Polymorphism at the self-incompatibility locus in Solanaceae predates speciation

(S locus/molecular evolution/shared polymorphism)

Thomas R. Ioerger^{*}, Andrew G. Clark[†], and Teh-hui Kao^{*†‡}

*Department of Molecular and Cell Biology and [†]Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, PA 16802

Communicated by Michael T. Clegg, September 24, 1990 (received for review July 5, 1990)

ABSTRACT Sequences of 11 alleles of the gametophytic self-incompatibility locus (S locus) from three species of the Solanaceae family have recently been determined. Pairwise comparisons of these alleles reveal two unexpected observations: (i) amino acid sequence similarity can be as low as 40% within species and (ii) some interspecific similarities are higher than intraspecific similarities. The gene genealogy clearly illustrates this unusual pattern of relationships. The data suggest that some of the polymorphism at the S locus existed prior to the divergence of these species and has been maintained to the present. In support of this hypothesis, the number of shared polymorphic sites was found to exceed the number found in simulations with independent accumulation of mutations. Strictly neutral evolution is exceedingly unlikely to maintain the polymorphism for such a long time. The allele multiplicity and extreme age of the alleles is consistent with Wright's classic one-locus population genetic model of gametophytic self-incompatibility. Similarities between the plant S locus and the mammalian major histocompatibility complex are discussed.

Self-incompatibility is a natural, genetically determined barrier to inbreeding. In many plant species, self-incompatibility is controlled by a single locus, the S locus, which elicits recognition and rejection of self-pollen (1). In gametophytic self-incompatibility, inhibition of pollen germination or pollen-tube growth results when the S allele expressed by the pollen matches one of the two S alleles expressed in the pistil. A highly expressed, pistil-specific protein, the S protein, has been identified as a product of the S locus. The cDNA sequences of several alleles of the S protein from three species in the Solanaceae family have been determined, and extraordinary divergence, even between alleles from the same species, is apparent (T.R.I. and T.-h.K., unpublished work).

The self-incompatibility systems of many plant species have long been of interest to evolutionary biologists, and the extreme levels of polymorphism in S-protein sequences suggest that an unusual array of evolutionary forces operate at the S locus. Early appreciation of the allelic diversity of Oenothera organensis, with 37 alleles identified in a population of 1000 plants (2, 3), prompted Wright (4) to examine population genetic models of self-incompatibility loci. A consequence of gametophytic self-incompatibility is that all individuals in a population are heterozygotes. Wright noticed that rare alleles have a reproductive advantage because pollen bearing a rare allele is less likely to land on a stigma with the same allele, so that rare alleles suffer fewer aborted pollinations. This frequency dependence made it less likely for a new mutation to be rapidly lost compared with a neutral allele, and Wright's model produced a balance between mutation and random genetic drift that could maintain many alleles in the population. Additionally, this model predicts that the alleles could be far older than neutral alleles. The availability of molecular sequence data provides us with unprecedented power to test these models and allows us to explore the evolutionary past of the locus by reconstructing the genealogy of S alleles.

MATERIALS AND METHODS

Sequences and Alignment. The 11 inferred S-protein sequences examined in this study are S_{F11} , S_z , S_{1nic} , S_{2nic} , S_{3nic} , and S_{6nic} from Nicotiana alata (5, 6), an ornamental tobacco species; S_{1pet} , S_{2pet} , and S_{3pet} from Petunia inflata (7), a species of wild petunia; and S_{2xol} and S_{3xol} from Solanum chacoense (8), a wild potato species. The alignment of amino acid sequences was taken from unpublished work (T.R.I. and T.-h.K.). The percentage of amino acid identity among pairs of alleles was calculated from the aligned sequences.

Gene Genealogy. A distance matrix based on amino acid similarities was calculated from the amino acid alignment, and the neighbor-joining algorithm (9) was used to generate a gene genealogy. The aligned amino acid sequences were also used to generate maximum parsimony trees (10) whose robustness was tested by "bootstrapping."

Shared Polymorphism Test. A test of the common ancestry of polymorphisms was performed based on nucleotide or amino acid sequences. A site that has the same two or more nucleotides segregating in two species is defined as a "shared polymorphic site." Shared polymorphic sites can occur either by chance or by common ancestry, so the test hinges on whether the observed number of shared polymorphic sites exceeds the number of shared polymorphic sites that would occur by independent accumulation of polymorphisms. The null distribution of numbers of shared polymorphisms was generated with a computer by randomly drawing one of the observed sequences and introducing as many random polymorphisms as were observed in the real data. After these randomly mutated sequences were constructed, the number of shared polymorphic sites was tallied exactly as for the real data. If the number of observed shared polymorphisms is greater than some critical fraction of the distribution with random mutations, then the null hypothesis is rejected and the test indicates that there is an excess of shared polymorphism. The most reasonable conclusion in such a case is that the polymorphisms have common ancestry. The same approach was also applied to the amino acid sequences.

RESULTS

The matrix of amino acid similarities indicates an extraordinary degree of divergence among alleles within each species (Table 1). The degree of sequence divergence within species

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. \$1734 solely to indicate this fact.

[‡]To whom reprint requests should be addressed.

			Ν.	alata				P. inflata		S. cha	coense
	S _{FII}	S _z	Sinic	S _{2nic}	S _{3nic}	S _{6nic}	Sipet	S _{2pet}	S _{3pet}	S _{2sol}	S _{3sol}
S _{F11}	_										
S_z	59.7										
Sinic	44.7	45.0	—								
S _{2nic}	44.7	43.6	67.7								
S _{3nic}	44.8	45.4	68.9	63.3	_						
Sonic	43.1	45.1	61.0	60.0	69.9	—					
Sipet	53.5	62.4	42.5	39.5	43.3	42.6	_				
S _{2pet}	53.2	61.7	40.8	40.3	41.7	39.9	73.1	_			
S _{3pet}	52.5	66.3	41.5	41.0	42.4	41.1	73.5	80.1	—		
S _{2sol}	40.1	41.6	46.7	44.7	46.0	45.3	38.8	38.8	39.3	_	
S _{3sol}	42.4	44.6	50.0	48.5	49.0	47.5	47.0	43.8	42.6	41.5	

Table 1. Pairwise amino acid similarity

These values represent percent amino acid identity. Note the unusually low similarity, even between alleles from the same species. For example, S_z and S_{Inic} are only 45.0% similar. Also note that some alleles from the same species are even less similar than alleles from different species (e.g., 45.0% between S_z and S_{Inic} vs. 62.4% between S_z and S_{Iper}).

is on the same order as the sequence divergence between species. In fact, some alleles appear to be more similar to alleles in other species than to other alleles in the same species.

