

S-allele diversity in a natural population of *Physalis crassifolia* (Solanaceae) (ground cherry) assessed by RT-PCR

ADAM D. RICHMAN*, MARCY K. UYENOYAMA† & JOSHUA R. KOHN

Department of Biology 0116, University of California at San Diego, La Jolla, CA 92093–0116 and †Department of Zoology, Box 90325, Duke University, Durham, NC 27708–0325, U.S.A.

Allelic diversity at the self-incompatibility (*S*-) locus in the ground cherry, *Physalis crassifolia* (Solanaceae), was surveyed in a natural population occurring in Deep Canyon, CA, using a molecular assay to determine the genotype of individual plants. A total of 28 different *S*-alleles were identified and sequenced from a sample of 22 plants. All plants examined were heterozygous, as expected under gametophytic self-incompatibility (GSI). The estimated number of alleles in this population is 43–44, comparable to allelic diversity reported for other species, as determined by the standard diallel crossing method. Allele frequencies in the sample deviated from the expectation of equal frequency under GSI; it is suggested that this deviation may result from sampling of related individuals. Molecular analysis of genotypes within single pollen donor families indicates that, for all alleles examined, segregation is consistent with predictions for single-locus GSI. The implications of a reliable and efficient molecular assay for determining the *S*-genotype of plants are discussed.

Keywords: frequency-dependent selection, *Physalis crassifolia*, RT-PCR, *S*-locus, self-incompatibility.

Introduction

S-locus polymorphism has long attracted the interest of evolutionary biologists as an example of extreme diversifying selection. The conceptual basis for polymorphism in gametophytic self-incompatibility (GSI) systems is particularly simple because dominance in the determination of pollen specificity, typical of sporophytic systems, is absent. Single pollen donor crosses fall into one of three categories: compatible, half-compatible, or incompatible, corresponding to zero, one or two *S*-alleles shared by the parents. The opportunity for mating associated with any allele is inversely related to its frequency, promoting the maintenance of many alleles within populations (Emerson, 1939; Ockendon, 1974; Campbell & Lawrence, 1981; Levin, 1993).

S-allele diversity in population samples has been investigated with diallel crosses to determine compatibility, as assessed by examination of pollen tube growth in the style (Emerson, 1938; Lawrence,

1975) or seed-set (Levin, 1993). Richman *et al.* (1995) reported the first direct determination of the *S*-locus genotypes of plants sampled from natural populations using the polymerase chain reaction (PCR) procedure. This approach circumvents the necessity of large crossing experiments and provides sequence information on different alleles segregating in a population. Other advantages of the direct assay include unambiguous comparisons of studies of *S*-diversity carried out at different times and in different species, and the ability to study plants not amenable to greenhouse crossing experiments, such as tree species or plants producing limited numbers of flowers.

Here we report on *S*-allele diversity in a natural population of another GSI species, the ground cherry *Physalis crassifolia* (Solanaceae), as determined by reverse transcription (RT)-PCR and identification of different alleles by RFLP analysis and DNA sequencing. The pattern of inheritance of a subset of these sequences was investigated by carrying out compatible and half-compatible single-donor crosses. In addition, the predictions of the molecular assay were compared to the results of diallel crosses

*Correspondence.

among progeny arrays from each kind of cross, evaluated by fruit-set. The discovery rate of new *S*-alleles in the sample was used to estimate population *S*-allele diversity, and the sample distribution of allelic frequencies compared to that expected for a sample from a finite population.

Methods

Collection of plant materials

Ground cherry is a perennial subshrub of south-west North American deserts, occurring in rocky areas at between 300 and 800 m elevation, and blooming from November to August. At Deep Canyon Desert Reserve in Palm Desert, CA, widely scattered individuals occur in rocky areas embedded in alluvial fan habitat at 400 m elevation. An area 200 m by 2.5 km was carefully searched, and 54 individuals were located and tagged. Styles for molecular analyses were collected in May 1993 and 1994 and immediately frozen in liquid nitrogen. Open pollinated fruits were collected from these plants for propagation in the greenhouse.

RT-PCR

Styles collected one to two days prior to anthesis were either processed immediately or frozen at -80°C . Stylar RNA was extracted using the protocol described by Richman *et al.* (1995). Subsequent cDNA synthesis was performed using the cDNA Cycle[®] kit (Invitrogen Corp., San Diego, CA) and oligo dT primer. A 2 μL aliquot of the cDNA synthesis reaction was used directly in a PCR reaction without further purification, using degenerate *S*-locus-specific primers. Three primers were used: one upstream primer (PR1), targeted to the conserved region designated as C2 by Ioerger *et al.* (1991), in combination with one of two downstream primers (PR2 and PR3), corresponding to the conserved regions C3 and C4. Primer sequences and PCR amplification conditions were as described by Richman *et al.* (1995).

