Historical events and allelic polymorphism at the gametophytic self-incompatibility locus in Solanaceae

Y Lu[1,2]

[1]*Department of Biology, Duke University, Durham, NC 27708, USA;* [2]*Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Xiangshan, Beijing 100093, China*

The historical migration rate of a species is often difficult to estimate with neutral markers, because the relationship between the turnover time of the markers and the age of the species commonly remains unknown. Compared with neutral markers, the plant self-incompatibility locus (S) provides a much better source of data for migration-rate estimation due to its high allelic polymorphism and antiquity. Here, the results from extensive surveys of S alleles in two wild solanaceous species, *Solanum carolinense* and *Physalis longifolia*, indicate that historical migration rates have differed significantly between the species; the higher migration rate found in *S. carolinense* appears to have interacted with the balancing selection at the S locus to result in fewer S alleles being maintained in the species. Historical population growth rates estimated via a modified coalescent approach also suggest a faster growing population for *S. carolinense* than for *P. longifolia*, which would have further widened their interspecific difference in S-allele polymorphism. These historical factors may have reduced the probability of new S alleles to prevailing in *S. carolinense*, leaving old ones segregating at the S locus with little signature of positive selection being currently detectable.
Heredity (2006) **96**, 22–28. doi:10.1038/sj.hdy.6800740; published online 28 September 2005

Introduction

Allelic diversification underlies many important biological phenomena such as disease resistance (eg, Jones, 2001; Feng *et al*, 2003) and immunological reactions (Rosenthal, 2001). It is conceivable that much of the current genetic diversity reflects the ecological circumstances the species has experienced. Understanding the ecological context in which allelic diversification takes place is thus fundamental to postgenomic analyses, although few examples are available so far (eg Wegner *et al*, 2003). Part of the difficulty comes from the lack of connection between genetic diversity and species ecology. Finding suitable systems for study is, therefore, a prerequisite for exploring evolutionary processes that run across different levels.

In flowering plants, the S locus of gametophytic self-incompatibility (GSI) offers an ideal system for examining ecological factors that may be responsible for different degrees of allelic polymorphism across species. The S locus is under strong balancing selection (Wright, 1960; Ewens, 1964; Yokoyama and Nei, 1979) and has typically been found to have multiple alleles segregating within species (eg, Emerson, 1939; Kato and Mukai, 2004), which can be interpreted using neutral theories of allelic polymorphism and their extensions. Species have

been found to differ in S allele number and also in allelic diversity (eg, Richman *et al*, 1995; Lu, 2001; Raspé and Kohn, 2002; Kato and Mukai, 2004). These observations have inspired hypotheses from both ecological and genetic perspectives including the involvement of population bottlenecks (Richman *et al*, 1996b), growth habit (Richman and Kohn, 1999), and lineage-dependent selection (Uyenoyama, 2003).

After the pioneering S-allele sequence surveys conducted on *Solanum carolinense* (Richman *et al*, 1995) and *Physalis crassifolia* (Richman *et al*, 1996a) of the Solanaceae, the bottleneck hypothesis was the first to be advanced to account for variation in S-polymorphism between the species. A bottleneck event was proposed in the lineage leading to *P. crassifolia*, because of the species' paucity of ancient genealogical branches and a smaller estimate of the long-term effective population size than that of *S. carolinense* (Richman *et al*, 1996b). Additional data indicated, however, that the same pattern of few ancient lineages was shared among *Physalis* congeners, which had a diverse range of S polymorphism (Richman and Kohn, 1999; Lu, 2001). Supposing that the bottleneck event was an ancient one, prior to the split of *Physalis* and *Witheringia* (Richman and Kohn, 2000; Igic *et al*, 2003), it would nevertheless be difficult to explain why the subsequent proliferation of S alleles occurred in some species but not in others. Furthermore, the estimate of the long-term effective population size might have been larger if the population's subdivision had been taken into account (Schierup *et al*, 2000).

In light of the newer data, a hypothesis emphasising the ephemeral and disturbed habitats and the growth

habits of weedy species was proposed to explain specific S polymorphism (Richman and Kohn, 1999). Although initially looking promising, this hypothesis showed little utility when examined in detail for three solanaceous species (Lu, 2001), as the null hypothesis was inconsistent with the patterns of S polymorphism observed across the species.

