Mol Gen Genet (1993) 238:419–427

MGG

© Springer-Verlag 1993

Molecular diversity at the self-incompatibility locus is a salient feature in natural populations of wild tomato (*Lycopersicon peruvianum*)

Bruce A. Rivers[1], Robert Bernatzky[1], Steven J. Robinson[2], Willi Jahnen-Dechent[3]

[1] Department of Plant and Soil Science, French Hall, University of Massachusetts, Amherst, MA 01003, USA
[2] Department of Biology, University of Massachusetts, Amherst, MA 01003, USA
[3] Institut für Physiologische Chemie und Pathobiochemie der Johannes-Gutenberg-Universität, Mainz, Germany

Received: 6 August 1992 / Accepted: 5 November 1992

Abstract. A cDNA encoding a stylar protein was cloned from flowers of self-incompatible wild tomato (*Lycopersicon peruvianum*). The corresponding gene was mapped to the *S* locus, which is responsible for self-incompatibility. The nucleotide sequence was determined for this allele, and compared to other *S*-related sequences in the Solanaceae. The *S* allele was used to probe DNA from 92 plants comprising 10 natural populations of *Lycopersicon peruvianum.* Hybridization was conducted under moderate and permissive stringencies in order to detect homologous sequences. Few alleles were detected, even under permissive conditions, underscoring the great sequence diversity at this locus. Those alleles that were detected are highly homologous. Sequences could not be detected in self-incompatible *Nicotiana alata*, self-compatible *L. esculentum* (cultivated tomato) or self-compatible *L. hirsutum.* However, hybridization to an individual of self-incompatible *L. hirsutum* revealed a closely related sequence that maps to the *S* locus in this reproductively isolated species. This supports the finding that *S* locus polymorphism predates speciation. The extraordinarily high degree of sequence diversity present in the gametophytic self-incompatibility system is discussed in the context of other highly divergent systems representing several kingdoms.

Key words: Tomato – Lycopersicon peruvianum – Lycopersicon hirsutum – Self incompatibility (SI)

Introduction

Self-incompatibility (SI) prevents self-fertilization in flowering plants, effectively ensuring genetic diversity by promoting outcrossing. In gametophytic SI, pollen (the male gametophyte) expresses its own genotype. Whenever the *S*-allele of the haploid pollen is the same as either of the alleles present in the diploid pistil, pollen tube growth is arrested and fertilization is prevented. Gametophytic SI (far more common than the sporophytic system) is controlled by a single locus (*S*-locus; deNettancourt 1977) encoding a glycoprotein necessary for the incompatibility response.

An intriguing characteristic of the *S*-locus is the extraordinarily high degree of sequence divergence within a species. Sequences which code for the SI glycoproteins have been isolated from *Nicotiana alata* (Anderson et al. 1989). Comparisons of the S_1, S_2, S_3, and S_6 alleles in *N. alata* revealed an overall homology of only 51.5%, attesting to the tremendous diversity at this locus. Ioerger and colleagues (1990) compared 11 gametophytic *S*-alleles representing 3 genera from the family Solanaceae, including the *S*-alleles of *N. alata.* These comparisons revealed that amino acid sequence similarity can be as low as 40% within a species. Furthermore, interspecific similarities are often higher than intraspecific similarities, indicating that polymorphism at this locus predates speciation (Ioeger et al. 1990).

Wright (1939) observed that all species within a population are necessarily heterozygous at the *S* locus; any rare allele arising would be reproductively selected for, since pollen possessing this allele would be less likely to land on a stigma which shared that allele, so that these alleles would be positively selected for when compared to neutral alleles of other genes. Wright's model predicts that functional *S*-alleles would remain in the gene pool for great lengths of time. *S*-alleles within a population should therefore be numerous, ancient, and diverse.

Little work has focused on the genetic diversity found at the *S-locus* within natural populations. Emerson (1939) estimated the presence of a minimum of 45 alleles in populations of *Oenothera organensis*; Atwood (1944) located over 30 alleles in natural populations of *Trifolium repens*, and Campbell and Lawrence (1981) uncovered 32 alleles in natural populations of *Papaver rhoeas.* To date, no studies incorporating molecular biological techniques have been conducted on *S-locus* diversity found within natural populations.

Communicated by E. Meyerowitz

Correspondence to: R. Bernatzky

In this study, an S allele from *Lycopersicon peruvianum* was cloned, sequenced and mapped to the S-locus. It was then used to probe 92 individuals representing 10 reproductively isolated populations of wild tomato (*L. peruvianum*). Southern hybridizations were conducted under different stringencies in order to examine the diversity of the S-alleles within these natural populations.

Materials and methods

Plant material. Ninety-two *L. peruvianum* plants representing 10 populations were used in this study (Table 1). Seed was generously provided by Dr. Charles M. Rick at the University of California, Davis. These morphologically and ecologically diverse populations are from primary collections and have not been regrown nor subjected to selection. Dr. Rick also provided *L. peruvianum* LA2157 used in one of the mapping studies. *Lycopersicon hirsutum* (LA1777 and LA1675) seed was obtained from Dr. Steven D. Tanksley at Cornell University.

mRNA isolation. Styles from mature (fully opened) flowers from one individual of *L. peruvianum* (LA2163, Table 1) were frozen in liquid nitrogen and ground to a fine powder using a mortar and pestle. Total RNA was extracted from the powder using the guanidinium isothiocyanate procedure (Chirgwin et al. 1979) in combination with phenol/chloroform extraction and $LiCl_2$ fractionation (Sambrook et al. 1989). RNA was recovered between fractionation steps by ethanol precipitation. RNA concentration was determined spectrophotometrically. Total RNA yield was 2.3 mg from 1.3 g mature styles and 0.14 mg from 0.2 g of styles from closed flowers.

Poly(A)$^+$RNA was prepared from total RNA by chromatography on oligo-dT cellulose spun columns (Pharmacia) following the manufacturer's instructions. Twenty micrograms of poly(A)$^+$RNA were isolated from 1.5 mg of total RNA after two spun column fractionations.