PCR amplification and RFLP analysis

RT-PCR from stylar RNA from individual plants resulted in a single band on an agarose gel of the size expected from published *S*-allele cDNA sequences. In most cases, restriction enzyme analysis of the products indicated the presence of two abundant sequences. In some instances, using the primer combination PR1 and PR3, it appeared that one product was amplified in greater amounts than

another, as determined by RFLP analyses. In these instances, a second set of primers (PR1, PR2) was used to obtain both products in approximately equal concentrations in order to obtain clones of both alleles.

Cloning and sequencing

PCR products were cloned, without purification, using the Invitrogen TA[®] cloning kit, in order to separate amplification products prior to sequencing. Cloned sequences were amplified from individual bacterial colonies, as described by Richman *et al.* (1995). Clones were sequenced using the BRL Cycle Sequencing[®] kit (Life Technologies, Gaithersburg, MD) and P³³ end-labelled m13 universal primers. Sequencing gels were immersed in 5 per cent acetic acid for 10 min, vacuum dried on Whatman 3 MM paper and exposed to X-ray film for 1–2 days.

Analysis of compatibility of single-donor crosses

Single-donor families were raised in the greenhouse to determine the pattern of inheritance of PCR markers and their association with the physiological expression of incompatibility. Individuals grown from open-pollinated fruits were crossed, the progeny raised to flowering, and the *S*-locus genotypes of the offspring determined by RT-PCR and RFLP analyses. Two different kinds of families were examined, fully compatible and half-compatible crosses. Compatible crosses were made between unrelated plants with easily distinguishable RFLP genotypes. Half-compatible crosses were made among full- or half-sibs, because of the rarity of this kind of mating among unrelated plants.

Results

S-locus genotypes and sequences

In all, 28 different sequences were found (GenBank accession numbers L46653–L46680) among the 22 plants sampled from Deep Canyon, each of which had a unique, heterozygous *S*-genotype (Table 1). In all cases the sequences obtained were consistent with the observed RFLP patterns. Inferred amino acid sequences for the 28 ground cherry sequences, along with 10 *S*-allele cDNAs from other taxa, and a homologous *S*-related RNase, are shown in Fig. 1.

Number of alleles in the population

The number of sequences recovered in a finite sample underestimates the total diversity within the

Table 1 *S*-genotypes of 22 plants of *Physalis crassifolia* sampled from Deep Canyon, CA

Ref no.	<i>S</i> -genotype
DC1.73	$S_{15}S_{13}$
DC1.9	S_3S_{20}
DC2.71	S_4S_5
DC2.13	$S_{28}S_9$
DC2	$S_{16}S_{19}$
DC6	S_2S_{12}
DC8	S_6S_{22}
DC9	$S_{11}S_{13}$
DC10	$S_{14}S_{20}$
DC16	$S_{11}S_{22}$
DC17	$S_{15}S_{22}$
DC19	$S_{10}S_{27}$
DC21	$S_{23}S_{24}$
DC23	S_9S_{12}
DC24	$S_{12}S_{25}$
DC28	S_6S_{11}
DC29	$S_{16}S_{17}$
DC36	S_8S_{20}
DC44	S_1S_{11}
DC46	S_9S_7
DC51	$S_{18}S_{26}$
DC53	$S_{21}S_{11}$

population. An estimate of the total number of alleles in the population (N) is obtained by implicit solution of $n = N[1 - (1 - 2/N)^r]$, where n is the number of alleles observed and r the number of plants examined (Paxman, 1963; O'Donnell & Lawrence, 1984). Using $r = 22$ and $n = 28$, the estimated number of alleles in the Deep Canyon population is 43–44.

The above estimator assumes that all genotypes are present at equal frequency, as expected under frequency-dependent selection at a gametophytic *S*-locus. This assumption was examined using Mantel's (1974) statistic

$$\chi_{n-1}^2 = (n-1)(\sum C_j^2 - 4r^2/n)/(2r - 4r/n)$$

where C_j is the number of times an allele occurs, n the number of alleles found, and r the number of plants sampled. The small sample size relative to the number of alleles detected suggested that the assumption of normality, on which the χ^2 statistic depends, might not hold. Therefore, the significance level of the test statistic was evaluated by Monte Carlo simulation. One thousand random samples of size $r = 22$ were drawn with replacement from a pool of $N = 44$ alleles at equal frequency, and Mantel's χ^2 calculated for each sample (Fig. 2). The value

obtained for the ground cherry population is marginally significant ($\chi^2 = 17.5$, $P = 0.046$).