Strong frequency-dependent selection at the S locus may affect its linked loci (eg, Strobeck, 1980; Glémin et al, 2001), and permit accumulation of deleterious mutations near the S locus. This led to the hypothesis of lineage-dependent selection (Uyenoyama, 2003), which suggests that the longer that two alleles had coexisted before their split, the less likely they would be to form a zygote, owing to the selection at the fitness loci linked to the S locus. This hypothesis has been tested via crosses between genotyped parents (Stone, 2004) but awaits more data before a definitive conclusion is reached.

Concomitant with the empirical investigations, conceptual advances have also shed some new lights on species-specific polymorphism. One of them suggests that the interaction between the migration among subpopulations and balancing selection at the S locus may lead to a minimum allele number at an intermediate migration rate (Schierup, 1998; Muirhead, 2001). Albeit based on simulation work, this new result raises the likelihood that different migration rates of GSI species may interact with the balancing selection at the S locus leading to different degrees of allelic polymorphism in different species. The pattern of G_{st} at the S locus is expected to differ from that of allele number (Schierup et al, 2000); a parallel effect is seen at neutral loci where the genetic structure, measured by F_{st}, is hardly affected by the number of polymorphic alleles at the locus (see the data of Dhuyvetter et al, 2004). The dynamic interaction between selection and migration first noticed by Schierup (1998) may have captured an important facet of the process of allelic evolution at the S locus, and is referred to as interaction hypothesis hereafter. As migration rates may leave distinctive hallmarks in the distribution of allele frequencies among populations (Muirhead, 2001), the interaction hypothesis may be tested with S alleles randomly sampled from natural populations.

This study reports the analysis of two extensive surveys on S alleles sampled among natural populations of previously known GSI species (species showing GSI), *S. carolinense* and *Physalis longifolia*, of the Solanaceae. Based on both coding and noncoding sequences and allelic distributions of the samples, we find evidence that the natural populations of GSI species have indeed experienced their species-specific regimes of migration, which appears to have been mediated by species-specific pollination, seed dispersal, and interaction with local fauna.

Materials and methods

Population sampling
S. carolinense and *P. longifolia* are both weedy perennial herbs with extensive underground growth. They are frequently found at roadsides, in fields, and other disturbed habitats. The distribution of *S. carolinense* spans most of the United States, whereas *P. longifolia* is

Figure 1 The geographic distributions of samples collected from *S. carolinense* (■) and *P. longifolia* (●) in the United States.

found largely in the central and southern United States. A total of 13 populations of *S. carolinense* and 11 populations of *P. longifolia* were randomly sampled in 1998 and 1999 in the eastern and central United States (Figure 1, Table 1). Individual plants were sampled at an approximately 2 m interval within each population exhaustively. If two adjacent plants showed the same S genotype, only one of them was used in the data analysis to avoid double sampling of the same genet.

Sequencing S alleles
Healthy styles of individuals raised in the greenhouses were collected with clean forceps, placed in 1.5 ml vials, and stored in a liquid nitrogen canister. Extraction of total RNA and subsequent RT-PCR were performed as described previously (Lu, 2001, 2002). S introns were obtained from leaf gDNAs by PCR using allelic-specific primers. Newly obtained sequences in this study have been deposited in GenBank (accession numbers AY706443–AY706474).

Estimating species-level migration rate
Two methods were applied to the estimation of migration rates in *S. carolinense* and *P. longifolia*. The first one was based on the analytical model developed by Muirhead (2001), estimating the ratio of migration rate (m) and mutation rate (u) for each species from allelic distribution classes. If mutation rate at the S locus differs little between species, the ratio m/u would reflect the migration rate of each species. Since there is little evidence for species-specific mutation rates at the same

Table 1 Estimates of migration and total alleles from samples of two self-incompatible species

Species	Population	Sample size	Allele number
Physalis longifolia			
	A	9	9
	Am	9	12
	Bf	11	12
	E	14	14
	F	15	13
	K	12	8
	L	13	13
	N	9	7
	P	2	3
	V	4	6
	X	8	8
Observed total:	11	106	36
	Estimated m/u: 10.6; total alleles: 46		
Solanum carolinense			
	Ak	7	8
	Bf	8	10
	C	2	3
	E	13	12
	F	5	6
	H	9	9
	K	11	7
	Kn	12	12
	L	6	9
	Lm	10	12
	M	7	7
	Nj	12	11
	X	6	6
Observed total:	13	108	22
	Estimated m/u: 19.9; total alleles: 35		

locus, the estimation of m/u becomes equivalent to the estimation of m (rescaled by the mutation rate).

