) Population genetics of the homomorphic self-incompatibility polymorphisms in flowering plants. *Annals of Botany*, **85**, 221–226.

Lawrence MJ, Franklin-Tong VE (1994) The population-genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. IX. Evidence of an extra effect of selection acting on the *S*-locus. *Heredity*, **72**, 353–364.

Les DH, Reinartz JA, Esselman EJ (1991) Genetic consequences of rarity in *Aster furcatus* (Asteraceae), a threatened, self-incompatible plant. *Evolution*, **45**, 1641–1650.

Levin DA (1993) *S*-gene polymorphism in *Phlox drummondii*. *Heredity*, **71**, 193–198.

Lu Y (2001) Roles of lineage sorting and phylogenetic relationship in the genetic diversity at the self-incompatibility locus. *Heredity*, **86**, 195–205.

Mable BK, Schierup MH, Charlesworth D (2003) Estimating the number, frequency, and dominance of *S*-alleles in a natural population of *Arabidopsis lyrata* (Brassicaceae) with sporophytic control of self-incompatibility. *Heredity*, **90**, 422–431.

Matton DP, Luu DT, Xike Q *et al.* (1999) Production of an S RNase with dual specificity suggests a novel hypothesis for the generation of new S alleles. *Plant Cell*, **11**, 2087–2098.

May G, Shaw F, Badrane H, Vekemans X (1999) The signature of balancing selection: fungal mating compatibility gene evolution. *Proceedings of the National Academy of Sciences of the USA*, **96**, 9172–9177.

Mayo O (1966) On the problem of self-incompatibility alleles. *Biometrics*, **22**, 111–120.

Mehlenbacher SA (1997) Revised dominance hierarchy for *S*-alleles in *Corylus avellana* L. *Theoretical and Applied Genetics*, **94**, 360–366.

Miege C, Ruffio-Châble V, Schierup MH *et al.* (2001) Intra-haplotype polymorphism at the *Brassica S* locus. *Genetics*, **159**, 811–822.

Moran PAP (1962) *The Statistical Processes of Evolutionary Theory*. Clarendon Press, Oxford.

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

Nagylaki T (1975) The deterministic behavior of self-incompatibility alleles. *Genetics*, **79**, 545–550.

Nagylaki T (1985) Homozygosity, effective number of alleles, and interdeme differentiation in subdivided populations. *Proceedings of the National Academy of Sciences of the USA*, **82**, 8611–8613.

Nasrallah JB, Kandasamy MK, Chang MC *et al.* (2000) Identifying genes for pollen–stigma recognition in crucifers. *Annals of Botany*, **85**, 125–132.

de Nettancourt D (2001) *Incompatibility and Incongruity in Wild and Cultivated Plants*. Springer-Verlag, Berlin.

Neuhauser C (1999) The ancestral graph and gene genealogy under frequency-dependent selection. *Theoretical Population Biology*, **56**, 203–214.

Nou IS, Watanabe M, Isuzugawa K, Isogai A, Hinata K (1993) Isolation of S-alleles from a wild population of *Brassica campestris* L. at Balcesme, Turkey and their characterization by S-glycoproteins. *Sexual Plant Reproduction*, **6**, 71–78.

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.

Peters AD, Hallignan DL, Whitlock MC, Keightley PD (2003) Dominance and overdominance of mildly deleterious induced mutations for fitness traits in *Caenorhabditis elegans*. *Genetics*, **165**, 589–599.

Raspé O, Kohn JR (2002) *S*-allele diversity in *Sorbus aucuparia* and *Crataegus monogyna* (Rosaceae: Maloideae). *Heredity*, **88**, 458–465.

Richards AJ (1986) *Plant Breeding Systems*. Allen & Unwin, London.

Richman A (2000) Evolution of balanced genetic polymorphism. *Molecular Ecology*, **9**, 1953–1963.

Richman AD, Kohn JR (1999) Self-incompatibility alleles from *Physalis*: implications for historical inference from balanced genetic polymorphisms. *Proceedings of the National Academy of Sciences of the USA*, **96**, 168–172.

Richman AD, Kohn JR (2000) Evolutionary genetics of self-incompatibility in the Solanaceae. *Plant Molecular Biology*, **42**, 169–179.

Richman AD, Schaeffer S, Uyenoyama MK (1995) *S*-allele sequence diversity in natural populations of *Solanum carolinense* (Horsenettle). *Heredity*, **75**, 405–415.

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

Richman AD, Uyenoyama MK, Kohn JR (1996b) *S*-allele diversity in a natural population of *Physalis crassifolia* (Solanaceae) (ground cherry) assessed by RT-PCR. *Heredity*, **76**, 497–505.

Rousset F (2004) *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press, Princeton.

Samaha RR, Boyle TH (1989) Self-incompatibility of *Zinnia augustifolia* HBK (Compositae). II. Genetics. *Journal of Heredity*, **80**, 368–372.

Sampson DR (1964) A one-locus self-incompatibility system in *Raphanus raphanistrum*. *Canadian Journal of Genetics Cytology*, **6**, 435–445.

Sampson DR (1967) Frequency and distribution of self-incompatibility alleles in *Raphanus raphanistrum*. *Genetics*, **56**, 241–251.

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. *Nature Genetics*, **20**, 43–45.

Sassa H, Hirano H, Ikehashi H (1992) Self-incompatibility-related RNAses in styles of Japanese pear (*Pyrus serotina* Rehd). *Plant Cell Physiology*, **33**, 811–814.

Schierup MH (1998) The number of self-incompatibility alleles in a finite, subdivided population. *Genetics*, **149**, 1153–1162.

Schierup MH, Vekemans X, Christiansen FB (1997) Evolutionary dynamics of sporophytic self-incompatibility alleles in plants. *Genetics*, **147**, 835–846.