If allele frequencies are unequal, then Paxman's (1963) maximum likelihood (ML) method will tend to underestimate the true number of alleles in the population, because of failure to sample rare alleles. O'Donnell & Lawrence (1984) proposed alternative estimators which are less biased than the ML method when frequencies are significantly uneven. In the present case, because of the marginal significance of the deviation in ground cherry, these equations provide the same estimated total number of alleles (43–44).

For comparison with other studies, we summarized the thoroughness of sampling of *S*-allelic diversity using a repeatability statistic R (Campbell & Lawrence, 1981) calculated as $R = (m - n)/(m - 3)$ where m is the number of alleles examined (equal to $2r$) and n is the number of different alleles found. If n equals m , no allele has been sampled more than once, and $R = 0$. R converges on 1 as n converges on N . If only three different alleles are sampled (the minimum required within a population if it is to persist), then $R = 1$ regardless of the number of alleles examined. For the ground cherry data, $R = 0.39$.

Compatible crosses

Under GSI, a cross between individuals which share no *S*-alleles in common is expected to generate four different *S*-genotypes among the resulting progeny. Progeny sets from two such compatible crosses were examined by determining the RFLP *S*-genotypes of the parents and offspring and the pattern of fruit-set resulting in crosses among the progeny (Tables 2 and 3). No attempt was made to assign the alleles to sequences identified in Fig. 1, although it was determined by RFLP analysis that the two compatible families shared no alleles in common, so that the segregations of a total of eight alleles were examined, designated $S_a - S_h$. In both families, progeny could be classified by their RFLP patterns as having received one or the other maternal and paternal sequences, consistent with the expectation that the RFLP patterns identified in the parents are allelic. In Family 1, all four possible genotypes were observed among the six offspring. In Family 2, three different genotypes among the six offspring were observed, with the fourth expected genotype missing, presumably because of the small size of the progeny array.

The physiological expression of self-incompatibility within these families was investigated by fruit-set.

LE
S2Mdo WPM--NNDGTYVSNCDPNSFYDQSIQD---LISMQNQWPTLACP---SGSGSTFWSHEWEKHGTCABSLTNOHAYFKKALDLKQO---IDLLLSILQAGADIHPDGE--SVDLVNIRNAIKSAI--GYTPWIOCNVDQSGNSOL
S3Mdo WPS--NMRSELEFNCS--NVTYAKIQN---IRTOLEMIWENFVN---RNLHGFNWRNEMKNGHAGCGYPTIRNDLHYFQTVIEMYTQKQNSVDISLAKAIEPDG--KFRTRQKEIVDAIRGHIHGKEPNLKCQKNT--QITEL
TOMS6 WPS--NWGSDPKKCTT--ILNPOUTIN---LTAQLELIIWENVLN---RKAHARFWRQWRKHGTCGPTIADDMHYFSTVIEMYTQKQNSVDISLAKAIPEK--KFRTRQKEIVDAIRGHIHGKEPNLKCQKNT--QITEL
CHACOS2 WPD--NRKYLANNRCSYA---YNALJTN--FRQSLDORWDLTNSKMTMKEKQGTWEYKNGKCCCEKLVNQO--QVFNLIAMKIKDK--FDLLGLTRKRHGINPFGS--TYELNDRERAIKTVSIEV--PSLKCIIRKPGQNVEL
S1NIC WPD--NVTSELANVCDROK--KFKLIFED--DKKQNDLDRWPDLDTRDCKNGGQWFSYKXKHGTCCLPYNQO--QVFNLIAMKIKDK--FDLLGLTRKRHGINPFGS--TYELNDRERAIKTVSIEV--PSLKCIIRKPGQNVEL
S1PET WPS--ITGFRLEFCFTGD--KYETFKD--NNIVDYLDRHWQMKFDENYAKHQPLWSYKXKHGTCCLPYNQO--QVFNLIAMKIKDK--FDLLGLTRKRHGINPFGS--TYELNDRERAIKTVSIEV--PSLKCIIRKPGQNVEL
S2PET WPK--NKHFRLEFCFTGD--KYSRFEK--DNIINVLERHWIQMRFDEKASTQPLWEHYNRHHGICCKNLYDQE--AYFLAIARLKER--FDLLGLTRKRHGINPFGS--TYELNDRERAIKTVSIEV--PSLKCIIRKPGQNVEL
S1SC WPD--KNGVILLHDCPFVKKDRYFTLTD--HKKLALDALKRFPQLKDFSGINQDMLRHFQKGGSCGKRYKQ--AYFLAIARLKER--FDLLGLTRKRHGINPFGS--TYELNDRERAIKTVSIEV--PSLKCIIRKPGQNVEL
Physal-5 WPD--KNGVILLHDCPFVKKDRYFTLTD--HKKLALDALKRFPQLKDFSGINQDMLRHFQKGGSCGKRYKQ--AYFLAIARLKER--FDLLGLTRKRHGINPFGS--TYELNDRERAIKTVSIEV--PSLKCIIRKPGQNVEL
Physal-27 WPD--KNGVILLHDCPFVKKDRYFTLTD--HKKLALDALKRFPQLKDFSGINQDMLRHFQKGGSCGKRYKQ--AYFLAIARLKER--FDLLGLTRKRHGINPFGS--TYELNDRERAIKTVSIEV--PSLKCIIRKPGQNVEL
Physal-1 WPD--KNGVILLHDCPFVKKDRYFTLTD--HKKLALDALKRFPQLKDFSGINQDMLRHFQKGGSCGKRYKQ--AYFLAIARLKER--FDLLGLTRKRHGINPFGS--TYELNDRERAIKTVSIEV--PSLKCIIRKPGQNVEL
Physal-12 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-15 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-16 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-1 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-4 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-26 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-18 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-11 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-17 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-1 Physal-1 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-13 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-3 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-24 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-6 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-25 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-7 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-19 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-2 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-9 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-10 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-8 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-14 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-21 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-20 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-22 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI
Physal-23 WPD--KAGVTLMDCCSPTP--NVTNFR--GKMLDLDLDTHTQTLTLLTKGTGLAEGRIWNYQYI