Using the notation of Muirhead (2001), we define $\hat{h}_k$ the expected fraction of alleles in a population that are shared among k populations at equilibrium, then the vector $\{\hat{h}_1, \hat{h}_2, ..., \hat{h}_n\}$ would describe the equilibrium condition of the species of n populations. These distributions may be compared between species using nonparametric methods such as Miller's test (Hollander and Wolfe, 1999). Assuming migration rate m and mutation rate to new allele u, the recurrent relation has been proposed to be

$$\hat{h}_k = \hat{h}_1 \left(\frac{m}{u}\right) \left(\frac{n-(k-1)}{n}\right) \hat{h}_{k-1}$$

To estimate m/u for each population, one may obtain

$$\frac{m}{u} = \frac{\hat{h}_k n}{\hat{h}_1 \hat{h}_{k-1}[n-(k-1)]}$$

A species-wide estimate of m/u is therefore obtainable from the averages of $\hat{h}_1$, $\hat{h}_{k-1}$, and $\hat{h}_k$ of the sampled populations.

As several assumptions in the Muirhead model (eg, equilibrium populations, symmetrical migration among populations, selective equivalence among alleles) might be violated in natural populations and cause biases of unknown degrees, a second method of estimating a species-wide migration rate was devised, and is presented here. This method simply employs the percentage of populations an S allele reaches within a species as a crude measure of migration rate (in arbitrary units). Let x_i be the observed number of populations where allele i is found, and n be sampled alleles among k populations, then $1/k \sum_{i=1}^{n} x_i/n$ estimates the percentage of populations an S allele reaches species wide. The rational for the estimate is straightforward. The dispersal of an S allele across natural populations is via migration. Irrespective of population sizes and interpopulation distances, a higher migration rate is expected to increase the dispersal of S alleles across populations, therefore leading to a broader dispersion within the species. This method assumes random dispersal of S alleles and equal frequency of alleles in the ancestral population. Using the consequent distribution of S alleles to estimate migration rate necessarily yields a long-term average. The percentage is qualitatively comparable to the migration rate obtained from the first method because both estimation methods rely on the same information – the S allelic distribution across populations.

Estimating total alleles
Using a model of overdominance to approximate the selection at the S locus, the total allele number in a species may be estimated by $\hat{i} n \sum_{k=1}^{n} \hat{h}_k/k$, where n is the number of populations surveyed and $\hat{i}$ the average number of alleles in a population (Muirhead, 2001).

Estimating historical population changes
According to the coalescence theory developed for neutral genes (Watterson, 1974; Kingman, 1982), for m randomly sampled genes in a population size of N, the time for two genes coalesce back to their most recent common ancestor is geometrically distributed with the mean of $2N$ generations. The genealogy of k alleles under balancing selection is expected to have the same topology as that of neutral genes but on a different time scale (Takahata, 1990). Let the scaling factor be f_s for the S-locus (Vekeman and Slatkin, 1994) and assume that it changes on a smaller scale than that of N, the expected time for all sampled S alleles (m) to coalesce to the most recent common ancestor becomes