A phylogenetic tree based on neighbor-joining analysis (9) of the distance matrix calculated from the amino acid alignment reveals two major evolutionary branches (Fig. 1). One lineage contains S_{FII} and S_z of N. alata along with three alleles of P. inflata, and the other lineage includes S_{Inic} , S_{2nic} , S_{3nic} , and S_{6nic} of N. alata along with two alleles of S. chacoense. A genealogy of nearly identical topology was obtained by the parsimony method (10). The unrooted par-

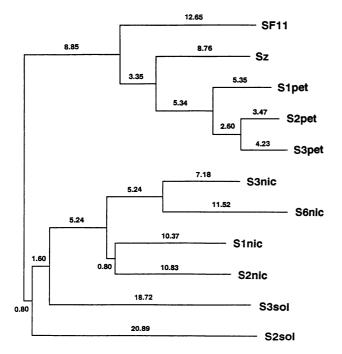


FIG. 1. S-locus gene genealogy. Note that the tree indicates that some alleles are more related to alleles in other species than to other alleles in the same species. For example, S_z and S_{FII} are clustered with the *P. inflata* alleles rather than with the other four *N. alata* alleles. Similarly, the other four *N. alata* alleles cluster with the two *S. chacoense* alleles rather than with S_z and S_{FII} . The tree was constructed from a distance matrix by the neighbor-joining algorithm (9), using software provided by J. H. Ferguson, N. Saitou, and M. Nei. Amino acid similarity values were taken from the 250 PAM table (11), appropriate for highly diverged sequences and summed over aligned residues. The distances, labeled as branch lengths, were calculated by subtracting each pairwise similarity value from 100.

simony tree shows that the smallest cluster with all six N. alata alleles necessarily contains the two S. chacoense alleles. The significance of this pattern was assessed by bootstrap resampling (12), which showed no cases out of 100 randomly generated trees in which all six N. alata alleles clustered together. Thus, the phylogenetic analyses indicate that some N. alata alleles are more closely related to alleles in other species than to the other N. alata alleles.

Based on the results of the phylogenetic analysis, we hypothesize that the extant polymorphism at the S locus arose prior to the time of divergence of the three solanaceous species. According to this hypothesis, one of the ancestral alleles was the most recent common ancestor of the P. inflata alleles as well as of S_{FII} and S_{z} from N. alata, and another was the most recent common ancestor of the two S. chacoense alleles and the remaining four N. alata alleles. Further support comes from a comparison of the average number of synonymous substitutions per site $[d_s;$ calculated by the method of Nei and Gojobori (13) and corrected for underestimation by the Jukes-Cantor formula (14)] between S alleles from the same species and between sequences of the small subunit of ribulose-1,5-bisphosphate carboxylase (RuBCs) (15-17) from each pair of these three genera. The average number of synonymous substitutions per site among S alleles from the same species exceeds that from interspecific comparisons of the RuBCs sequences (0.82 vs. 0.47, respectively). Under the assumption of equal rates of synonymous substitution, this result suggests that the S alleles within a species diverged before speciation. RuBCs has an estimated synonymous substitution rate of 6.6×10^{-9} substitutions per site per year (18), which is typical of plant nuclear genes (19). With this rate of synonymous substitution, *RuBCs* sequences data indicate a Petunia-Nicotiana split at 27 million years ago, a Nicotiana-Solanum split at 28 million years, and a Petunia-Solanum split at 36 million years, making the shared polymorphisms in the S locus extraordinarily old.

The shared polymorphism test provides a means of statistically testing whether common ancestry of polymorphisms is an appropriate conclusion to draw from these data. Table 2 gives a count of the number of shared polymorphic sites among all species pairs and the count of polymorphic sites common to all three species. In all cases, the observed number of shared polymorphic sites is significantly greater than the number that would be expected had the polymorphisms occurred independently. As Fig. 2 shows, in some cases none of the randomly generated sequences had as many shared polymorphisms as the observed number.

A more conservative test generates random mutations only at those sites that are observed to be polymorphic in the All three species

Table 2.	Significance tests of the numbers of shared	
nucleotide	e polymorphisms	

	Sample	Random pairs of alleles
No. of polymorphic sites		
N. alata	360	208
P. inflata	154	114
S. chacoense	235	235
No. of shared polymorphisms		
N. alata and P. inflata	75	16.6
N. alata and S. chacoense	102	35.5
P. inflata and S. chacoense	25	16.7
All three species	16	3.11
No. of randomly mutated sequen than the observed number of	U	
		•
N. alata and P. inflata	0/1000	0/1000
N. alata and S. chacoense	0/1000	0/1000
P. inflata and S. chacoense	5/1000	2/1000

This analysis used the alignment of the 11 nucleotide sequences without the leader peptide (a total of 618 base pairs). The first block of the table reports the number of polymorphic sites and the average number of differences between pairs of alleles within each species. The second block reports the number of polymorphisms shared among all observed alleles for each pair of species and for pairs of alleles drawn from each species. A site is considered to have a shared polymorphism if, among the alleles in the sample, two species share the same pair of distinct nucleotides. The third block reports the significance test ascertained by a computer simulation that generated random, independent mutations. One allele from each species was chosen and was randomly assigned as many mutations as were observed in the first block. The number of shared polymorphisms in this randomly generated sample was compared to the observed number. The third block of the table reports the number of such random samples that have more shared polymorphisms than did the actual data

0/1000

25/1000

sample. This test is appropriate if monomorphic sites are strongly constrained, and results in a larger number of shared polymorphic sites by chance. Even by this more conservative approach, all of the tests reported in Table 2 remain significant. Comparable results were obtained from an analysis of shared amino acid polymorphisms (Fig. 3). We conclude that the excess shared polymorphism is due to common ancestry and that the polymorphisms must have predated speciation.

DISCUSSION

The findings of the molecular analyses are consistent with predictions of Wright's classic one-locus population genetic model of gametophytic self-incompatibility (4). In this model, functionally new alleles generated by newly arisen mutations are initially favored over more frequent alleles. Opposing this spontaneous introduction of new variation is the loss of alleles by random genetic drift. The model attains an equilibrium in which many more alleles are maintained in a finite population than could be maintained at a strictly neutral locus. Recent theoretical analyses indicate that alleles at a gametophytic self-incompatibility locus will have very long residence times and expected coalescence times far exceeding those of neutral alleles (20, 21). The inevitable consequence of the extreme age of alleles is that they will have highly divergent sequences.

Explanations for the extraordinary allelic diversity that do not invoke the role of self-incompatibility fail to explain the observed pattern of interspecific shared polymorphism. Although a high mutation rate can also generate high levels of allelic diversity, it would not result in the large number of shared polymorphisms observed between alleles. The uniqueness of the S locus has been demonstrated by genomic

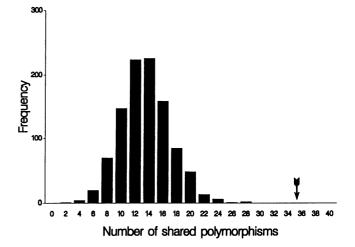


FIG. 2. The test of shared nucleotide-site polymorphism. The histogram indicates the distribution of number of shared nucleotide polymorphisms that are obtained from 1000 computer-generated samples for the tobacco-potato comparison. For each sample, one tobacco sequence and one potato sequence were drawn at random. Random pairs of tobacco sequences had an average of 208 polymorphic sites, and the potato sequences had an average of 235 polymorphic sites, so 208 mutations were introduced to the tobacco sequence and 235 mutations to the potato sequence. The number of shared polymorphic sites was then tallied from this sample, and the entire process was repeated 1000 times. The arrow indicates the average observed number of shared polymorphisms between tobacco and potato alleles. Because the observed number is greater than the randomly generated sample, we conclude that the shared polymorphisms are not due to chance coaccumulation of mutations, but rather that they represent common ancestry.

Southern blots of N. alata (6) and P. hybrida (22), which excludes the possibility of intergenic exchange as a mechanism for generating allelic diversity. Because there is only one gene, analysis of gametophytic S alleles is much simpler than that of the multigenic sporophytic self-incompatibility system of *Brassica* (23).