Table 1. List of *L. peruvianum* populations

Accession No.	Collection site
LA2163	Yamaluc, Cajamarca, Peru
LA2326	Balsas, Amazonas, Peru
LA1981	Huallaca, Ancash, Peru
LA1365	Caranquilloc, Ancash, Peru
LA1274	Pacaibamba, Lima, Peru
LA1283	Santa Cruz de Laya, Lima, Peru
LA1952	Loruas de Camana, Arequipa, Peru
LA1955	Matanari, Arequipa, Peru
LA2774	Azapa, Tarapaca, Chile
LA2770	Lluta, Tarapaca, Chile

Accessions are ordered in approximately ascending South latitude. These plants represent primary collections and were provided by Dr. Charles M. Rick of the University of California at Davis. LA2163 contains the plant from which the S_5 allele was cloned

cDNA synthesis. cDNA was synthesized from 5 μg of poly(A)$^+$RNA using a cDNA synthesis kit (Pharmacia). After addition of *Eco*RI/*Not*I adaptors (Pharmacia) the cDNA was ligated into the plasmid pUC18. Competent *Escherichia coli* cells (DH5α, Bethesda Research Laboratories) were transformed with the plasmids and plated onto LB plates containing ampicillin, 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal) and isopropyl-β-D-thio-galactopyranoside (IPTG). Individual transformed clones were replated as streak colonies and further used for DNA mini-preps as outlined by Willimzig (1989).

Library screening. From each of 170 plasmid mini-preps, 1 μg of DNA was digested with 1U of *Eco*RI (BRL). Samples were separated on 0.9% agarose gels in a buffer containing 100 mM TRIS-HCl, 25 mM EDTA, 12.5 mM sodium acetate, pH 8.1. Gels were run at 25 V overnight. Gels were dry-blotted (using only the liquid in the gel to transfer the DNA after soaking in 0.4 N NaOH) onto Zeta-probe membrane (Bio-Rad). All filters were pre-hybridized and hybridized as outlined by Bernatzky and Tanksley (1986b), with minor modifications. Plasmid filters were bathed in plastic boxes in hybridization buffer at a rate of 50 ml buffer/100 cm^2 of membrane (buffer: 5 × SSC (0.75 M sodium chloride, 0.125 M sodium citrate), 0.6% SDS, 50 mM sodium phosphate, pH 7.2, 5 × Denhardt's solution, 2.5 mM EDTA, 100 μg/ml salmon sperm DNA, 5% dextran sulfate). Hybridization was carried out overnight in an air bath incubator (Labline) at 60° C using the S_6 allele of *N. alata* as a probe (Anderson et al. 1989). The inserts from the putative S-clones were isolated from an agarose gel (Tautz and Renz 1983), and labeled with [α^{32}P]dCTP according to the random primer technique of Feinberg and Vogelstein (1983). Plasmid filters were washed thrice for 20 min each at 60° C in 0.1% SDS at the following salt concentrations: 2 ×, 2 ×, 1 × SSC.

Nucleotide sequence determination. Double-stranded plasmid DNA was sequenced using ^{35}S label and a modified T7 DNA polymerase (Sequenase, US Biochemicals) and a modified version of the dideoxy sequencing protocol described by the supplier. Sequencing compression artifacts were avoided by adding terminal deoxynucleotidyl transferase in a complete dNTP cocktail, as recommended by Fawcett and Bartlett (1990). DNA fragments from the dideoxy sequencing reactions were separated by electrophoresis on an IBI ST45 sequencer at a constant power of 80 W. A 3 μl sample from each of the reactions was loaded and fractionated for approximately 3 h. The dried gel was exposed to film for 1–2 days; the sequence was then read visually. Nucleotide sequence was fully determined for both strands, using fragments of the original cDNA subcloned into Bluescript KS$^+$ (Stratagene).

Plant DNA extraction. DNA was extracted from 2 g of leaves according to a modified procedure of Bernatzky and Tanksley (1986c) and Murray and Thompson (1980). Instead of filtering the plant homogenate and

deriving a crude nuclear pellet as the source of DNA, the DNA was extracted by directly lysing the homogenate.

Segregation analysis in L. peruvianum. A cross between the mRNA source plant (LA2163, Table 1) and a plant from LA2157 produced a hybrid, which was shown by Southern hybridization to contain the cloned *S*-allele. This hybrid was used as the female in a backcross to the LA2163 parent, resulting in a population of 38 plants which segregated for the cloned *S*-allele and corresponding *S*-related protein. Proteins were analyzed by SDS PAGE. DNA was extracted from these plants, digested with *Eco*RI, electrophoresed and blotted onto Zetaprobe. Hybridization was conducted at 68° C as described below.

Southern hybridization of population samples. Using the cloned *S*-allele from LA2163 as a probe, two series of hybridizations were conducted among the 10 populations of *L. peruvianum*. Southern blotting, hybridization buffer, and probe preparation were the same as for the library screening. Filters were pre-hybridized in a hybridization incubator (Robbins Scientific) at a rate of 25 ml buffer/500 cm^2 membrane for 1 h and hybridized at a rate of 5 ml/500 cm^2 membrane for 16 h. One series was conducted under moderately stringent conditions (68° C, 2 ×, 1 ×, 0.5 × SSC washes). After exposure to film, the filters were stripped (two 10 min washes in 0.1 × SSC, 0.1% SDS at 80° C) and a second hybridization was conducted under more permissive conditions (60° C; 2 ×, 2 ×, 1 × SSC washes). *S*-related sequences from the populations that did hybridize under moderately stringent conditions were further tested under higher stringency washes (final wash 68° C; 0.2 × SSC or 0.1 × SSC). A genomic clone (TG184) of tomato (*Lycopersicon esculentum*) that is linked to the *S*-locus (R. Bernatzky, unpublished data) was used as a control to confirm quantity of DNA loaded on gel, DNA transfer during blotting, and quality of Southern hybridization.