Fig. 1 Inferred amino acid sequences for 28 cDNA sequences isolated from *Physalis crassifolia* by RT-PCR, seven published *S*-allele cDNAs, and an *S*-locus related RNase (*LE*). In ground cherry, shorter sequences were obtained using primer pairs PR1 and PR2, longer sequences using PR1 and PR3. Sequences were aligned using CLUSTAL v 1.5 (Thompson *et al.*, 1994). Published sequences were derived from *Mahus x domestica* Borkh. (S2Mdo, S3Mdo: Broothaerts *et al.*, 1995), *Lycopersicon peruvianum* (TOMS6: Royo *et al.*, 1994), *L. esculentum* (LE: Jost *et al.*, 1991), *Nicotiana glauca* (S1NIC: Khey-Pour *et al.*, 1990), *Petunia inflata* (S1PET, S2PET: Ai *et al.*, 1990), *Solanum carolinense* (S1SC: Richman *et al.*, 1995), *S. chacoense* (CHACOS2: Xu *et al.*, 1990). Amino acids shown in bold typeface indicate primer sequence included for alignment purposes only.

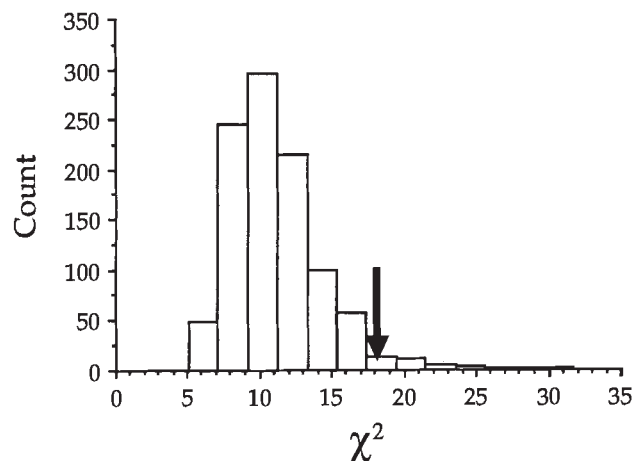


Fig. 2 Sampling distribution of Mantel's (1974) χ^2 statistic, generated by Monte Carlo simulation, for 1000 samples of 22 plants, from a population containing 44 equally frequent *S*-alleles.

Only crosses between individuals of identical RFLP genotype fail to set fruit. Individuals with *S*-genotypes S_aS_c and S_bS_d occurred twice within Family 1, and in both cases crosses within genotypes were mutually incompatible. Three individuals of genotype S_eS_h occurred in Family 2, and these were also mutually incompatible. All other crosses within these families set fruit as predicted.