The clearest case of shared polymorphism would have alleles from each of two different species on each of two sides of a deep branch in the gene genealogy. The pattern observed here has only Nicotiana alleles on two sides of a deep branch. A tree constructed from linked shared polymorphic sites would have alleles from each species in both of the major clusters. However, hybridization and introgression or recombination/gene conversion (24) among these ancient alleles could have broken down associations among polymorphic sites, resulting in *P. inflata* alleles that cluster together while still exhibiting an excess of shared polymorphisms with N. alata. Genomic sequence data, which show extreme intraspecific sequence diversity for ≈ 5 kilobases downstream from the S locus, suggest that there is a mechanism for restricting intragenic recombination in this region (unpublished results). The extraordinary age of alleles and sequence divergence among alleles could only be found for a neutral locus in an implausibly large population.

Comparable evidence for ancient polymorphism that predates speciation has previously been found in the mammalian major histocompatibility complex (MHC) class I locus (24– 27). Both the MHC locus and the S locus are highly polymorphic and have highly divergent alleles. The extraordinary age of the polymorphism in both systems appears to have been maintained by natural selection through a mechanism involving self-/non-self-recognition. In the case of the human MHC, evidence suggests that either overdominance or frequency-dependent selection maintains the high allelic diversity (28). The operation of diversifying natural selection at the S locus should result in an excess of nonsynonymous sub-