Segregation of S-related proteins and DNA sequences in L. hirsutum. Major stylar proteins were mapped to the *S*-locus in an SI accession of *L. hirsutum* (LA1777). These *S*-related proteins were of similar molecular weight to *S*-related proteins found in *L. peruvianum*. A semi-compatible cross between two *L. hirsutum* plants from accession LA1777 was made. Twenty-five progeny were classified into 2 incompatibility groups (17 in one and 8 in the other) on the basis of controlled pollinations. The co-segregation of these stylar proteins with the breeding behavior was determined by SDS-PAGE. A population segregating for an allele in *L. hirsutum* with sufficient homology to the cloned allele from *L. peruvianum* was developed from a cross between individuals of LA1777 and LA1675. Stylar proteins from 49 plants were used as markers for the *S*-locus and analysis of segregating DNA fragments was as described for *L. peruvianum*.

Results

An *S*-allele clone was isolated from the *L. peruvianum* style cDNA library by screening with the *N. alata* S_6 probe. One clone, p76, had a convenient insert size of approximately 700 bp and was used in subsequent sequencing reactions and hybridization experiments.

The derived amino acid sequence of the putative *L. peruvianum* *S*-allele contained in p76 is presented in Fig. 1. Comparisons of the amino acid sequences of p76 and the S_6 allele of *N. alata* indicate shared, conserved

```
CGACCGCTTGTTTTTTCTCCCGTTTGTGGGGACTTCGATTACTTCCAACTTGTTTTACAA     60
R  P  L  V  F  S  P  V  C  G  D  F  D  Y  F  Q  L  V  L  Q

TGGCCAAGATCTTTTTGCAAAACAAGATATTGCCCAAATCCAGTTCCAAGAAACTTCACG    120
W  P  R  S  F  C  K  T  R  Y  C  P  N  P  V  P  R  N  F  T
                                                   Δ
ATCCATGGGCTATGGCCTGATAAACAGAGAATTATGCCGATTAACTGCCCGCGGAAAGAG    180
I  H  G  L  W  P  D  K  Q  R  I  M  P  I  N  C  P  R  K  E
   *
AGCTACAAAAGTATTACGGACTCTAAGAAAATCAAACTACTGGAGCAACACTGGCCCGAT    240
S  Y  K  S  I  T  D  S  K  K  I  K  L  L  E  Q  H  W  P  D

TTGACCTCCAATCAAGGCAGTGCAGAATTCTGGAGATATCAATACAAGAAGCACGGAACG    300
L  T  S  N  Q  G  S  A  E  F  W  R  Y  Q  Y  K  K  H  G  T
                                                   *
TGTAGTGTGGACCTATATAATCAAGAACAATATTTTGATTTAGCCATTGAATTAAAAGAG    360
C  S  V  D  L  Y  N  Q  E  Q  Y  F  D  L  A  I  E  L  K  E

AAGTTTGATCTTTTGAAAACTCTCAAAAATCACGGAATTACTCCATCAAAAACTAATACA    420
K  F  D  L  L  K  T  L  K  N  H  G  I  T  P  S  K  T  N  T

GTTATAGATGTCGAAGAAGCTATCAAGGCCGTAACAAAAGAGGTTCCTAACCTCAACTGC    480
V  I  D  V  E  E  A  I  K  A  V  T  K  E  V  P  N  L  N  C

ATTGGTGACTCTAGTCAAACCATGGAACTACTGGAGATAGGCATATGTTTTAACAAAGAG    540
I  G  D  S  S  Q  T  M  E  L  L  E  I  G  I  C  F  N  K  E

GGAACTACGGTGATTGCTTGTCGTCGACGTTGGAATAATCATCCTAACGGGAACCAGAAG    600
G  T  T  V  I  A  C  R  R  R  W  N  N  H  P  N  G  N  Q  K

ATTACTCTTCCACCATGA                                              618
I  T  L  P  P  stop
```

Fig. 1. Nucleic acid and derived amino acid sequences of the S_5 allele of *L. peruvianum*. Numbers refer to the base and letters represent the amino acids coded for by the sequence of nucleotides. Amino acid residues that are either identical or conservatively exchanged in an alignment with the S_2 allele of *N. alata* (Anderson et al. 1986) are written in **boldface**. Cys residues that are at similar positions are underlined. Proposed RNAse active site His residues as reported in McClure et al. (1989) are marked with *asterisks*. The alignment was done using the program GAP (Devereux et al. 1984)

residues. These areas of conservation are also present in the 11 S gene sequences presented by Ioerger et al. (1990). Comparisons of the 11 published derived amino acid sequences with our sequence indicate several cysteine residues in common, suggesting that the tertiary structure of these S-related proteins may be similar. The *L. peruvianum* allele is designated as S_5 since four other alleles from this species have been characterized to some extent at the protein level (Mau et al. 1986).

Previous denaturing electrophoresis analysis had shown that populations of *L. peruvianum* are polymorphic for discrete stylar proteins in the molecular weight range of 20–30 kDa. These proteins have been mapped to the S locus through controlled pollinations of 77 progeny from a fully compatible cross between two plants from accession LA2163 (R. Bernatzky, unpublished results). In the present work, 38 progeny from the *L. peruvianum* backcross population (LA2163 × LA2157) × LA2163, were segregating for the S_5 allele. The S-related proteins from the LA2163 parent were used as markers for the S-locus in the backcross progeny. The presumed protein product of the cloned S-allele was found in a ratio of 19 with:19 without. This is expected since the progeny were derived from a cross between male and female parents each heterozygous for the cloned allele. Pollen bearing the cloned allele would be eliminated by incompatibility, and the cloned allele donated by the eggs would segregate in a 1:1 ratio. Southern analysis of the DNA from these same plants indicates that p76 was hybridizing to only the cloned allele donated from the LA2163 parent (data not shown). This lack of hybridization at moderate stringencies to other alleles is consistent with the observed lack of hybridization between S alleles of *N. alata* (Bernatzky et al. 1988; Anderson et al. 1989). Homologous DNA fragments co-segregated exactly with the stylar protein, confirming that the cloned sequence maps to the S-locus, and probably codes for the stylar protein, and thus supporting the sequence data which indicate that p76 is an S-related gene product.