Half-compatible crosses

In a half-compatible cross, only two *S*-genotypes are expected among the progeny. All progeny should carry the compatible paternal allele, and one or the other maternal allele. Three such families were investigated, and the results are summarized in Tables 4 and 5. Family 3 contains a total of three alleles, designated S_i – S_k , which on the basis of RFLP analyses are different from the eight previous alleles examined in compatible crosses. All eight progeny carry the compatible paternal allele S_k , as expected under GSI. Progeny 2 and 8 carry the maternal allele S_j and the remaining six plants the alternative maternal allele S_i . Fruit-set within the family is consistent with the assignments of *S*-genotype; crosses between different *S*-genotypes set fruit, whereas crosses between identical *S*-genotypes fail. Families 4 and 5 are reciprocal crosses between the female parent of Family 2, DC24.4 with genotype S_gS_h , and sibling DC24.3 with genotype S_iS_j . Assignments of *S*-genotypes and fruit-set within these families are shown in Table 5. All progeny within Family 4, in which DC24.3 is the pollen donor, share

Table 2 Fruit-set in Family 1 of *Physalis crassifolia*, a cross between DC8.1 (female) \times DC6.2

Ovule parent		Pollen parent					
		1 S_aS_c	3 S_aS_c	2 S_bS_d	4 S_bS_d	5 S_aS_d	10 S_bS_c
1	S_aS_c		0	–	+	+	+
3	S_aS_c	0		+	+	+	+
2	S_bS_d	–	+		0	+	+
4	S_bS_d	+	+	0		+	+
5	S_aS_d	+	+	+	+		+
10	S_bS_c	+	+	+	+	+	

– Indicates results from cross not available.

Table 3 Fruit-set in Family 2 of *Physalis crassifolia*, a cross between DC24.4 (female) \times DC24.1

Ovule parent		Pollen parent				
		1 S_eS_g	3 S_fS_h	4 S_eS_h	5 S_eS_h	6 S_eS_h
1	S_eS_g		+	+	+	+
3	S_fS_h	+		+	+	+
4	S_eS_h	+	+		0	0
5	S_eS_h	–	–	–		0
6	S_eS_h	+	+	0	0	

– Indicates results from cross not available.

Table 4 Fruit-set in Family 3 of *Physalis crassifolia*, a cross between DC42.1 (female) \times DC42.2

Ovule parent		Pollen parent						
		1 S_iS_k	4 S_iS_k	5 S_iS_k	6 S_iS_k	7 S_iS_k	2 S_jS_k	8 S_jS_k
1	S_iS_k		0	0	0	0	+	+
4	S_iS_k	0		0	0	0	+	+
5	S_iS_k	0	0		0	0	+	+
6	S_iS_k	0	0	0		0	+	+
7	S_iS_k	0	0	0	0		+	+
2	S_jS_k	+	+	–	+	+		0
8	S_jS_k	+	+	+	–	+	0	

– Indicates results from cross not available.

the compatible allele S_i , and all progeny in Family 5, in which DC24.4 is the pollen donor, share the compatible allele S_g . Data on fruit-set within families are again consistent with the molecular analyses, with individuals of identical *S*-genotype mutually

Table 5 Fruit-set in Families 4 and 5 of *Physalis crassifolia*, crosses between DC24.4 (female) × DC24.3 (1–4) and DC24.3 (female) × DC24.4 (R1–3)

Ovule parent	Pollen parent						
	1 S_gS_l	2 S_gS_l	3 S_hS_l	4 S_hS_l	R1 S_gS_h	R2 S_gS_h	R3 S_gS_l
1	S_gS_l	0	+	+	–	–	–
2	S_gS_l	0	+	+	–	–	–
3	S_hS_l	+	+	0	–	–	–
4	S_hS_l	+	+	0	–	–	–
R1	S_gS_h	–	–	–	–	0	+
R2	S_gS_h	–	–	–	0	–	+
R3	S_gS_l	–	–	–	+	+	–

– Indicates results from cross not available.

incompatible, whereas all half-compatible crosses between individuals of two different genotypes set fruit.

Data on fruit-set and compatibility within sibships bear on the issue of whether compatibility is determined by a single locus. Pandey (1957) reported that in *Physalis ixocarpa* some crosses were incompatible in one direction but not the other, a result not consistent with single-locus GSI. He postulated a second multiallelic locus with epistatic effects. We found no evidence for unidirectional failure of reciprocal crosses in any family examined in this study. The families studied here are not large, and the effects Pandey postulates are difficult to detect if there are many alleles segregating at the second locus.