$$E(T_m) \approx 4Nf_s \left(1 - \frac{1}{m}\right) \qquad (1)$$

Assuming that N has been large enough for both solanaceous species, we may regard N as a variable rather than a parameter. This is feasible since the geometrical distribution, on which the coalescent theory is based, holds true for each generation as long as N is far larger than m. It follows that the sum of geometrical distribution throughout generations should also apply for a large and variable N, therefore the Equation (1). Now, taking m as a variable as well, each $E\{Tk\}$, where $k = 2, ..., m-1$, may be approximated by the corresponding tree depth, a measure directly taken from the allelic genealogy of S alleles (Figure 2). The observed relationship between the values of $E\{T_k\}$ and their k's provides a guide for fitting the best model(s) depicting the changes of N. The fitted pattern of N is, in theory, reflective of historical changes of the population size at the species level, given that all the assumptions made in the coalescent theory are not seriously violated. The inferred
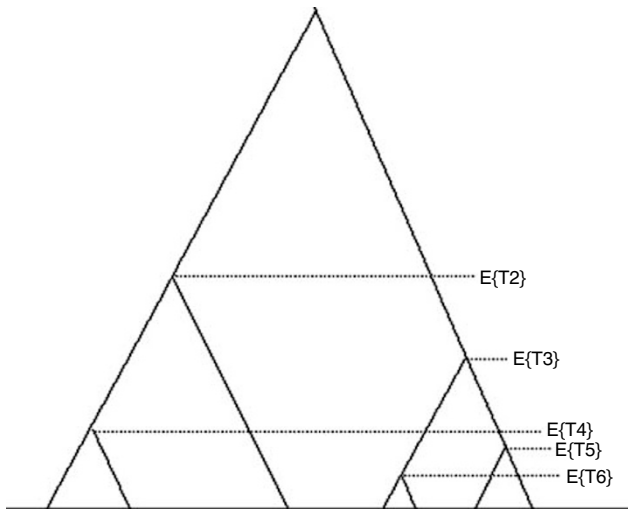
Figure 2 Illustration of the relationship between the sum of expected coalescent time ($E\{T_k\}$) and number of alleles (k) based on the coalescence theory.

Figure 3 Frequency distributions of allelic classes in *S. carolinense* (Sc) and *P. longifolia* (Pl) observed from sampled populations.

population size N is necessarily in arbitrary units but may be translated into a meaningful value when a reference population size (such as the current population size) becomes available.

The total data points collected from an allelic genealogy are the number of internal nodes i ($i = 2,\ldots,m-1$). After visual comparisons of the fitting between growth models and the data via graphic software, one may empirically obtain a model that best fits the observed data. If the data suggest that more than one model is needed, one may fit different growth models for segments of the genealogy. This approximate method is valid for subdivided GSI populations as well as panmictic ones because the growth pattern is inferred from the shape of allelic genealogy, which has been found to change little from a panmictic population to a subdivided one (Schierup *et al*, 2000).

Owing to the method's sensitivity to branch lengths, an allelic genealogy should be reconstructed with little distortion. Sequences of comparable region and length and a sound tree-building approach are essential for a reliable genealogy. Since S alleles are highly selected, the existing neutral models for nucleotide substitutions may not be appropriate. I chose amino-acid distances empirically estimated by the PAM matrix and the Neighbor-Joining algorithm implemented in Phylip (Felsenstein, 2000) for constructing the allelic genealogies of the two species concerned here.

Testing hypotheses

The large numbers of genotypes gathered from the two solanaceous species also allowed for direct tests of the hypothesis of lineage-dependent selection in this study. Invoking genealogy-dependent selection at the fitness loci linked to the S locus, the hypothesis argues that recently diverged S haplotypes, when forming a zygote, would have a higher level of homozygosity at the linked region because of their recently shared history of the region. Deleterious and recessive mutations thus exposed as homozygotes may reduce the viability of the zygotes, making suc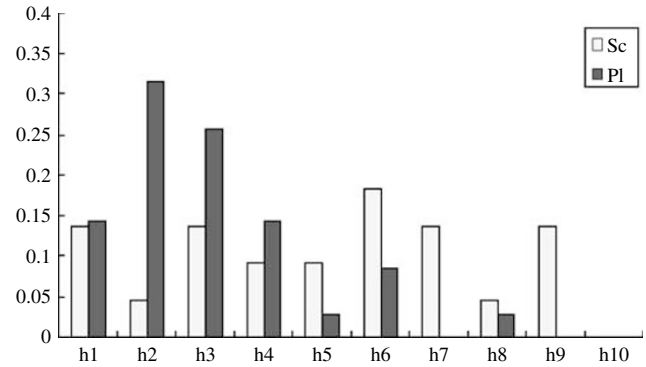h genotypes less frequent in natural populations. Under this hypothesis, one would observe a positive correlation between the viability of a given genotype and the divergence time between the two constituent alleles (Uyenoyama, 2003). A correlation coefficient (r) was therefore calculated between paralinear distance (estimated for two constituent alleles of a genotype; Lake, 1994) and the genotypic frequency in each species. The correlation was evaluated using a t-test. The power of the test, approximated assuming a normal distribution (z), was calculated as $1-p[z_{\alpha/2}-t_s]$, where $t_s = r\sqrt{(n-2)/(1-r^2)}$, α the type I error set at 0.05, and n is the sample size.