	10 20 30 40 50
SF11	DFEYLQLVLTWPASFCYANH.CERIAPNNFTIHGLWPDNVKTRLHNC
Sz	DFDYHQLVLTWPASFCYPKNFCSRIAPKNFTIHGLWPDKVRGRLOFC
Sinic	NFEYMOLVLTWPTAFC.N VMN. CERT. PTNFTIHGLWPDNVSTELNYC
S2nic	AFEYMQLVLTWPITFC.R. IKH.CERT.PTNFTIHGLWPDHHTTMLNYC
S3nic	AFEYMOLVLOWPAAFC.HTTPSP.CKRI.PMNFTIHGLWPDWVSTMLNYC
S6nic	AFEYHQLVLQWPTAFC.HTTPCKNI.PSNFTIHGLWPDHVSTTLNFC
Slpet	NFEYLQLVLTWPASFCFRPKNICKRPA.ENFTIHGLWPEITGFRLEFC
S2pet	NFDYFQLVLTWPASFCY PKNFCKRK.SMNFTIHGLWPKMKHFRLEFC
S3pet	NFDYIQLVLTWPASFCYRPKNICRRI.PMNFTIHGLWPEKEHFRLEFC
S2sol	TFDYMKLVLQWPPMYC.RNKFCERI.PRNFTVHGLWPDNKKYLLNNC
S2sol S3sol	TFEHLQLVLTWPTSFC.HKERCIRS.SSNFTIHGLWPDNTSTRLNFC
55301	II MINER DIVISION MERCIRS . SOM TIMOLAT DAISIRLATO
	60 70 80 90 100
SF11	KPKPTYSYFT.GKMLHDLDKHWMQLKFEQDYGRTEQPSWKYQYIKHGSCC
Sz	TSEKYVNFAQDSPILDDLDHHWMELKYHRDFGLKNQFLWRGQYQKHGTCC
Sinic	DEQEKFELFEDDKEQEDLDDRWPDLTLDRDDCKEGQGFWSTEYKKHGTCC
S2nic	DRSKPYNMFTDGKKKNDLDERWPDLTKTKFDSLDKQAFWKDEYVKHGTCC
S3nic	SGEDEYEKLDDDKKKKDLDDRWPDLTIARADCIKHQVFWKHEYNKHGTCC
Sónic	GKEDDYNIIHDGPEKNGLYVRWPDLIREKADCHKTQNFWRREYIKHGTCC
61	TO DESCRIPTION TO THE DESCRIPTION OF THE STREET OF THE STR
Slpet S2pet	TGDFKTETFKDNNIVDYLERHWVQMKFDENYAKYHQPLWSYEYRKHGMCC TGDK.TSRFKEDNIIWVLERHWIOMRFDEKYASTKOPLWEHEYNRHGICC
S2pet	DGDKFVSFSLKDRIVNDLERHWVQMKFDEKFAKIKQPLWEHEINKHGICS
Sober	Dodar vərəladili vadlerin vçirri derfirinde in the tarigitus
S2sol	. RSYAYNALTNVREQSKLDDRWPDLTSNKSMTHKEQKFWEYEYNKHGTCC
S3sol	. KIVKYNKIEDEHKIDALEYGWPNLTTTEAVSKEDQVFWGKQYTKHGSCC
0011	
SF11	QERYNQNTYFGLALRLKDKFDLLETLQTHRIIPGSSY. TFQDIFDAIKTV
Sz	QERYBONTYFGLALRLKDKFDLLETLOTHRIIPGSSY. TFODIFDAIKTV IPRYBOMOYFLLAMRLKDKFDLLATLRTHGITPGTKH, TFNETRDAIKTV
Sz Slnic	QERYNQNTYFGIALLLKNKFDLLETLOTHRIIFGSSY. TFQDIFDAIKTV IFRYNQMQYFLLAMRLKDKFDLLATLATHGITFGTKH. TFNETRDAIKTV LFSYNQEQYFDLAMALKOKFDLLESFRNHGITFTSY. TVOKYNNTVKAI
Sz	QERYHQNTYFGLALRIKDKFDLLETLQTHRI I PGSSY. TFQDI FDAIKTV I PRYHQMYFFLAMRIKDKFDLLATIRHGI I PGSSY. TFNETRDAIKTV LPSYHQEQYFDLAMALKDKFDLLSSFRNHGI I PTESY. TVQEYNNTVKAI SDKFDRQYFDLAMTIRDKFDLLSSIRNHGI SRGFSY. TVQHLNNTIKAI
Sz Slnic S2nic	QERYNQNTYFGIALLLKNKFDLLETLOTHRIIFGSSY. TFQDIFDAIKTV IFRYNQMQYFLLAMRLKDKFDLLATLATHGITFGTKH. TFNETRDAIKTV LFSYNQEQYFDLAMALKOKFDLLESFRNHGITFTSY. TVOKYNNTVKAI
Sz Slnic S2nic S3nic	QERYHQNTYFGLALRIKDKFDLLETLQTHRIIFGSSY.TFQDIFDAIKTV IFXYHQMYYFLLAMRIKDKFDLLATIRTHGITFGTKH.TFNETBDAIKTV LFSYHQEQYFDLAMALKDKFDLLSSFRHGIIFTKSY.TVQKYNTYKAI SDKFDREQYFDLAMALKDKFDLLSSLRNHGISRGFSY.TVQKINTIKAI SESYHLTQYFDLAMALKDKFDLLTSLRKHGIIFGNSY.TVQKINNTIKAI SEIYHQVQYFRLAMALKDKFDLLTSLRKHGIIRGYKY.TVQKINNTIKTV
Sz Slnic S2nic S3nic S6nic S1pet	QERYMQNTYFGIALLIKDKFDLLETLQTHRI IFGSSY. TFQDIFDAIKTV IFRYMQMQYFLIAMRLKDKFDLLETLQTHRI IFGSSY. TFQDIFDAIKTV LFSYMQRQYFDLAMALKDKFDLLESFRNHGIFTESSY. TVQETNNTVKAI SDKFDREQYFDLAMILRDKFDLLESFRNHGIFTESSY. TVQETNNTVKAI SESYMLTQYFDLAMALKDKFDLLSSLRNHGIFGENSY. TVQETNNTIKAI SEIYMQVQYFRLAMALKDKFDLLTSLKNHGIIRGYKY. TVQETNNTIKTV SKIYMQKAYFLLATRLKEKFDLLTTLRTHGITPGTKH. TFGDIQKAIKTV
Sz Slnic S2nic S3nic S6nic S1pet S2pet	QERYNQNTYFGIALBLENKFDLLETLQTHRIIFGSSY.TFQDIFDAIKTV IFRYNDMQYFLAMRLENKFDLLETLQTHRIIFGSSY.TFQDIFDAIKTV LFSYNDRQYFDLAMALKNKFDLLESFRNHGIIFTKSY.TVQKINNTVKAI SDKFDREQYFDLAMALKNKFDLLSSIRNHGISRGFSY.TVQKINNTIKAI SEYNDLQYFDLAMALKNKFDLLTSLRKHGIIFGSY.TVQKINNTIKTV SEYNDVQYFRLAMALKNKFDLLTSLKNHGIIFGTKY.TVQKINNTIKTV SKIYNQKAYFLLATRLKEKFDLLTTLRTHGITFGTKH.TFGDIQKAIKTV KNLYDQKAYFLLATRLKNKLDLLTTLRTHGITFGTKH.TFGDIQKAIKTV
Sz Slnic S2nic S3nic S6nic S1pet	QERYMQNTYFGIALLIKDKFDLLETLQTHRI IFGSSY. TFQDIFDAIKTV IFRYMQMQYFLIAMRLKDKFDLLETLQTHRI IFGSSY. TFQDIFDAIKTV LFSYMQRQYFDLAMALKDKFDLLESFRNHGIFTESSY. TVQETNNTVKAI SDKFDREQYFDLAMILRDKFDLLESFRNHGIFTESSY. TVQETNNTVKAI SESYMLTQYFDLAMALKDKFDLLSSLRNHGIFGENSY. TVQETNNTIKAI SEIYMQVQYFRLAMALKDKFDLLTSLKNHGIIRGYKY. TVQETNNTIKTV SKIYMQKAYFLLATRLKEKFDLLTTLRTHGITPGTKH. TFGDIQKAIKTV
Sz Slnic S2nic S3nic S6nic S1pet S2pet S3pet	QERYMQNTYFGIALELKDKFDLLETLOTHRI IFGSSV. TFQDIFDAIKTV IFRYMQNQYFLIAMRLKDKFDLLETLOTHRI IFGSSV. TFQDIFDAIKTV IFSYMQNQYFDLAMALKDKFDLLESFRNHGIFTESSY. TVQETNNTVKAI SDKFDREQYFDLAMILKDKFDLLESFRNHGIFTESSY. TVQETNNTIKAI SESYMLTYFDIAMALKDKFDLLTSIRKHGIFGENSY. TVQETNNTIKAI SEIYMQVQYFRLAMALKDKFDLLTSIRKHGIIRGTNY. TVQETNNTIKAV SKIYMQKAYFLLATRLKEKFDLLTTIRTHGITPGTKH. TFGDIQKAIKTV KNLYDQEAYFLLATRLKDKLDLLTTIRTHGITPGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTIRTHGITPGTKH. TFGEIQKAIKTV
Sz Slnic S2nic S3nic S6nic S1pet S2pet S3pet S2sol	QERYMQNTYFGLALRLKDKFDLLETLQTHRIIFGSSY.TFQDIFDAIKTV IFRYNOMQYFLAMRLKDKFDLLETLQTHRIIFGSSY.TFQDIFDAIKTV IFSYNOMQYFLAMRLKDKFDLLESFRNHGIIFTKSY.TVQEYNNTVKAI SDKFDREQYFDLAMALKDKFDLLSSLRNHGISRGFSY.TVQEINNTIKAI SEYMLTQYFDLAMALKDKFDLLTSLRKHGIIFGSTY.TVQEINNTIKAI SEYMQVQYFRLAMALKDKFDLLTSLRKHGIIFGTKY.TVQEINNTIKTV SKIYMQKAYFLLATRLKEKFDLLTTLRTHGITFGTKH.TFGDIQKAIKTV KNLYDQEAYFLLATRLKDKHDLLTTLRTHGITFGTKH.TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTLRTHGITFGTKH.TFGEIQKAIKTV
Sz Slnic S2nic S3nic S6nic S1pet S2pet S3pet	QERYMQNTYFGIALELKDKFDLLETLOTHRI IFGSSV. TFQDIFDAIKTV IFRYMQNQYFLIAMRLKDKFDLLETLOTHRI IFGSSV. TFQDIFDAIKTV IFSYMQNQYFDLAMALKDKFDLLESFRNHGIFTESSY. TVQETNNTVKAI SDKFDREQYFDLAMILKDKFDLLESFRNHGIFTESSY. TVQETNNTIKAI SESYMLTYFDIAMALKDKFDLLTSIRKHGIFGENSY. TVQETNNTIKAI SEIYMQVQYFRLAMALKDKFDLLTSIRKHGIIRGTNY. TVQETNNTIKAV SKIYMQKAYFLLATRLKEKFDLLTTIRTHGITPGTKH. TFGDIQKAIKTV KNLYDQEAYFLLATRLKDKLDLLTTIRTHGITPGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTIRTHGITPGTKH. TFGEIQKAIKTV
Sz Slnic S2nic S3nic S6nic S1pet S2pet S3pet S2sol	QERYMQNTYFGLALRLKDKFDLLETLQTHRIIFGSSY.TFQDIFDAIKTV IFRYNOMQYFLAMRLKDKFDLLETLQTHRIIFGSSY.TFQDIFDAIKTV IFSYNOMQYFLAMRLKDKFDLLESFRNHGIIFTKSY.TVQEYNNTVKAI SDKFDREQYFDLAMALKDKFDLLSSLRNHGISRGFSY.TVQEINNTIKAI SEYMLTQYFDLAMALKDKFDLLTSLRKHGIIFGSTY.TVQEINNTIKAI SEYMQVQYFRLAMALKDKFDLLTSLRKHGIIFGTKY.TVQEINNTIKTV SKIYMQKAYFLLATRLKEKFDLLTTLRTHGITFGTKH.TFGDIQKAIKTV KNLYDQEAYFLLATRLKDKHDLLTTLRTHGITFGTKH.TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTLRTHGITFGTKH.TFGEIQKAIKTV
Sz Slnic S2nic S3nic S6nic S1pet S2pet S3pet S2sol	QERYMQNTYFGLALRLKDKFDLLETLQTHRIIFGSSY.TFQDIFDAIKTV IFRYNOMQYFLAMRLKDKFDLLETLQTHRIIFGSSY.TFQDIFDAIKTV IFSYNOMQYFLAMRLKDKFDLLESFRNHGIIFTKSY.TVQEYNNTVKAI SDKFDREQYFDLAMALKDKFDLLSSLRNHGISRGFSY.TVQEINNTIKAI SEYMLTQYFDLAMALKDKFDLLTSLRKHGIIFGSTY.TVQEINNTIKAI SEYMQVQYFRLAMALKDKFDLLTSLRKHGIIFGTKY.TVQEINNTIKTV SKIYMQKAYFLLATRLKEKFDLLTTLRTHGITFGTKH.TFGDIQKAIKTV KNLYDQEAYFLLATRLKDKHDLLTTLRTHGITFGTKH.TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTLRTHGITFGTKH.TFGEIQKAIKTV