Discussion

Specificity of the PCR assay

RT-PCR using stylar RNA as the template for cDNA synthesis resulted in amplification of a single band of expected size in each individual. Each band was subsequently found to contain two different sequences, as determined by RFLP analyses and DNA sequencing, consistent with the expectation that all individuals are heterozygous under GSI. In single-donor crosses between compatible plants involving eight different parental sequences, the segregation of the molecular sequences among the progeny was consistent with the interpretation that these sequences were alleles at a single locus. The transmission pattern observed for half-compatible crosses involving a total of seven alleles confirmed that this locus was in fact the gametophytic self-

incompatibility locus. In addition, the predictions of the molecular assay were in accord with patterns of fruit-set among the progeny. We conclude that the molecular assay is a reliable method for determination of the *S*-locus genotype of individual plants.

The number and frequency of *S*-alleles in the population

Twenty-eight different sequences were detected in 22 plants sampled in the Deep Canyon population yielding a population estimate of 43–44 alleles. This number is considerably higher than that found in a previous study using RT-PCR to assess *S*-allele diversity in another Solanaceous species, horsenettle (*Solanum carolinense*). In that study the total number of alleles held among two widely separate populations was not much greater than 14 (Richman *et al.*, 1995). Low *S*-allele diversity in horsenettle was attributed to the population structure of this rhizomatous colonizing species, consisting of many small and probably short-lived patches. By contrast, the high diversity of the ground cherry sample is comparable to reports based on crossing experiments for other species with GSI. Models of the relationship between population size and the number of *S*-alleles maintained at equilibrium under GSI suggest that a population size of about 6000–10 000 is required to maintain 44 alleles at equilibrium, assuming mutation rates to new specificities between 10^{-6} and 10^{-9} per generation (Wright, 1960; Vekemans & Slatkin, 1994). Several lines of evidence suggest that the origination rate is probably not higher than 10^{-6} (Vekemans & Slatkin, 1994). Origination rates lower than 10^{-9} appear inconsistent with high *S*-diversity observed in small, isolated populations (Vekemans & Slatkin, 1994).

A test of the hypothesis of equality of allele frequencies, expected under GSI, found that the allele frequencies in ground cherry are significantly uneven. In order to examine the basis for this result, we generated the expected sampling distribution for a sample of size $r = 22$ from a population of 8000 individuals, which, assuming an origination rate of 10^{-7} , will contain 44 alleles at equilibrium (see eqn 6 of Yokoyama & Hetherington, 1982). The particular combination of population size and origination rate used is not important for the present purpose, so long as the expected number of alleles maintained is near 44. To show this, we also generated the sampling distribution assuming a sample of size $r = 22$ from a population of infinite size containing exactly 44 alleles in equal frequency. The expected number of alleles sampled from an infinite popula-

tion is $n = 27.99$, and for the finite population $n = 27.54$, both of which are quite close to the observed $n = 28$. The expected and observed sampling distributions for the Deep Canyon sample are shown in Fig. 3. As suggested by the marginal significance of the test ($P = 0.046$), the deviation from equal frequency is not particularly marked

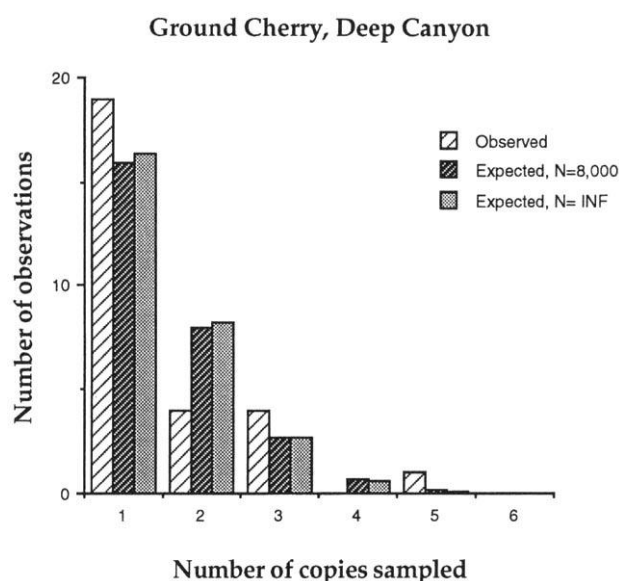


Fig. 3 Observed and expected numbers of copies of different *S*-alleles in a sample. Sample size is 22 plants. Expected numbers were calculated (1) assuming a population size of 8000 and origination rate of 10^{-7} using eqn 6 in Yokoyama & Hetherington (1982) and (2) assuming an effectively infinite population containing exactly 44 alleles at equal frequency.

(Fig. 3). There appear to be rather more rare alleles than expected, and one allele was sampled five times, an event with low probability ($P = 0.0027$ for the finite population discussed above).