Schierup MH, Vekemans X, Christiansen FB (1998) Allelic genealogies in sporophytic self-incompatibility systems in plants. *Genetics*, **150**, 1187–1198.

Schierup MH, Vekemans X, Charlesworth D (2000a) The effect of subdivision on variation at multi-allelic loci under balancing selection. *Genetical Research*, **76**, 51–62.

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

Schierup MH, Mable BK, Awadalla P, Charlesworth D (2001a) Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. *Genetics*, **158**, 387–399.

Schierup MH, Mikkelsen AM, Hein J (2001b) Recombination, balancing selection and phylogenies in MHC and self-incompatibility genes. *Genetics*, **159**, 1833–1844.

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

Sijacic P, Wang X, Skirpan AL, Wang Y, Dowd PE, McCubbin AG, Huang S, Kao T-H (2004) Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature*, **429**, 302–305.

Slatkin M, Charlesworth D (1978) The spatial distribution of transient alleles in a subdivided poppulation: a simulation study. *Genetics*, **89**, 793–810.

Stein J, Howlett B, Boyes DC, Nasrallah ME, Nasrallah JB (1991) Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brasscia olearacea*. *Proceedings of the National Academy of Sciences of the USA*, **88**, 8816–8820.

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

Steiner W, Gregorius H-R (1994) Single-locus gametophytic incompatibility: the symmetric equilibrium is globally asymptotically stable. *Journal of Mathematical Biology*, **32**, 515–520.

Stevens JP, Kay QON (1989) The number, dominance relationships and frequencies of self-incompatibility alleles in a natural population of *Sinapis arvensis* L. in South Wales. *Heredity*, **62**, 199–205.

Stone JL (2004) Sheltered load associated with *S*-alleles in *Solanum carolinense*. *Heredity*, Advanced online publication.

Strobeck C (1983) Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. *Genetics*, **103**, 545–555.

Takahata N (1990) A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proceedings of the National Academy of Sciences of the USA*, **87**, 2419–2423.

Takasaki T, Hatakeyama K, Suzuki G, Watanabe M, Isogai A, Hinata K (2000) The *S*-receptor kinase determines self-incompatibility in *Brassica* stigma. *Nature*, **403**, 913–916.

Takayama S, Shiba H, Iwano M *et al.* (2000) The pollen determinant of self-incompatibility in *Brassica campestris*. *Proceedings of the National Academy of Sciences of the USA*, **97**, 1920–1925.

Thien LB, White DA, Yatsu LY (1983) The reproductive biology of a relict-*Illicium floridanum* Ellis. *American Journal of Botany*, **70**, 719–727.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Uyenoyama MK (1988) On the evolution of genetic incompatibility systems. III. Introduction of weak gametophytic self-incompatibility under partial inbreeding. *Theoretical Population Biology*, **34**, 47–91.

Uyenoyama MK (1989) On the evolution of genetic incompatibility systems. V. Origin of sporophytic self-incompatibility in response to overdominance in viability. *Theoretical Population Biology*, **36**, 339–365.

Uyenoyama MK (1995) A generalized least-squares estimate for the origin of sporophytic self-incompatibility. *Genetics*, **139**, 975–992.

Uyenoyama MK (1997) Genealogical structure among alleles regulating self-incompatibility in natural populations of flowering plants. *Genetics*, **147**, 1389–1400.

Uyenoyama MK (2000a) A prospectus for new developments in the evolutionary theory of self-incompatibility. *Annals of Botany*, **85**, 247–252.

Uyenoyama MK (2000b) Evolutionary dynamics of dual-specificity self-incompatibility alleles. *Plant Cell*, **12**, 310–312.

Uyenoyama MK (2000c) Evolutionary dynamics of self-incompatibility alleles in *Brassica*. *Genetics*, **156**, 351–359.

Uyenoyama MK (2003) Genealogy-dependent variation in viability among self-incompatibility genotypes. *Theoretical Population Biology*, **63**, 281–293.

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

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

Vekemans X, Schierup MH, Christiansen FB (1998) Mate availability and fecundity selection in multi-allelic self-incompatibility systems in plants. *Evolution*, **52**, 19–29.

Vieira CP, Charlesworth D, Vieira J (2003) Evidence for rare recombination at the gametophytic self-incompatibility locus. *Heredity*, **91**, 262–267.

Wang Y, Wang X, Skirpan AL, Kao T-H (2003) *S*-RNase-mediated self-incompatibility. *Journal of Experimental Botany*, **54**, 115–122.

Wright S (1939) The distribution of self-sterility alleles in populations. *Genetics*, **24**, 538–552.

Wright S (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Wright SI, Lauga B, Charlesworth D (2003) Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. *Molecular Ecology*, **12**, 1247–1263.

Xue YB, Carpenter R, Dickinson HG, Coen ES (1996) Origin of allelic diversity in *Antirrhinum S* locus RNases. *Plant Cell*, **8**, 805–814.

Young AG, Brown AHD, Murray BG, Thrall PH, Miller C (2000) Genetic erosion, restricted mating and reduced viability in fragmented populations of the endangered grassland herb *Rutidosis leptorrhynchoides*. In: *Genetics, Demography and Viability of Fragmented Populations* (eds Young AG, Clarke GM), pp. 335–359. Cambridge University Press, Cambridge.

This is the inaugural paper of the research team 'Evolution of self-incompatibility systems' set up by X.V. in University of Lille to investigate population biology and molecular evolution of self-incompatibility systems in plants. VC is a postdoc interested in the evolution of genetic diversity in both plant and animal species. XV is a professor in plant population genetics and is also involved in investigations on plant population genetic structure.