Sz Slnic S2nic S3nic S6nic S1pet S2pet S3pet S2sol S3sol SF11	QERYMQNTYFGIALELKDKFDLLETLJTHRIIFGSS. TFQDIFDAILTV IFRYMQNQYFLAMRLKDKFDLLATLRTHGITFGTKH. TFNETKDAIKTV LFSYMQRQYFDLAMALKDKFDLLSSTRINGIJFTSY. TVQKINNTVKAI SDKFDREQYFDLAMALKDKFDLLSSTRINGIJRGFSY. TVQKINNTIKAI SEXYMLQYFDLAMALKDKFDLLTSLKNIGIJRGTSY. TVQKINNTIKAI SEXYMLQYFRLAMALKDKFDLLTSLKNIGIJRGTKY. TVQKINNTIKTV SKIYMQKAYFLLATRLKEKFDLLTTLRTHGITFGTKH. TFGDIQKAIKTV KNLYDQRAYFLLATRLKEKFDLLTTLRTHGITFGTKH. TFGDIQKAIKTV SNLYDQRAYFLLATRLKEKFDLLTTLRTHGITFGTKH. TFGDIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTLRTHGITFGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTILRTHGITFGTKH. TFGEIQKAIKTV EKLYMQAQYFNLTMNLKDKFDLLEILRNHGIVFGSLA. LLSNSGRPLRQL TDLYDKDAYFDLAMNLKDRFDLLKILAMHGITFGTSHHTSSNIQNAVKSV 160 170 180 190 200 SQE.NPDIKCAEV. TKGTFELYEIGICFTFNADSMFRCQSDTCDKT. AKVLFRR
Sz Slnic S2nic S3nic S6nic S1pet S2pet S3pet S2sol S3sol SF11 Sz	QERYMQNTYFGLALLIKDKFDLLETLQTHRIJFGSSV. TFQDIFDAIKTV IFRYMQMQYFLAMRLKDKFDLLETLQTHRIJFGSSV. TFQDIFDAIKTV LFSYMQRQYFDLAMALKDKFDLLESTRNHGIJFTKSY. TVQETNNTVKAI SDKFDRQYFDLAMLKDKFDLLESTRNHGIJFTKSY. TVQETNNTVKAI SESYMLAUXALKDKFDLLESTRNHGIJFGSSY. TVQETNNTIKAI SESYMLAUXQYFLAMALKDKFDLLTSIKNHGIJFGSSY. TVQETNNTIKAI SEIYMQVQYFRLAMALKDKFDLLTSIKNHGIJFGTKH. TFGDIQKAIKTV KNLYDQEAYFLLATRLKDKFDLLTTIRTHGITFGTKH. TFGDIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTILTHRGITFGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTILTHRGITFGTSH. TFGEIQKAIKTV 100 170 180 190 200 SQE.NPDIEGAEV. TKGTPELYEIGICFTFNADSMFRCPQSDTCDKTAKVLFRR
Sz Slnic S2nic S3nic S5nic S1pet S2pet S3pet S2sol S3sol SF11 Sz S1nic	QERYMQNTYFGIALELKDKFDLLETLOTHRIIFGSS. TFODIFDAIKTV IFRYMQNQYFLIAMRLKDKFDLLETLOTHRIIFGSS. TFODIFDAIKTV IFSYMQNQYFLIAMRLKDKFDLLESTRINGIIFTSS. TVQETNNTVKAI SDKFDREQYFDLAMILKDKFDLLSSIRINGISRGFSY. TVQETNNTVKAI SESYMLTYFDIAMILKDKFDLLSSIRINGISRGFSY. TVQETNNTIKAI SEIYMQQYFRLAMALKDKFDLLTSIRKHGIIFGTSY. TVQETNNTIKAI SEIYMQQYFRLAMALKDKFDLLTSIRKHGIIFGTSY. TVQETNNTIKTV SKIYMQKAYFLLATRLKEKFDLLTTIRTHGITFGTKH. TFGDIQKAIKTV KNLYDQEAYFLLATRLKEKFDLLTTIRTHGITFGTKH. TFGDIQKAIKTV SNLYDQAYFLLARRVKDKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQAYFLLARRVKDKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQAYFLLARRVKDKFDLLTTIRTHGITFGTSH. TFGEIQKAIKTV 160 170 180 190 200 SQE.NPDIECAEV.TKGTPELYEIGICFTPTADSFFCCPSDTCDKT. AKVLFRR TMGVDPDL&CVEH. IKGYRELYEIGICFTPTADSFFCCPSDTCDETGITKILFRR
Sz Slnic S2nic S3nic S5nic S1pet S2pet S3pet S2sol S3sol SF11 Sz Slnic S2nic	QERYMQNTYFGLALLLKUKFDLLETLOTHRIJFGSSY.TFQDIFDAIKTV IPRYMQNQYFLLAMRLKDKFDLLETLOTHRIJFGSSY.TFQDIFDAIKTV IPSYMQRQYFDLAMALKDKFDLLESTRINGIJFTSSY.TVQETNNTVKAI SDKFDREQYFDLAMALKDKFDLLSSIRNHGISRGFSY.TVQETNNTVKAI SDKFDREQYFDLAMALKDKFDLLSSIRNHGISRGFSY.TVQETNNTIKAI SEYMQVAYFLAMALKDKFDLLTSIRKHGIJFGTKY.TVQETNNTIKTV SKIYMQVAYFLAMALKDKFDLLTSIRKHGIJFGTKH.TFGDIQKAIKTV KNLYDQEAYFLLATRLKEKFDLLTTIRTHGITFGTKH.TFGEIQKAIKTV SNLYDQEAYFLLATRLKEKFDLLTTIRTHGITFGTKH.TFGEIQKAIKTV SNLYDQAYFLLAMRVKDKFDLLITIRTHGITFGTKH.TFGEIQKAIKTV SNLYDQAYFLLAMRVKDKFDLLITIRTHGITFGTKH.TFGEIQKAIKTV SNLYDQAYFLLAMRVKDKFDLLITIRTHGITFGTKH.TFGEIQKAIKTV EELYMQAQYFNLTMNLKDKFDLLITIRTHGITFGTKH.TFGEIQKAIKTV 160 170 180 190 200 SQE.NPDIKCAEV.TKGTPELYEIGICFTFTADSHFRCPQSDTCDKT.AKVLFRR TMGVDPDLKCUEH.KGVRELYEIGICFTFTADSHFRCPQSDTCDETCITHILFRR TMGVDPDLKCAEV.TKGTPELYEIGICFTFTADSHFRCPQSDTCDETCITHILFRR TMGVDPDLKCAEV.TKGTPELYEIGICFTPTADSFFPCFNSWTCDETCITHILFRR TMGVDPDLKCAEV.MEUQEIGICFDFVKWWIDCPRPKTCKATRN.GITFF.
Sz Slnic S2nic S3nic S3nic S1pet S2pet S3pet S2sol	QERYMQNTYFGIALLLKDKFDLLETLQTHRIIFGSSV. TFQDIFDAIKTV IFRYMQNQYFLIAMRLKDKFDLLETLQTHRIIFGSSV. TFQDIFDAIKTV IFRYMQNQYFLIAMRLKDKFDLLESTRNHGIFTKSY. TVQETNNTVKAI SDKFDREQYFDLAMILKDKFDLLESTRNHGIFTKSY. TVQETNNTVKAI SESYMLAUXPFLIAMALKDKFDLLSSTRNHGIFGSSY. TVQETNNTIKAI SESYMLAUXPFLIAMALKDKFDLLTSILKNHGIIRGTSY. TVQETNNTIKAI SEIYMQVQYFRIAMALKDKFDLLTSILKNHGIIRGTSY. TVQETNNTIKAI SEIYMQVQYFRIAMALKDKFDLLTSILKNHGIIRGTSY. TVQETNNTIKAI SEIYMQVQYFRIAMALKDKFDLLTTIRTHGITPGTKH. TFGEIQKAIKTV KNLYDQEAYFLLATRLKEKFDLLTTIRTHGITPGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTIRTHGITPGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTIRTHGITPGTSH. TFGEIQKAIKTV 160 170 180 190 200 SQE.NPDIKCAEV. TKGTPELYEIGICFTFNADSHFRCPQSDTCDKTAKVLFRR TMCVDPDLKCVEH. IKGGVELYEIGICFTFNADSHFRCPQSDTCDKTAKVLFRR TMCVDPDLKCVEM. IKGGVELYEIGICFTPTADSFPCCFNBUTDCTCITKILFRF TKG. FPNLTCNKQ MELQEIGICFDQKVKWVIDCPNPETCKATRN.GITFP. TGG. FNLTSCRL MELQEIGICFDQKVKWVDCPNPETCKATRN.GITFP.
Sz Slnic S2nic S3nic S5nic S1pet S2pet S3pet S2sol S3sol SF11 Sz Slnic S2nic	QERYMQNTYFGLALLLKUKFDLLETLOTHRIJFGSSY.TFQDIFDAIKTV IPRYMQNQYFLLAMRLKDKFDLLETLOTHRIJFGSSY.TFQDIFDAIKTV IPSYMQRQYFDLAMALKDKFDLLESTRINGIJFTSSY.TVQETNNTVKAI SDKFDREQYFDLAMALKDKFDLLSSIRNHGISRGFSY.TVQETNNTVKAI SDKFDREQYFDLAMALKDKFDLLSSIRNHGISRGFSY.TVQETNNTIKAI SEYMQVAYFLAMALKDKFDLLTSIRKHGIJFGTKY.TVQETNNTIKTV SKIYMQVAYFLAMALKDKFDLLTSIRKHGIJFGTKH.TFGDIQKAIKTV KNLYDQEAYFLLATRLKEKFDLLTTIRTHGITFGTKH.TFGEIQKAIKTV SNLYDQEAYFLLATRLKEKFDLLTTIRTHGITFGTKH.TFGEIQKAIKTV SNLYDQAYFLLAMRVKDKFDLLITIRTHGITFGTKH.TFGEIQKAIKTV SNLYDQAYFLLAMRVKDKFDLLITIRTHGITFGTKH.TFGEIQKAIKTV SNLYDQAYFLLAMRVKDKFDLLITIRTHGITFGTKH.TFGEIQKAIKTV EELYMQAQYFNLTMNLKDKFDLLITIRTHGITFGTKH.TFGEIQKAIKTV 160 170 180 190 200 SQE.NPDIKCAEV.TKGTPELYEIGICFTFTADSHFRCPQSDTCDKT.AKVLFRR TMGVDPDLKCUEH.KGVRELYEIGICFTFTADSHFRCPQSDTCDETCITHILFRR TMGVDPDLKCAEV.TKGTPELYEIGICFTFTADSHFRCPQSDTCDETCITHILFRR TMGVDPDLKCAEV.TKGTPELYEIGICFTPTADSFFPCFNSWTCDETCITHILFRR TMGVDPDLKCAEV.MEUQEIGICFDFVKWWIDCPRPKTCKATRN.GITFF.
Sz Slnic S2nic S3nic S3nic S1pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pic S3pic S2pic S2pic S3pic S2pic S2pic S3pic S2pet S2pet	QERYMQNTYFGLALLLKUKFDLLETLQTHRIIFGSSY.TFQDIFDAIKTV IPRYMQNQYFLLAMRLKDKFDLLETLQTHRIIFGSSY.TFQDIFDAIKTV IPSYMQRQYFDLAMALKDKFDLLESTRINGIIFTKSY.TVQKINNTYKAI SDKFDREQYFDLAMALKDKFDLLSSIRINGISRGFSY.TVQKINNTIKAI SDKFDREQYFDLAMALKDKFDLLSSIRINGISRGFSY.TVQKINNTIKAI SEYMLQYFPLAMALKDKFDLLTSIRKHGIIFGTKY.TVQKINNTIKTV SKIYMQKAYFLLATRLKEKFDLLTTIRTHGITFGTKH.TFGDIQKAIKTV KNLYDQKAYFLLATRLKEKFDLLTTIRTHGITFGTKH.TFGEIQKAIKTV SNLYDQRAYFLLATRLKEKFDLLTTIRTHGITFGTKH.TFGEIQKAIKTV SNLYDQRAYFLLATRLKEKFDLLTIRTHGITFGTKH.TFGEIQKAIKTV SNLYDQRAYFLLANRVKDKFDLLITIRTHGITFGTKH.TFGEIQKAIKTV EELYMQAQYFNLTMNLKDKFDLLITIRTHGITFGTKH.TFGEIQKAIKTV 160 170 180 190 200 SQE.NPDIKCAEV.TKGTPELYEIGICFTPTADSHFRCPSDTCDKT.AKVLFRR TMQVDPDL&CVEH.IKGVRELYEIGICFTPTADSFFQCPHSWTCDETGITKILFRR TKG.FNLTCSRLRELKEIGICFDQKVKWVIDCFPHFTCKKTN.GITFF. TGG.FYNLTCSRLRELLEIGICFDSKVKWVIDCFHFTCKKPHGNBGIKFP. TKG.YPNLSCTKRQMELLEIGICFDSKVKWVIDCPHFTCKKASNQGIMFF.