The significance level of the χ^2 statistic used to test the hypothesis of equal frequencies was evaluated by the use of Monte Carlo simulation. This approach was motivated by the observation that the *P*-value associated with the Mantel χ^2 statistic for $n - 1$ degrees of freedom was quite large ($P = 0.90$), indicating that the distribution was somewhat more even than expected, whereas the sampling distribution in Fig. 3 suggests otherwise. Levin (1993) also reported a surprisingly high *P*-value in a test of *S*-allele frequency in *Phlox drummondii*. It appears that these *P*-values are inflated because of the small sample size relative to the number of alleles recovered. Probability values obtained using the Monte Carlo approach, sample sizes and *R*-values for several studies are given in Table 6. When sample size is small the Monte Carlo significance level can differ greatly from the table value, indicating that the assumption of normality is not appropriate in these instances (Mantel, 1974).

It is important to point out that the population sample of ground cherry appears to have been collected on a spatial scale which is different from previous studies. The present sample represents somewhat fewer than half of the individuals located after careful searching of an area of 0.5 km². Thus it seems likely that individuals sampled in this study were actually relatively close to each other compared to individuals sampled from dense populations occurring over smaller areas. This would

Table 6 Sample size (r), estimated number of alleles in the population (N) using Paxman's (1963) ML method, thoroughness of sampling (R), and significance levels for Mantel's χ^2 reported in the literature (citations below) and as evaluated by Monte Carlo simulation, using the appropriate r and N

Taxon	r	N	R	Mantel	Monte Carlo
<i>Papaver rhoeas</i> ¹	36	28	0.67	$P = 0.009$	$P < 0.001$
	36	34	0.61	$P = 0.022$	$P < 0.001$
	51	32	0.72	$P < 0.001$	$P < 0.001$
<i>Oenothera organensis</i> ²	37	31	0.65	$P = 0.995$	$P = 0.981$
<i>Phlox drummondii</i> ³	24	44	0.40	$P = 0.995$	$P = 0.402$
<i>Solanum carolinense</i> ⁴	14	12	0.68	$P > 0.500$	$P = 0.150$
	12	14	0.57	$P > 0.975$	$P = 0.918$
<i>Physalis crassifolia</i>	22	44	0.39	$P = 0.900$	$P = 0.046$

¹Lawrence & O'Donnell (1981).

²Emerson (1939) cited by Lawrence & O'Donnell (1981).

³Levin (1993).

⁴Richman *et al.* 1995.

seem to enhance the potential for an unequal frequency distribution arising from sampling of relatives. However, although the present sample is small, there is little indication of such neighbourhood effects, at least at the scale of the transect. For example, allele S_{11} , which was sampled five times, occurred at widely scattered locations in each of the four equal area quadrants within the transect. Finally, the population estimate of 6000–10 000 implies that a conservative estimate of the area occupied by this population studied in the Deep Canyon transect is more than 50 km².

Conclusions

Recent advances in the molecular genetics of self-incompatibility now permit a new approach to the study of S -allele diversity in natural populations. In this and a previous study (Richman *et al.*, 1995), we have presented evidence that a PCR assay is an efficient and accurate method for determining the S -locus genotype of individual plants and for estimating S -allele diversity in natural populations of Solanaceous species. The use of a molecular assay presents several advantages over previous methodologies. The method utilizes small amounts of stylar tissue collected in the field. Surveys of S -allele diversity can now, in principle, be carried out in long-lived taxa which cannot be maintained in the greenhouse in any numbers, or in those producing insufficient numbers of flowers for crossing experiments. In addition, the direct determination of S -alleles by sequence differences permits comparison of the relationships among S -alleles from different populations and species. The use of genetic variation under balancing selection, of which frequency-dependent selection at the S -locus is a special case, has been suggested as an alternative to neutral genetic variation for the inference of history of populations and species (Takahata, 1993a,b). The S -locus appears an ideal candidate in which to explore the utility of this approach, because the nature of selection acting on variation at this locus is well understood.

Acknowledgements

We thank Dr T. D. Price for helpful discussion, and S. Rupert for laboratory assistance. The research was supported by an Alfred P. Sloan Postdoctoral Fellowship (A.D.R.), Public Health Service grant GM 37841 (M.K.U.), NSF Postdoctoral Fellowship DEB-94-04386 (A.D.R.), and NSF grant DEB 93-06473 (J.R.K.).