The degree of sequence homology of the cloned allele to other alleles was further investigated in 92 other *L. peruvianum* plants representing 10 diverse populations (Table 1). These plants are derived from seed that was collected in the field. Hybridization of p76 to genomic DNA from these individuals revealed highly varied levels of homology to other S-related sequences in natural populations of these plants. In fact, the p76 probe hybridized to only six of the 92 plants under conditions of moderate stringency (hybridization at 68° C, final wash at 68° C, 0.5 × SSC). The population distribution of these six plants, and the sizes of hybridizing restriction

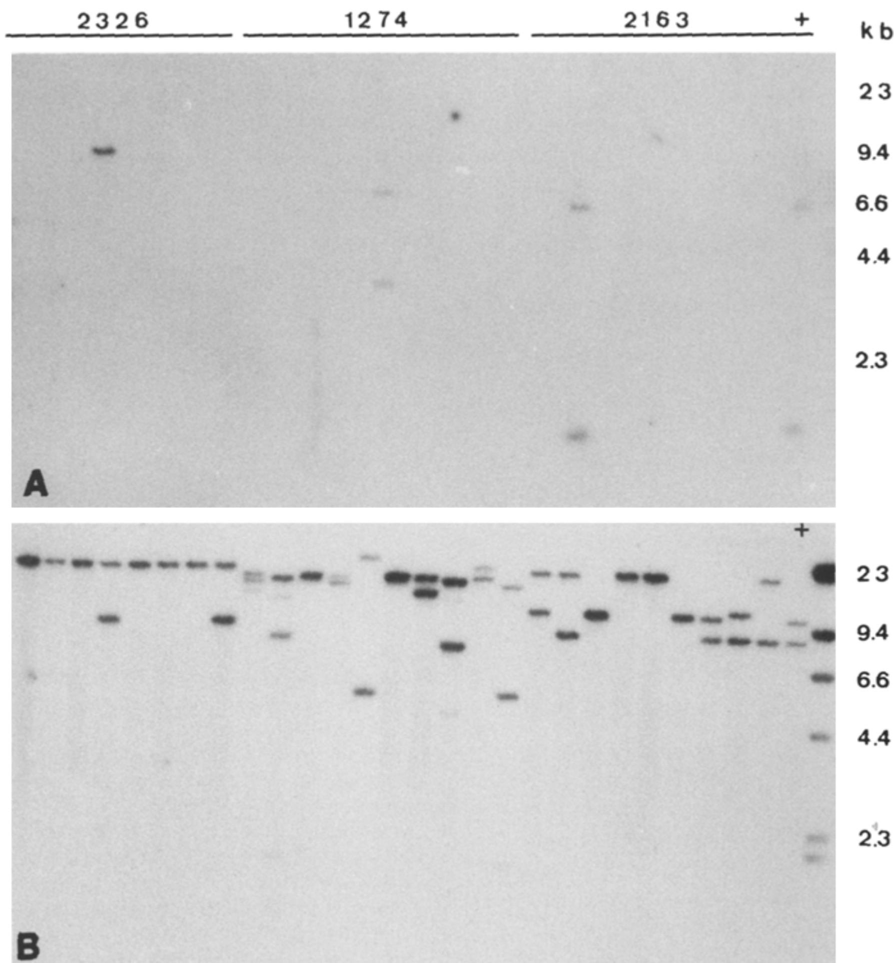

Fig. 2A. Autoradiograph of the S_5 allele hybridized to 3 different natural populations of *L. peruvianum*. Hybridization was conducted at 68° C under the conditions described in the text. Note that only 3 additional alleles were detected. The lane at extreme right (+) contains DNA from the plant from which the S_5 allele was cloned. All samples (~3 μg) of genomic DNA were digested with *Eco*RI. Plants with two-banded phenotypes contain a conserved internal *Eco*RI site (see text). Note the similarity in fragment size between the two alleles in population LA2163; this plant has been shown to contain the S_5 allele through controlled pollinations. Numbers at the top refer to accession and numbers on the right indicate DNA fragment sizes in kb. **B** Autoradiograph of the same filter probed with a genomic fragment linked to the S locus on *chromosome 1*. The probe (*TG184*) was used as a control to confirm that the lack of positive signals using the S_5 probe was not due to the blotting and hybridization procedure. The slight variation in signal intensity between individuals is likely to be due to variation in amounts of DNA loaded

Table 2. Results of filter hybridization of *L. peruvianum* genomic DNA to the S_5 probe conducted under moderate (68° C) and permissive (60° C) conditions

Accession no.	No. of plants sampled	Homologous fragments (kb) at 68° C	Additional fragments (kb) detected at 60° C
LA2163	10	6.0, 1.3 (1)	0
LA2326	8	8.0 (1)	0
LA1981	9	0	0.5 (1) 10.0 (1)
LA1365	10	6.2, 3.5 (1)	18.5 (1)
LA1274	10	6.2, 3.3 (1)	0
LA1283	10	0	0
LA1952	9	6.2, 3.5 (1)	4.0, 1.3 (2) 10.0 (1) 0.4 (1)
LA1955	10	18.0 (1)	9.0, 3.3 (1) 14.0 (1)
LA2774	10	0	2.0 (3) 0.4 (1)
LA2770	8	0	0

Sizes of homologous fragments (kb) were compared to λ DNA cut with *Hin*dIII. Numbers in parentheses represent the number of plants within the accession possessing the indicated homologous fragments