Sz Slnic S2nic S3nic S3nic S1pet S2pet S2pet S2sol S3sol S2sol	QERYMQNTYFGIALLLKDKFDLLETLQTHRIIFGSSV. TFQDIFDAIKTV IFRYMQNQYFLIAMRLKDKFDLLETLQTHRIIFGSSV. TFQDIFDAIKTV IFRYMQNQYFLIAMRLKDKFDLLESTRNHGIFTKSY. TVQETNNTVKAI SDKFDREQYFDLAMILKDKFDLLESTRNHGIFTKSY. TVQETNNTVKAI SESYMLAUXPFLIAMALKDKFDLLSSTRNHGIFGSSY. TVQETNNTIKAI SESYMLAUXPFLIAMALKDKFDLLTSILKNHGIIRGTSY. TVQETNNTIKAI SEIYMQVQYFRIAMALKDKFDLLTSILKNHGIIRGTSY. TVQETNNTIKAI SEIYMQVQYFRIAMALKDKFDLLTSILKNHGIIRGTSY. TVQETNNTIKAI SEIYMQVQYFRIAMALKDKFDLLTTIRTHGITPGTKH. TFGEIQKAIKTV KNLYDQEAYFLLATRLKEKFDLLTTIRTHGITPGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTIRTHGITPGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTIRTHGITPGTSH. TFGEIQKAIKTV 160 170 180 190 200 SQE.NPDIKCAEV. TKGTPELYEIGICFTFNADSHFRCPQSDTCDKTAKVLFRR TMCVDPDLKCVEH. IKGGVELYEIGICFTFNADSHFRCPQSDTCDKTAKVLFRR TMCVDPDLKCVEM. IKGGVELYEIGICFTPTADSFPCCFNBUTDCTCITKILFRF TKG. FPNLTCNKQ MELQEIGICFDQKVKWVIDCPNPETCKATRN.GITFP. TGG. FNLTSCRL MELQEIGICFDQKVKWVDCPNPETCKATRN.GITFP.
Sz Slnic S2nic S3nic S3nic S1pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pet S2pic S3pic S2pic S2pic S3pic S2pic S2pic S3pic S2pet S2pet	QERYMQNTYFGIALLLKDKFDLLETLJTHRIIFGSSV. TFQDIFDAIKTV IFRYMQNQYFLIAMRLKDKFDLLETLJTHRIIFGSSV. TFQDIFDAIKTV IFRYMQNQYFLIAMRLKDKFDLLESTRINGIFTKSY. TVQETNNTVKAI SDKFDREQYFDLAMILKDKFDLLESTRINGIFTKSY. TVQETNNTVKAI SDKYDREQYFDLAMILKDKFDLLSSTRINGIFTKSY. TVQETNNTIKAI SESYMLAUKVFDLLTSIRKHGIFGSNY. TVQETNNTIKAI SESYMLAUKVFDLLTSIRKHGIFGSNY. TVQETNNTIKAI SEIYMQAYFLIATRLKEKFDLLTTIRTHGITFGTKH. TFGDIQKAIKTV KNLYDQEAYFLIATRLKEKFDLLTTIRTHGITFGTKH. TFGDIQKAIKTV SNLYDQRAYFLIATRLKEKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQRAYFLIAMRVKDKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQRAYFLIAMRVKDKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQRAYFLIAMRVKDKFDLLTIRTHGITFGTSH. TFGEIQKAIKTV SNLYDQRAYFLIAMRVKDKFDLLTIRTHGITFGTSH. TFGEIQKAIKTV SNLYDQRAYFLIAMRVKDKFDLLEILRINGIVFGSLA. LLSNSGRPLRQL TDLYDKDAYFDLAMNLKDKFDLLKILAMHGITFGTSHTSSNIQNAVKSV 160 170 180 190 200 SQE.NPDIKCAEV. TKGTPELYEIGICFTFTADSHFRCPQSDTCDKTAKVLFRR TMGVDPDLKCVEH. IKGVRELYEIGICFTPTADSFPCPHSTCKTMT.GITFF. TGG. FFNLTCNKQMELLEIGICFDEVKWIDCPMPETCKKTN.GITFF. TGG. FFNLSCTKRQMELKEIGICFDETVKWIDCPMPETCKATASNQGIMFF. TKG.YPNLSCTKRGQELWEVGIGFDSTAKWVIDCPMPETCKTASNQGIMFF. TNQVDPDLKCVEH. IKGVQELNEIGICFTPAADSFYCCHSTCDETGISTILFR
Sz Slnic S2nic S3nic S1pet S2pet S2pet S2pet S3sol S5sol S1nic S2nic S1nic S5nic S5nic S1pet S2pet S3pet	QERYMQNTYFGIALELKDKFDLLETLOTHRIJFGSSV. TFODIFDAIKTV IFRYMQNQYFLAMRLKDKFDLLETLOTHRIJFGSSV. TFODIFDAIKTV IFRYMQNQYFLAMRLKDKFDLLESTRINGGIFTESSY. TVOETNNTYKAI SDKFDREQYFDLAMALKDKFDLLSSTRINGGIFTESSY. TVOETNNTYKAI SDKFDREQYFDLAMALKDKFDLLSSTRINGGIRGTSY. TVOETNNTIKAI SESYMLTYFDLAMALKDKFDLLTSIRKHGIFGTSY. TVOETNNTIKAI SEIYMQQYFRLAMALKDKFDLLTSIRKHGIFGTSY. TVOETNNTIKAI SEIYMQQYFRLAMALKDKFDLLTSIRKHGIFGTSY. TFOEIQKAIKTV KNLYDQEAYFLLATRLKEKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQAYFLLATRVKDKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQAYFLLATRVKDKFDLLTTIRTHGITFGTSH. TFGEIQKAIKTV SNLYDQAYFLLATRVKDKFDLLTILTINGITFGTSH. TFGEIQKAIKTV SNLYDQAYFLLATRVKDKFDLLTILTINGITFGTSH. TFGEIQKAIKTV SNLYDQAYFLLATRVKDKFDLLEILENINGIVFGSLA. LLSNSGPIRQL TDLYDKDAYFDLANNLKDRFDLLKILAMHGITFGTSHTSSNIQNAVKSV 160 170 180 190 200 SQE.NPDIECAEV. TKGTPELYEIGICFTFTADSFFQCPSDTCDKT. AKVLFRR TMG, FFNLTCNKQ MELGEIGICFDYKNWIDCPNFETCKATR.GITFT. TGG, FFNLTCSRL RELKEIGICFDETVKNWIDCPNFETCKAFT. NEGVMFF. TGG, FFNLTCSRL RELKEIGICFDETVKNWIDCPNFETCKAFT.NGGIFFF. TGG, FFNLTSCTKRQ MELGEIGICFDSKVKNVIDCPNFETCKAFT.NGGIFFF. TGG, FFNLTSCTKRQ
Sz Slnic S2nic S2nic S2nic S2pet S2pet S2pet S2sol S3sol SF11 Sz S1nic S2nic S2nic S2nic S1nic S2nic S2nic S2nic S2nic S2nic S2nic S2nic	QERYMQNTYFGLALLLKUKFDLLETLGTHRIIFGSSY. TFQDIFDAILTV IFRYMQNQYFLLAMRLKDKFDLLATLRTHGITFGSSY. TFQDIFDAILTV IFYNDONQYFLAMALKDKFDLLASTRHGIFJTKSY. TVQKINNTYKAI SDKFDREQYFDLAMALKDKFDLLSSTRHGIFJTKSY. TVQKINNTYKAI SDKFDREQYFDLAMALKDKFDLLSSTRHGIFGTSY. TVQKINNTYKAI SEYMLQYFFLAMALKDKFDLLTSIRKHGIFGTSY. TVQKINNTYKAI SEYMLQYFFLAMALKDKFDLLTSIRKHGIFGTKH. TFGDIQKAIKTV KNLYDQEAYFLLATRLKEKFDLLTTIRTHGITFGTKH. TFGDIQKAIKTV SNLYDQRAYFLLATRLKEKFDLLTTIRTHGITFGTKH. TFGDIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTILRTHGITFGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLTILRTHGITFGTKH. TFGEIQKAIKTV SNLYDQRAYFLLAMRVKDKFDLLEILRNHGIVFGSLA. LLSNSGRPLRQL TDLYDKDAYFDLAMNLKDRFDLLKILAMHGITFGTSHHTSSNIQNAVKSV 160 170 180 190 200 SQE.NPDIKCAEV.TKGTPELYEIGICFTFTADSHFRCFQSDTCDKT. AKVLFRR TMQVDPDLKCVEH. IKGVRELYEIGICFTPTADSHFRCFQSDTCDKT. AKVLFRR TMGVPFDLKCKGH. MELQEIGICFDSKVEWIDCFNPHTCKATEN.GITFP. TGG.FPNLTCSRL. MELLEIGICFDSKVEWIDCFNPHTCKATEN.GITFP. TGG.YPNLSCTKRG. MELLEIGICFDSTAKEVIDCFNPHTCKATSNQGIMFP. TNQVDPDLKCVEH. IKGVRELNEIGICFNPAADSHFDCRNSTCDETDSTYTLFRF TNNKDPDLKCVEH. IKGVKELNEIGICFNPAADSFHDCRNSTCDETDSTTLFRF TNNKDPDLKCVEH. IKGVKELNEIGICFNPAADSFHDCRNSTCDETDSTTTLFFR TNNKDPDLKCVEH. IKGVKELKEVGICFTPAADSFHDCRNSTCDETDSTTLFFRF TNNKDPDLKCVEH. IKGVKELKEVGICFTPAADSFHDCRNSTCDETDSTTLFFRF TNNKDPDLKCVEH. IKGVKELKEVGICFTPAADSFHDCRNSTCDETDSTTLFFRF TNNKDPDLKCVEH. IKGVKELKEVGICFTPAADSFHDCRNSTCDETDSTTTTFFT TNNKDPDLKCVEH. IKGVKELKEVGICFTPAADSFHDCRNSTCDETDSTTTTFFT TNNKDPDLKCVEH. IKGVKELKEVGICFTPAADSFHDCRNSTCDETDSTTTTFFT TNNKDPDLKCVEH. IKGVKELKEVGICFTPAADSFHDCRNSTCDETDSTTTTIFTT
Sz Slnic S2nic S3nic S1pet S2pet S2pet S2pet S3sol S5sol S1nic S2nic S1nic S5nic S5nic S1pet S2pet S3pet	QERYMQNTYFGIALELKDKFDLLETLOTHRIJFGSSV. TFODIFDAIKTV IFRYMQNQYFLAMRLKDKFDLLETLOTHRIJFGSSV. TFODIFDAIKTV IFRYMQNQYFLAMRLKDKFDLLESTRINGGIFTESSY. TVOETNNTYKAI SDKFDREQYFDLAMALKDKFDLLSSTRINGGIFTESSY. TVOETNNTYKAI SDKFDREQYFDLAMALKDKFDLLSSTRINGGIRGTSY. TVOETNNTIKAI SESYMLTYFDLAMALKDKFDLLTSIRKHGIFGTSY. TVOETNNTIKAI SEIYMQQYFRLAMALKDKFDLLTSIRKHGIFGTSY. TVOETNNTIKAI SEIYMQQYFRLAMALKDKFDLLTSIRKHGIFGTSY. TFOEIQKAIKTV KNLYDQEAYFLLATRLKEKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQAYFLLATRVKDKFDLLTTIRTHGITFGTKH. TFGEIQKAIKTV SNLYDQAYFLLATRVKDKFDLLTTIRTHGITFGTSH. TFGEIQKAIKTV SNLYDQAYFLLATRVKDKFDLLTILTINGITFGTSH. TFGEIQKAIKTV SNLYDQAYFLLATRVKDKFDLLTILTINGITFGTSH. TFGEIQKAIKTV SNLYDQAYFLLATRVKDKFDLLEILENINGIVFGSLA. LLSNSGPIRQL TDLYDKDAYFDLANNLKDRFDLLKILAMHGITFGTSHTSSNIQNAVKSV 160 170 180 190 200 SQE.NPDIECAEV. TKGTPELYEIGICFTFTADSFFQCPSDTCDKT. AKVLFRR TMG, FFNLTCNKQ MELGEIGICFDYKNWIDCPNFETCKATR.GITFT. TGG, FFNLTCSRL RELKEIGICFDETVKNWIDCPNFETCKAFT. NEGVMFF. TGG, FFNLTCSRL RELKEIGICFDETVKNWIDCPNFETCKAFT.NGGIFFF. TGG, FFNLTSCTKRQ MELGEIGICFDSKVKNVIDCPNFETCKAFT.NGGIFFF. TGG, FFNLTSCTKRQ