References

- AI, Y., SINGH, A., COLEMAN, C. E., IOERGER, T. R., KHEYR-POUR, A. AND KAO, T.-H. 1990. Self-incompatibility in *Petunia inflata*: isolation and characterization of cDNAs encoding three S -allele-associated proteins. *Sex. Plant Repro.*, **3**, 130–138.
- BROOHTHAERTS, W., JANSSENS, G. A., PROOST, P. AND BROEKAERT, W. F. 1995. cDNA cloning and molecular analysis of two self-incompatibility alleles from apple. *Plant Mol. Biol.*, **27**, 499–511.
- CAMPBELL, J. M. AND LAWRENCE, M. J. 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.
- EMERSON, S. 1938. The genetics of self-incompatibility in *Oenothera organensis*. *Evolution*, **23**, 190–202.
- EMERSON, S. 1939. A preliminary survey of the *Oenothera organensis* population. *Evolution*, **24**, 524–537.
- IOERGER, T. R., GOHLKE, J. R., XU, B. AND KAO, T.-H. 1991. Primary structural features of the self-incompatibility protein in Solanaceae. *Sex. Plant Repro.*, **4**, 81–87.
- JOST, W., BAK, H., GLUND, K., TERPSTRA, P. AND BEINTEME, J. J. 1991. Amino acid sequence of an extracellular, phosphate-starvation-induced ribonuclease from cultured tomato (*Lycopersicon esculentum*) cells. *Eur. J. Biochem.*, **198**, 1–6.
- KHEYR-POUR, A., BINTRIM, S. B., IOERGER, T. R., REMY, R., HAMMOND, S. A. AND KAO, T.-H. 1990. Sequence diversity of pistil S -proteins associated with gametophytic self-incompatibility in *Nicotiana glauca*. *Sex. Plant Repro.*, **3**, 88–97.
- LAWRENCE, M. J. 1975. The genetics of self-incompatibility in *Papaver rhoeas*. *Proc. R. Soc. B.*, **188**, 275–285.
- LAWRENCE, M. J. AND O'DONNELL, S. 1981. The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. III. The number and frequency of S -alleles in two further natural populations (R102 and R104). *Heredity*, **47**, 53–61.
- LEVIN, D. A. 1993. S -gene polymorphism in *Phlox drummondii*. *Heredity*, **71**, 193–198.
- MANTEL, N. 1974. Approaches to a health research occupancy problem. *Biometrics*, **30**, 355–362.
- O'DONNELL, S. AND LAWRENCE, M. J. 1984. The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. IV. The estimation of the number of alleles in a population. *Heredity*, **53**, 495–507.
- OCKENDON, D. J. 1974. Distribution of self-incompatibility alleles and breeding structure of open-pollinated cultivars of Brussels sprouts. *Heredity*, **33**, 159–171.
- PANDEY, K. K. 1957. Genetics of self-incompatibility in *Physalis ixocarpa* Brot. – a new system. *Am. J. Bot.*, **44**, 879–887.
- PAXMAN, G. J. 1963. The maximum likelihood estimation of the number of self-sterility alleles in a population. *Genetics*, **48**, 1029–1032.
- RICHMAN, A. D., KAO, T.-H., SCHAEFFER, S. W. AND

- UYENOYAMA, M. K. 1995. S-allele sequence diversity in natural populations of *Solanum carolinense*. (Horse-nettle) *Heredity*, **75**, 405–415.
- ROYO, J., KOWYAMA, Y. AND CLARKE, A. E. 1994. Cloning and nucleotide sequence of two S-RNases from *Lycopersicon peruvianum* (L.) Mill. *Pl. Physiol.*, **105**, 751–752.
- TAKAHATA, N. 1993a. Allelic genealogy and human evolution. *Mol. Biol. Evol.*, **10**, 2–22.
- TAKAHATA, N. 1993b. Evolutionary genetics of human paleo-populations. In: Takahata, N. and Clark, A. G. (eds) *Mechanisms of Molecular Evolution*, pp. 1–21. Sinauer Associates, Sunderland, MA.
- THOMPSON, J. D., HIGGINS, D. G. AND GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.*, **22**, 4673–4680.
- VEKEMANS, X. AND SLATKIN, M. 1994. Gene and allelic genealogies at a gametophytic self-incompatibility locus. *Genetics*, **137**, 1157–1165.
- WRIGHT, S. 1960. On the number of self-incompatibility alleles maintained in equilibrium by a given mutation rate in population of a given size: a reexamination. *Biometrics*, **16**, 61–85.
- XU, B., MU, J., NEVINS, D. L., GRUN, P. AND KAO, T.-H. 1990. Cloning and sequencing of cDNAs encoding two self-incompatibility associated proteins in *Solanum chacoense*. *Mol. Gen. Genet.*, **224**, 341–346.
- YOKOYAMA, S. AND HETHERINGTON, L. E. 1982. The expected number of self-incompatibility alleles in finite plant populations. *Heredity*, **48**, 299–303.