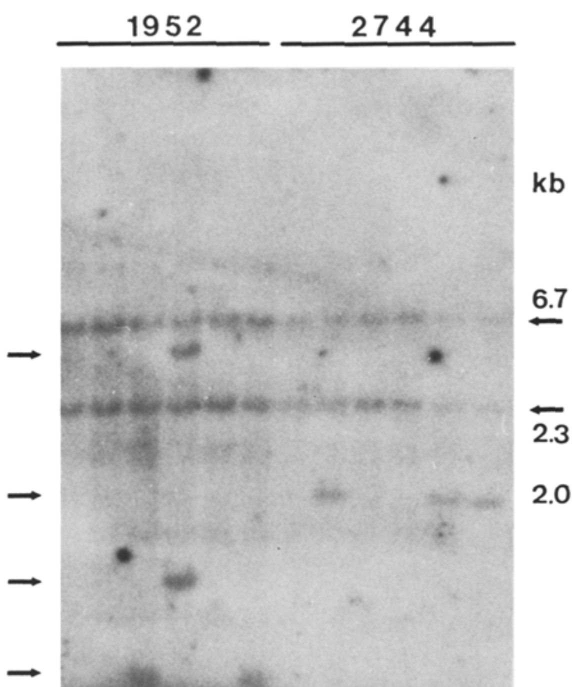

Fig. 3. Autoradiograph of *L. peruvianum* genomic DNA probed with the S_5 allele at lower stringency (60° C). It should be noted that this figure is a representative portion of the filter used in the hybridization and not all plants from the respective populations cited in Table 2 are shown. *Arrows* at left indicate additional potential alleles. *Arrows* at right indicate fragments common to all samples at 60° C. The identity of common fragments is unknown, though it is not likely to be chloroplast DNA, since our probe did not hybridize to chloroplast DNA from radish

fragments, are given in Table 2. This lack of detectable hybridization to the majority of plants in three populations is illustrated in Fig. 2. Even under conditions of low stringency (hybridization at 60° C, 5 × SSC, and final wash at 60° C, 1 × SSC) only an additional 13 plants showed any significant hybridization (Table 2, Fig. 3). Under these conditions, two additional fragments present in every plant DNA sample in the populations were observed, with sizes of approximately 6.6 and 3.3 kb, respectively (Fig. 3). Due to the conserved nature of these bands, the possibility that they represented chloroplast DNA sequences that contain regions of homology to the probe was considered. However, the S_5 probe did not hybridize to radish chloroplast DNA (provided by Chris Makaroff, Miami University) that was blotted and probed at 60° C, and so they are probably not conserved chloroplast sequences. The origin of these bands remains under investigation.

To demonstrate that the limited homology of the S_5 allele to other *S*-alleles was not due to the blotting or hybridization procedure, each filter was hybridized at moderate stringency to a control probe, TG184, consisting of a cloned genomic fragment from *L. esculentum* linked to the *S*-locus on chromosome 1 (Fig. 2; linkage of TG184 to the *S*-locus, R. Bernatzky, unpublished). Hybridization to the TG184 control varies only slightly among individuals, probably due to variation in the amount of DNA loaded. In contrast, the lack of hybridization of p76 to the majority of plants is most likely due to high sequence divergence between the cloned allele and other alleles in these natural populations.

Individuals whose genomes hybridize to p76 under conditions of low stringency are distributed throughout eight of the ten populations. It is noteworthy that the population from which the cloned allele was isolated (LA2163) did not contain a higher number of homologous alleles than the other poulations; only one additional LA2163 individual displays homology to p76. This one individual has an *Eco*RI restriction pattern similar to that of the plant from which the *S*-allele was cloned. The plant used for cloning is cross compatible with the other seven plants of the population. Progeny testing revealed that plants from this population contain an allele in common with the plant from which the clone was derived. One of these plants is the one that shows hybridization to p76, indicating that it contains the same S_5-allele.

Some of the plants that do hybridize to p76 contain more than one homologous DNA fragment. Four individuals produce homologous *Eco*RI fragments at moderate stringencies; two additional individuals had two-banded phenotypes at low stringency (Table 2). The multiply-banded phenotype does not result from two alleles highly homologous to p76 in each plant. Rather, in each of these plants, p76 hybridizes to a single allele containing an internal *Eco*RI site. To demonstrate this, p76, which contains an internal *Eco*RI site, was digested with *Eco*RI. The resulting 440 bp subfragment was used as a probe on Southern blots containing *Eco*RI-digested genomic DNA from the individuals with homologous

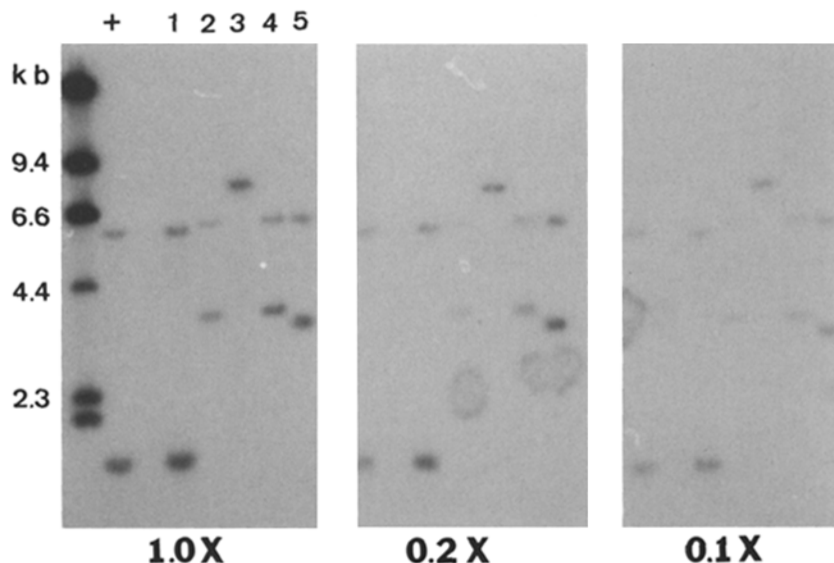

Fig. 4. Results of hybridization analysis at increased stringency of the five individuals that exhibited homology to the S_5 probe when hybridized under conditions of moderate stringency. After hybridizing individuals at 68° C, stringency during the washes (68° C for 20 min) was increased by reducing the SSC concentration from 1 × to 0.2 × for one blot, and to 0.1 × SSC for another. The effect of increasing stringency and/or additional washes is a reduction in signal for all plants uniformly, including the plant from which the S_5 allele was cloned (+) suggesting that all of these alleles are highly homologous. Plants are from the following populations: LA2163 (lane 1), LA1365 (lane 2), LA2326 (lane 3), LA1952 (lane 4), LA1274 (lane 5)

alleles; hybridization was conducted under conditions of moderate stringency. In each case the higher molecular weight band was absent, resulting in single banded phenotypes. This indicates that the p76 was hybridizing only to single alleles in these plants.