FIG. 3. Amino acid alignment of 11 S-allele proteins. The bold residues are sites at which two species share an amino acid polymorphism. Simulations were performed as in Table 2 to test the significance of the number of shared polymorphisms. One allele of each of two species was randomly drawn, and random mutations in nucleotide sequences were generated, based on the number of observed polymorphisms in each species. These sequences were then translated, and the amino acid sequences were compared to ascertain the number of shared amino acid polymorphisms. The numbers of shared amino acid polymorphisms baserved in the data (and the count of cases in the random simulations having a greater number of shared polymorphisms) were as follows: *N. alata–P. inflata*, 17 (8/1000); *N. alata–S. chacoense*, 23 (0/1000); *P. inflata–S. chacoense*, 6 (0/1000); polymorphisms shared by all three species, 4 (12/1000).

stitutions over synonymous substitutions. Further analyses of the rates of synonymous and nonsynonymous substitution and of the coalescence properties of self-incompatibility alleles will be presented elsewhere. Because of its wide distribution over nearly half of the families of flowering plants (1), gametophytic self-incompatibility may prove to be one of the most general mechanisms for generating trans-specific polymorphism.

We thank Marilyn A. Anderson and Adrienne E. Clarke for providing the nucleotide sequences of S_{3nic} and S_{6nic} and Michael T. Clegg, Dan E. Krane, Stephen W. Schaeffer, and Thomas S. Whittam for insightful discussions. This work was supported by grants from the National Science Foundation (T.-h.K.) and the National Institutes of Health (A.G.C.).