The parameters Ka (the number of nonsynonymous substitutions per nonsynonymous site) and Ks (the number of synonymous substitutions per synonymous site) were also computed for pairs of *S. carolinense* sequence using the method of Nei and Gobojori (1986) implemented in DnaSP3.5 (Rozas and Rozas, 1999), as nine more S alleles have been found since the last reported survey of the species by Richman *et al* (1995).

Results

S allele distribution and total allele number

The numbers of surveyed individuals and sites were comparable between *S. carolinense* and *P. longifolia* (Figure 1, Table 1), but significantly contrasting patterns of allelic distribution appear between species (Miller's test: statistic $Q = 26.5$, $P < 0.001$ for H_o:equal dispersion between species, and two-sided procedure). The observed allelic distribution provided a basis for qualitative and quantitative inferences on historical migration rate of a species. The qualitative inferences were drawn by comparing the observed allelic distribution to that of Muirhead's simulation (2001). For *P. longifolia*, classes of the observed allelic distribution clustered towards the left side of the spectrum, suggesting a low migration level among populations. Conversely, for *S. carolinense*, the distribution classes were much scattered, conforming to an intermediate level of migration (Figure 3). The observed number of total S alleles was lower in *S. carolinense* than in *P. longifolia*; so was the estimated total allele number (Table 1).

Migration rates between species

The quantitative estimation of m/u based on Muirhead's model suggests a two-fold higher average migration rate

in *S. carolinense* than in *P. longifolia* (Table 1). The second estimation method showed a similar pattern – the mean proportion of populations where an average *S* allele reached was 50% (SE 1.3%) in *S. carolinense*, and 35% (SE 1.6%) in *P. longifolia*; the difference in the migration rates was highly significant between the species (approximate *t*-test, $P < 0.001$). These quantitative results are in line with the qualitative estimates of the migration rates in *S. carolinense* and *P. longifolia*, respectively.

Historical population growth rates

The extension of the coalescent theory to the *S* locus suggests that a species' effective population size

$$N \approx \frac{mE(T_m)}{4f_s(m-1)}$$

for *m* *S* alleles. Applying the observed tree depth for *m* alleles to estimate $E\{T_m\}$ and taking f_s to be constant (or to have a smaller variance than that of N), the population's exponential growth rate was estimated from the branching pattern to be 1.2 for *S. carolinense* and 0.85 for *P. longifolia* (Figure 4).

Testing hypothesis of lineage-specific selection

No significant correlation was found between genotypic frequency and Lake's paralinear distance for allele pairs, either in *S. carolinense* ($n = 47$, $r = 0.07$, $P > 0.2$) or in *P. longifolia* ($n = 75$, $r = 0.02$, $P > 0.5$). The power of the test was approximately 0.93 in the case of *S. carolinense* and 0.96 in that of *P. longifolia*.

Comparison of selection intensities between species

In addition to the number of *S* alleles, the average *S* nucleotide diversity also differed between the species. The average of Lake's paralinear distances was 0.61 (SE 0.001) in *S. carolinense* and 0.53 (SE 0.001) in *P. longifolia*. Most introns of *S. carolinense* *S* alleles obtained in this study were not alignable owing to multiple indels and substitutions, particularly of genealogically distant ones. Using the seven pairs that were alignable (accession numbers AY706457; AY706443; AY706445; AY706446; AY706449; AY706454; AY706448), I estimated the selection intensity at the *S* locus for *S. carolinense*. The results indicated a negligible level of selection (−0.04, SE, 0.09; in comparison to 0.67, SE 0.055, reported previously for *P. longifolia* (Lu, 2002)). The remarkable gap between the two species in the level of positive selection was also found at the ratios of Ka/Ks; few *S. carolinense* allelic pairs have a ratio larger than one.

Discussion

The parallel sampling of *S* alleles in the two solanaceous species in this study has presented an interesting and powerful case for understanding how evolutionary forces shape the genetic polymorphism at the *S* locus. Through testing specific hypotheses and making connections between species ecology and genetic polymorphism, we now get a rare glimpse into the origins of its species-specific genetic diversity.