The individuals from LA2163, LA2326, LA1365, LA1274, and LA1952 (Table 2) that hybridized to p76 under moderate stringency were also tested under higher stringency. Stringency was increased during the washing steps at 68° C (20 min each) by reducing the SSC concentration from 1 × SSC to 0.2 × SSC for one blot and then further to 0.1 × SSC for another blot. Fig. 4 shows that the effect of increasing stringency and/or additional washes is a uniform reduction in the signal for all plants, including the plant from which the allele was cloned, suggesting that these alleles are highly homologous to p76.

The probe p76 was also hybridized to several other plants in the same genus or family. Two individuals of *Nicotiana alata* homozygous for the S_2 and S_6 alleles (Anderson et al. 1989), two individuals of cultivated tomato (*L. esculentum*, cvs. Rutgers and Grosse-Lisse), one individual of *L. hirsutum* (LA1675), and three individuals of *L. hirsutum* (LA1777) were hybridized under moderate stringency. Only one of the individuals of LA1777 gave a detectable signal under these conditions. By inspection, the signal was nearly equal to that of the control (the *L. peruvianum* individual from which p76 was cloned). This interspecific hybridization under moderate stringency underscores the great age of these *S*-related sequences and further supports findings of *S*-locus polymorphism that predates speciation in the Solanaceae (Ioerger et al. 1990).

To determine whether the homologous fragment of *L. hirsutum* was an authentic *S*-related sequence, the fragment was mapped to the *S*-locus through controlled pollinations and segregation analysis of stylar proteins. Major stylar proteins of molecular weight similar to those of *L. peruvianum* can be observed in extracts from styles of *L. hirsutum* LA1777. A semi-compatible cross was made between two *L. hirsutum* plants of accession LA1777. These plants contain a major protein in common and two polymorphic proteins, possibly representing three *S*-alleles. Twenty-five progeny were classified into two incompatibility groups (17 in one and 8 in the other) on the basis of controlled pollinations. The identification of two incompatibility groups in the progeny is consistent with a common parental allele. Segregation of the stylar proteins matched exactly that of the *S*-alleles based on breeding, indicating that these proteins are products of the *S*-locus.

Using these stylar proteins as markers for the *S*-locus, we determined whether the DNA fragment from *L. hirsutum* LA1777 that hybridized to p76 mapped to the

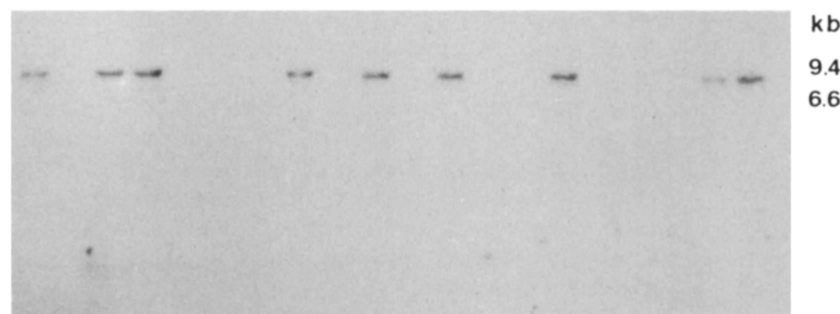

Fig. 5. Hybrization analysis of progeny of an individual of *L. hirsutum* that contains an allele with a high degree of homology to the *L. peruvianum* S_5 probe when hybridized at 68° C. The plants are segregating approximately 1:1 for presence/absence of the allele

S-locus. A second population was derived from a hybrid made between the LA1777 plant and an individual from LA1675. An individual containing the homologous allele was then backcrossed to the LA1777 parent in order to observe the segregation of the homologous allele. Forty-nine progeny were scored for the presence of the protein presumably encoded by the homologous *S*-allele. DNA was extracted from these plants and the presence of the homologous fragment was determined by Southern analysis using p76 as a probe. The segregation of the homologous DNA fragment (Fig. 5) matched exactly the segregation of the *S*-related protein. This suggests that the homologous sequence in the LA1777 plant maps to the *S*-locus, encoding a stylar protein.

Discussion

Our population survey provides evidence that extreme sequence diversity can be pervasive within as well as between populations of a single species. Although allelic comparisons of non-*S* plant genes are limited (since researchers rarely isolate multiple alleles of plant genes for comparison), a high degree of sequence conservation (Sachs et al. 1986) and good cross hybridization on Southern blots appear to be the general characteristics of most plant genes, especially within a species (Bernatzky and Tanksley 1986a; Tanksley et al. 1988). It has been estimated that sequence divergence of between 10 and 40% can be revealed by filter hybridization carried out at 68° C in Na^+ concentrations in the range of 0.5–0.05 M; with every 1° C drop in temperature corresponding to approximately 1% mismatch of sequence (Beltz et al. 1983). There are also practical considerations in the application of filter hybridization. If, for example, two strands of DNA differ by 50%, yet the mismatching is clustered so that the strands still share a relatively long portion of the overall sequence, a relatively high melting temperature may be realized. On the other hand, if the mismatching is dispersed throughout the length of the sequence such that every other base is mismatched, no hybridization will occur (Beltz et al. 1983). It is important to note, however, that interpretation of data based on Southern analysis for additional *S*-sequences must be tempered by the fact that, since diversity is so high, absence of hybridization signals cannot be taken as evidence for lack of additional SI-related sequences. Indeed, even at low stringencies, functional alleles cannot be detected in some plants (Fig. 2). Initially, our plant populations were hybridized at 68° C and washed twice at 68° C in 2 × SSC, 0.1% SDS. However, when hybridizations and washes were conducted at 60° C, an increase in the number of signals was observed, though many *S*-alleles were still undetected, underscoring the high degree of divergence characteristic of the *S*-locus. It is of special interest to note the high degree of homology between the cloned *L. peruvianum* allele and the allele found in *L. hirsutum* LA1777. Reproductive barriers preclude hybridization between these species (Rick 1986), yet the cloned allele and the allele from *L. hirsutum* are clearly more related than the cloned allele is to the majority of alleles in *L. peruvianum*, including the alleles found within LA2163, the population from which the allele was cloned.