Testing hypotheses

The interaction hypothesis predicts a lower total allele number at the intermediate migration rate (eg, $m/u \sim 30$)
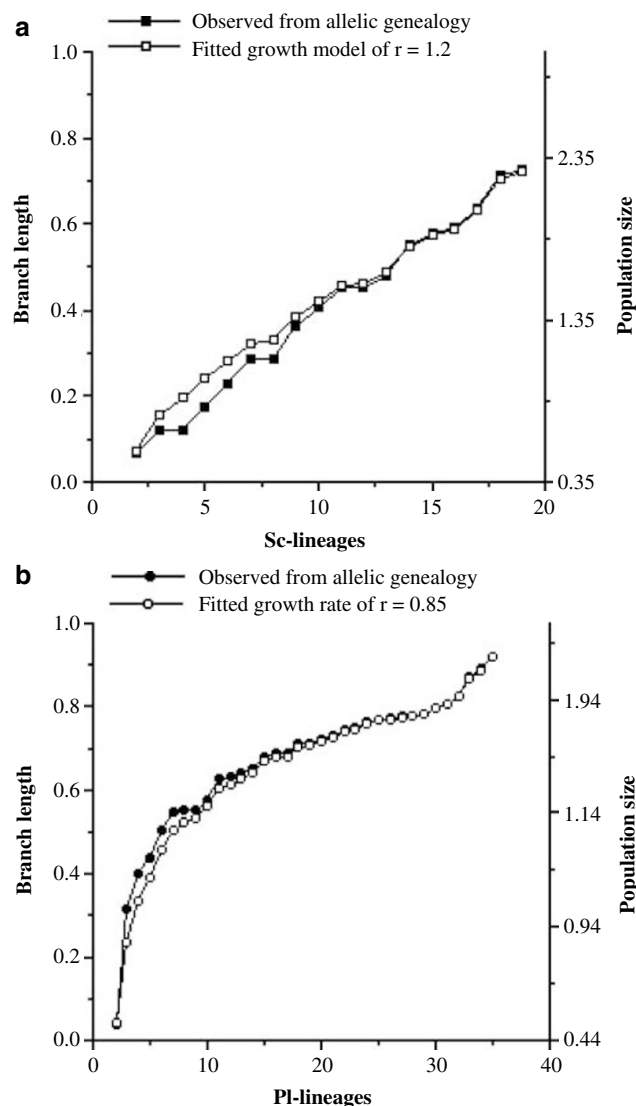

Figure 4 The estimations of the historical population size changes with 20 *S* alleles (accession numbers L40539-L4049; L40551; AY706463-706464; AY706466-706471) of *S. carolinense* (Sc) (**a**) and 36 *S* alleles (accession numbers AF281180-AF281201; AF374420-374430; AY706472-706474) of *P. longifolia* (Pl) (**b**); *r* is the exponential growth rate of the best-fit model. The units for population size are arbitrary.

and higher allele numbers at both high ($m/u > 1000$) and low ($m/u < 10$) migration rates. Two independent lines of support for the interaction hypothesis have been found. At the intermediate migration rate ($m/u \sim 20$) in *S. carolinense*, a smaller number of total alleles was observed than that in *P. longifolia*, which has a lower migration rate ($m/u \sim 11$). Since *S* allele distributions are the cumulative results of historical migration rates, the above estimates should reflect the historical patterns of migration in the two species respectively.

Ecological evidence corroborates these results. During our field work, we found that *S. carolinense* and *P. longifolia* inhabit nearly identical habitats. Frequent sharing of the same habitats would have potentially exposed these species to the same pollinators and fruit dispersers had they possess a similar reproductive syndrome. Interestingly, the major pollinators for

P. longifolia are solitary bees such as *Perdita halictoides* (Andrenidae) (Sullivan, 1986). These bees have limited patrol area as they burrow along at ground level. Large-bodied bumblebees, on the other hand, are the primary pollinators for *S. carolinense* (Connolly and Anderson, 2003). They typically move at longer distances, sometimes up to a dozen of kilometres (Heinrich, 1976). These long flights may contribute to interpopulation gene exchanges.

Following the pollination over the summer, the fruits of *S. carolinense* are relished by several species of birds as well as some small mammals (Martin *et al*, 1961; Cippolini and Levey, 1997). The bright yellow fruits of *S. carolinense* may stay on dry branches for months after frosts, making them easy targets for migrating birds in the later autumn. In comparison, fruits of *P. longifolia* are a much less important food source for birds than for small mammals (Martin *et al*, 1961), as they are generally less conspicuous to birds.