Very reduced hybridization on Southern blots under the same conditions between alleles of *N. alata* (Bernatzky et al. 1988) provides an estimate of divergence for our current findings. The *N. alata* sequences share between 60 and 70% amino acid identity, yet give very weak cross hybridization at the DNA level. Our hybridization results suggest that the majority of alleles in our *L. peruvianum* samples share an even lower degree of amino acid identity.

In striking contrast to the low, intraspecific homology of the *L. peruvianum* alleles, amino acid identity between the cloned *L. peruvianum* allele and the allele from *L. hirsutum* is probably greater than 70%, based on signal strength under similar conditions of hybridization. This finding of higher sequence homology between taxa, as opposed to within taxa, is similar to results reported by Ioerger et al. (1990). For example, they report amino acid identity between two alleles of *N. alata* to be 45%, less than that between an *N. alata* allele and an allele from *Petunia inflata* (62.4%) based on sequence analysis.

The homologous alleles, present in the five accessions (LA2163, LA2326, LA1365, LA1274, and LA1952), which did hybridize to the S_5 probe under moderate conditions were subsequently subjected to washes of increasing stringency (1 ×, 0.2 ×, 0.1 × SSC, 68° C). Significantly, these alleles exhibited uniform reduction in signal intensity when compared with the original S_5 clone plant (present on the filters as a control). Washes at 0.1 × SSC have previously been shown to melt hybrids of the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rbcs) fragments in tomato, which are 89% homologous (Pichersky et al. 1986). In light of the data presented by Pichersky and colleagues (1986), it is indeed likely that these particular *S*-related sequences share an unusually high degree of homology in relation to our present understanding of the sequence divergence present at this locus.

The plants that contained highly homologous alleles are no longer available for fertilization tests to determine if they were indeed functionally the same alleles. Some may have been the same allele since they exhibited similar *Eco*RI restriction patterns. In fact, an internal *Eco*RI site appears to be a conserved feature of these alleles. However, a second cloned allele (S_6, unpublished) also contains a similar internal *Eco*RI site, but does not cross hybridize with the S_5 allele on Southern blots of plant genomic DNA, even under low stringency. Therefore, although the internal *Eco*RI site is conserved, it does not necessarily indicate that alleles that contain the site are closely related.

What mechanism(s) are responsible for the generation and maintenance of the large numbers of functional *S*-alleles within even relatively small reproductively isolated populations? The reproductive selective advantage that a novel *S*-allele provides may successfully maintain such novel alleles within a population, independent of any mechanism responsible for generating these alleles. Forced inbreeding has been reported to generate new

S-alleles (deNettancourt et al. 1971). However, a low frequency of reversion to parental types has also been reported (deNettancourt et al. 1971), indicating the presence of a mechanism which may restore *S*-alleles to the inbred populations, and possibly maintain the allelic diversity within the plant genome (Ebert et al. 1989).

Mechanisms reported to generate gene diversity in other single locus multiallelic systems include forms of gene conversion, intra-allelic/intragenic recombination, and/or accumulation of point mutations. These DNA rearrangements can be classified as incidental (unpredictable) or programmed (Borst and Greaves 1987).

Duplicative gene conversion, responsible for the mating type switching in yeast (reviewed in Borst and Greaves 1987), duplicative gene transposition, involved in trypanosome surface protein diversity (Pays and Steinert 1988), and allelic switching, involved in the pilus protein variability in the bacterium *Neisseria gonorrhoeae* (Segal et al. 1985, 1986) are mechanisms shown to produce high levels of genic diversity within living systems. These systems are distinct from gametophytic SI in plants in that they represent examples of gene activations that are programmed by gene rearrangement (Borst and Greaves 1987).

The major histocompatibility complex (MHC) of mammals however, offers some parallels to the gametophytic self-incompatibility system of plants. The class I and class II molecules of the MHC are important in the presentation of antigens to the vertebrate immune system and so are integral to recognition and response. These molecules are encoded by a small number of genes, possess a large number of alleles, and the diversity between alleles is relatively high, on the order of 10–15% at the DNA level (Wakeland et al. 1990; Kasahara et al. 1990). The gametophytic SI system is also based on a multi-allelic system, but the degree of diversity within a species is much higher.

The two dominant theories for the origins of diversity in the MHC are based on either a mechanism of generating gene diversification or on the retention, through selection, of ancient polymorphisms that are largely a result of point mutations (for reviews see Pease 1985; Geliebter and Nathenson 1987; Kasahara et al. 1990; Klein and Takahata 1990). The diversification mechanism involves micro-recombination between related sequences and results in clustering of mutations. The maintenance of ancient point mutations through selection is termed "trans-species mode of evolution" and states that polymorphisms are passed on from one species to the next and therefore predate speciation.

In the gametophytic self-incompatibility system of plants, mutations that alter allele specificity would be selected in a frequency-dependent manner, since any new allele would be accepted by all plants. In this system, allelic diversity is not only advantageous, it is necessary. Heterozygosity is required because homozygosity is precluded in natural populations. Recombination as a mechanism to generate new alleles was proposed by Fisher (1961) and is consistent with an observed clustering of mutations (Anderson et al. 1989; Ebert et al. 1989). Such recombination would be limited to a single locus since SI specificity is inherited as a single Mendelian factor, although the size of the locus is unknown. Work by Clark and Kao (1991) however indicates that unusually low rates of intragenic recombination are present in the *S*-alleles in the Solanaceae.

Support for the second theory on the origins of *S*-locus diversity, the maintenance of ancient polymorphisms, has been presented by Ioerger et al. (1990), and here, in the strong homology between the cloned allele from *L. peruvianum* and an allele from the reproductively isolated *L. hirsutum*. It is impressive that such a high level of diversity can exist at a single locus in geographically and reproductively isolated populations of a single species, yet presumably the mode of action of the gene products involved in inhibiting fertilization remains the same.