Seed dispersed by birds tends to be over a much longer distance than that of ground animals. Also, seeds passing through bird guts appear to be nearly 100% intact while those through the digestive systems of some mammals may be damaged to a great extent (Cippolini and Levey, 1997). The chances for an *S. carolinense* fruit to be carried outside its local population with intact seeds, therefore, appear to be much greater than those for a *P. longifolia* fruit. Consequently, both seed and pollen dispersal may well explain the higher migration rate estimated for *S. carolinense*. It is, then, likely that different interactions between levels of plant migration and balancing selection at the *S* locus have introduced different probabilities for new *S* specificities to be established at the population level, leading to difference in total allele numbers between the species.

No evidence has been found here, however, for the hypothesis of lineage-dependent selection. The data provided little evidence of the expected correlation between *S* genotypic frequency and genetic distance of two constituent alleles. Although more survey data are needed before the results of tests are beyond statistical doubt, the evidence leans towards species-specific rather than the systematic forces explaining the differences in *S* polymorphism. For systematic forces, such as balancing selection and lineage-dependent selection, it is unclear why the effects should manifest in one species but not in the other.

Historical population growth rates

The historical population size could affect the establishment of new alleles. Kimura (1962) showed that the probability of a selected mutation becoming fixed depends on population size, and the same pattern essentially holds true for selected mutations in subdivided populations (Takahata, 2001; Cherry and Wakeley, 2003). To evaluate this possibility for the *S* locus, the history of metapopulation size was estimated for the two species via the modified coalescent approach described above. The estimate of historical population growth rate is higher in *S. carolinense* than in *P. longifolia* (Figure 4). Although the difference between the species growth rates might be smaller, if the stronger positive selection in *P. longifolia* were taken into account, the current pan American distribution of *S. carolinense* and a much smaller species range for *P. longifolia* still argue for

a larger historical population size for *S. carolinense*, as was previously suggested by a different method (Richman *et al*, 1996b). A large metapopulation would have further diminished the chance of retaining new *S* alleles.

Contrasting patterns of molecular selection between species

The smaller number of *S* alleles would make the balancing selection at the *S* locus stronger (Lawrence, 2000), however, one does not observe a significant level of positive selection at the *S* locus in *S. carolinense*. This phenomenon reflects the antiquity of the *S* alleles in *S. carolinense*, rather than a lack of influence from positive selection. As positive selection is most pronouncing during the early diversification of alleles as seen in *P. longifolia* (Lu, 2002), lower recent allelic diversification would naturally be associated with a reduced signature of positive selection. There are at least two ramifications of this observation. One is that stronger balancing selection alone does not necessarily help to maintain more alleles in a species. The other is that balancing selection is not equivalent to positive selection. When positive selection is absent, sequence diversification of *S* alleles is mostly either neutral or under the influence of negative selection, as seen in *S. carolinense* where few allelic comparisons show their Ka/Ks > 1.

The different seed dispersal mechanisms and population histories match well the corresponding levels of *S* polymorphism in *S. carolinense* and *P. longifolia*. Linking ecological parameters to genetic polymorphism at the *S* locus gives us at least two insights. First, balancing selection, rather than acting alone, may interact with other evolutionary forces in natural populations to shape genetic polymorphism at the *S* locus in each species. Although the detail of the possible interactions is somewhat vague, even after the recent modeling effort (Muirhead, 2001), we know that interactions among migration rate, selection intensity, and population size may jointly alter the probability of fixing a new allele, as shown lately for beneficial alleles (Whitlock, 2003). Second, it is possible now to select species of dissimilar ecological attributes that may influence migration and population size, and to examine the corresponding levels of *S* polymorphism in order to identify likely players affecting species-specific genetic polymorphism. Even approximate estimation of ecological parameters (such as the historical change pattern in population size) may provide important insights into the genetic diversity bewildering us today.

Acknowledgements

References

Cherry JL, Wakeley J (2003). A diffusion approximation for selection and drift in a subdivided population. *Genetics* **163**: 421–428.

Cippolini ML, Levey DJ (1997). Why are some fruits toxic? Glycoalkaloids in *Solanum* and fruit choice by vertebrates. *Ecology* **78**: 782–798.

Connolly BA, Anderson GJ (2003). Functional significance of the androecium in staminate and hermaphrodite flowers of *Solanum carolinense* (Solanaceae). *Plant Syst Evol* **240**: 235–243.

Dhuyvetter H, Gaublomme E, Desender K (2004). Genetic differentiation and local adaptation in the salt-marsh beetle *Pogonus chalceus*: a comparison between allozyme and micro-satellite loci. *Mol Ecol* **13**: 1065–1074.

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

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

Felsenstein J (2000). PHYLIP 3.573 (Phylogeny inference package). *Department of Genetics*. University of Washington: Seattle.

Feng X, Carlton JM, Joy DA, Mu J, Furuya T, Suh BB *et al* (2003). Single-nucleotide polymorphisms and genome diversity in *Plasmodium vivax*. *Proc Natl Acad Sci USA* **100**: 8502–8507.

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

Heinrich B (1976). The foraging specializations of individual bumblebees. *Ecol Monogr* **46**: 105–128.

Hollander M, Wolfe DA (1999). *Nonparametric Statistical Methods*. John Wiley & Sons: New York. pp 158–161.

Igic B, Bohs L, Kohn JR (2003). Historical inferences from the self-incompatibility locus. *New Phytologist* **161**: 97–105.

Jones JD (2001). Putting knowledge of plant resistance genes to work. *Curr Opin Plant Biol* **4**: 281–287.

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

Kimura M (1962). On the probability of fixation of mutant genes in a population. *Genetics* **47**: 713–719.

Kingman JFC (1982). On the genealogy of large populations. *J Appl Prob* **19A**: 27–43.

Lake JA (1994). Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. *Proc Natl Acad Sci USA* **91**: 1455–1459.

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

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

Lu Y (2002). Molecular evolution at the self-incompatibility locus of *Physalis longifolia* (Solanaceae). *J Mol Evol* **54**: 784–793.

Martin AC, Zim HS, Nelson AL (1961). *American Wildlife & Plants*. Dover: New York, pp 414–415.

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

Nei M, Gobojori T (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. *Mol Biol Evol* **3**: 418–426.

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

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

Richman AD, Kohn JR (1999). Self-incompatibility alleles from *Physalis*: implications for historical inference from balanced genetic polymorphisms. *Proc Natl Acad Sci USA* **96**: 168–172.

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

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

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

Rosenthal BM (2001). Defining and interpreting intraspecific molecular variation. *Vet Parasitol* **101**: 187–200.

Rozas J, Rozas R (1999). DnaSP version3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.

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

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

Stone JL (2004). Sheltered load associated with S-alleles in *Solanum carolinense*. *Heredity* **92**: 335–342.

Strobeck C (1980). Heterozygosity of a neutral locus linked to a self-incompatibility locus or a balanced lethal. *Evolution* **34**: 779–788.

Sullivan JR (1986). Reproductive biology of *Physalis viscosa*. In: D'Arcy WG (ed) *Solanaceae: Biology and Systematics*. Columbia Univ. Press: New York, pp 274–283.

Takahata N (1990). A simple genealogical structure of strongly balanced allelic lines and *trans*-species evolution of polymorphism. *Proc Natl Acad Sci USA* **87**: 2419–2423.

Takahata N (2001). Genealogy of neutral genes and spreading of selected mutations in a geographically structured popula-tion. *Genetics* **129**: 585–595.

Uyenoyama MK (2003). Genealogy-dependent variation in variability among self-incompatibility genotypes. *Theor Popul Biol* **63**: 281–293.

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

Watterson GA (1974). The sampling theory of selectively neutral alleles. *Adv Appl Prob* **6**: 463–488.

Wegner KM, Kalbe M, Kurtz J, Reusch TBH, Milinski M (2003). Parasite selection for immunogenetic optimality. *Science* **301**: 1343.

Whitlock MC (2003). Fixation probability and time in sub-divided populations. *Genetics* **164**: 767–779.

Wright S (1960). On the number of self-incompatibility alleles maintained in equilibrium by a given mutation rate in a population of given size: a reexamination. *Biometrics* **16**: 61–85.

Yokoyama S, Nei M (1979). Population dynamics of sex-determining alleles in honey bees and self-incompatibility alleles in plants. *Genetics* **91**: 609–626.