The great diversity found at the *S*-locus may be likened to a multi-allele family. Each member of this family, while maintaining its functional ability, also demonstrates significant divergence at the molecular level. An important characteristic of this *S*-allele family is that members can be distinguished from each other on a functional level through compatibility behavior. This unique set of characteristics allows comparison between variation at the molecular level and the resulting breeding behavior, thus providing a unique system in which changes in sequence and their functional impact on a population can be studied.

Acknowledgements. Research was supported in part by Biomedical Research Support Grant no. RR07048.

References

Anderson MA, McFadden GI, Bernatzky R, Atkinson A, Orpin T, Dedman H, Tregear G, Fernley R, Clarke AE (1989) Sequence variability of three alleles of the self-incompatibility gene of *Nicotiana alata*. Plant Cell 1:483–491

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

Beltz GA, Jacobs KA, Eickbush TA, Cherbas PT, Kafatos FC (1983) Isolation of multigene families and determination of homologies by filter hybridization methods. Methods Enzymol 100:266–285

Bernatzky R, Tanksley SD (1986a) Genetics of actin-related sequences in tomato. Theor Appl Genet 72:314–321

Bernatzky R, Tanksley SD (1986b) Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. Genetics 112:887–898

Bernatzky R, Tanksley SD (1986c) Methods for detecting single or low copy sequences in tomato on Southern blots. Plant Mol Biol Rep 4:37–41

Bernatzky R, Anderson MA, Clarke AE (1988) Molecular genetics of self-incompatibility in flowering plants. Dev Genet 9:1–12

Borst P, Greaves DR (1987) Programmed gene rearrangements altering gene expression. Science 235:658–667

Campbell JM, Lawrence MJ (1981) The population genetics of the self-incompatibility polymorphism in *Papaver rhoeas*. I. The number and distribution of *S*-alleles in families from three localities. Heredity 46:69–79

Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry 18:5294–5299

Clark AG, Kao TH (1991) Excess nonsynonymous substitution at shared polymorphic sites among self-incompatibility alleles of the Solanaceae. Proc Natl Acad Sci USA 88:9823–9827

deNettancourt D, Ecochard R, Perquin MDG, van der Drift T, Westerhof M (1971) The generation of new *S*-alleles at the incompatibility locus of *L. peruvianum*. Theor Appl Genet 41:120–129

deNettancourt D (1977) Incompatibility in angiosperms. Springer-Verlag, New York

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Ebert PR, Anderson MA, Bernatzky R, Altschuler M, Clarke AE (1989) Genetic polymorphism of self-incompatibility in flowering plants. Cell 56:255–262

Emerson S (1939) A preliminary survey of the *Oenothera organensis* population. Genetics 24:538–552

Fawcett TW, Bartlett SG (1990) An effective method of eliminating "artifact banding" when sequencing double stranded DNA templates. Biotechniques 9:44–46

Feinberg AP, Vogelstein B (1983) A technique for radio-labeling DNA restriction fragments to high specific activity. Anal Biochem 132:6–13

Fisher R (1961) A model for the generation of self-sterility alleles. J Theoret Biol 1:411–414

Geliebter J, Nathenson SG (1987) Recombination and the concerted evolution of the murine MHC. Trends Genet 3:107–112

Ioerger TR, Clark AG, Kao TH (1990) Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proc Nat Acad Sci USA 87:9732–9735

Kasahara M, Klein D, Fan W, Gutknecht J (1990) Evolution of the class II major histocompatibility complex alleles in higher primates. Immunol Rev 113:65–82

Klein J, Takahata N (1990) The major histocompatibility complex and the quest for origins. Immunol Rev 113:5–26

Mau SL, Williams EG, Atkinson A, Anderson MA, Cornish EC, Greco B, Simpson RJ, Kheyr-Pour A, Clarke AE (1986) Style proteins of wild tomato (*Lycopersicon peruvianum*) associated with expression of self-incompatibility. Planta 169:184–191

McClure BA, Haring V, Ebert PR, Anderson MA, Simpson RJ, Sakiyama F, Clarke AE (1989) Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. Nature 342:955–957

Murray M, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325

Pays E, Steinert M (1988) Control of antigen gene expression in African trypanosomes. Annu Rev Genet 22:107–126

Pease LR (1985) Diversity in H-2 genes encoding antigen-presenting molecules is generated by interactions between members of the major histocompatibility complex gene family. Transplantation 39:227–231

Pichersky E, Bernatazky R, Tanksley SD, Cashmore AR (1986) Evidence for selection as a mechanism in the concerted evolution of *Lycopersicon esculentum* (tomato) genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Proc Natl Acad Sci USA 83:3880–3884

Rick CM (1986) Reproductive isolation in the *Lycoperiscon peruvianum* complex. In: D'arcy W (ed) Solanaceae: biology and systematics. Columbia Univ Press New York

Sachs MM, Dennis ES, Gerlach WL, Peacock WJ (1986) Two alleles of maize *Alcohol dehydrogenase-1* have 3′ structural and polyA addition polymorphisms. Genetics 113:449–467

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Segal E, Billyard E, So M, Storzbach M, Meyer TF (1985) Role of chromosomal rearrangement in *N. gonorrhoeae* pilus phase variation. Cell 40:293–298

Segal E, Hagblom P, Seifert HS, So M (1986) Antigenic variation of gonococcal pilus involves assembly of separated silent gene segments. Proc Natl Acad Sci USA 83:2177–2183

Tanksley SD, Bernatzky R, Lapitan NL, Prince JP (1988) Conservation of gene repertoire but not gene order in pepper and tomato. Proc Natl Acad Sci USA 85:6419–6423

Tautz D, Renz M (1983) An optimized freeze squeeze method for the recovery of DNA fragments from agarose gels. Anal Biochem 132:14–19

Wakeland EK, Boehme S, She JX (1990) The generation and maintenance of MHC class II gene polymorphism in rodents. Immunol Rev 113:207–226

Willimzig M (1985) LiCl-boiling method for plasmid mini-preps. Trends Genet 1:158

Wright S (1939) The distribution of self-sterility alleles in populations. Genetics 24:538–532