- 1. de Nettancourt, D. (1977) Incompatibility in Angiosperms (Springer, New York).
- 2. Emerson, S. (1938) Genetics 23, 190-202.
- 3. Emerson, S. (1939) Genetics 24, 524-537.
- 4. Wright, S. (1939) Genetics 24, 538-552.
- Kheyr-Pour, A., Bintrim, S. B., Ioerger, T. R., Remy, R., Hammond, S. A. & Kao, T.-h. (1990) Sex. Plant Reprod. 3, 88-97.
- Anderson, M. A., McFadden, G. I., Bernatzky, R., Atkinson, A., Orpin, T., Dedman, H., Tregear, G., Fernley, R. & Clarke, A. E. (1989) *Plant Cell* 1, 483-491.
- Ai, Y., Singh, A., Coleman, C. E., Ioerger, T. R., Kheyr-Pour, A. & Kao, T.-h. (1990) Sex. Plant Reprod. 3, 130–138.
- Xu, B., Mu, J., Nevins, D. L., Grun, P. & Kao, T.-h. (1990) Mol. Gen. Genet., in press.
- 9. Saitou, N. & Nei, M. (1987) Mol. Biol. Evol. 4, 406-425.
- Swofford, D. (1990) Mac/PAUP: Phylogenetic Analysis Using Parsimony (Univ. of Illinois, Urbana).
- Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1979) in Atlas of Protein Sequence and Structure, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 345-352.
- 12. Felsenstein, J. (1988) Annu. Rev. Genet. 22, 521-565.
- 13. Nei, M. & Gojobori, T. (1986) Mol. Biol. Evol. 3, 415-426.
- Jukes, T. H. & Cantor, C. R. (1969) in Mammalian Protein Metabolism, ed. Munro, H. N. (Academic, New York), Vol. 3, p. 21.
- 15. Mazur, B. J. & Chui, C.-F. (1985) Nucleic Acids Res. 13, 2373-2386.
- Tumer, N. E., Clark, W. G., Tabor, G. J., Hironaka, C. M., Fraley, R. T. & Shah, D. M. (1986) Nucleic Acids Res. 14, 3325-3342.
- Wolter, F. P., Fritz, C. C., Willmitzer, L., Schell, J. & Schreier, P. H. (1988) Proc. Natl. Acad. Sci. USA 85, 846–850.
- Meagher, R. B., Berry-Lowe, S. & Rice, K. (1989) Genetics 123, 845-863.
- Wolfe, K. H., Sharp, P. M. & Li, W.-H. (1989) J. Mol. Evol. 29, 208-211.
- 20. Takahata, N. (1990) Proc. Natl. Acad. Sci. USA 87, 2419-2432.
- 21. Takahata, N. & Nei, M. (1990) Genetics 124, 967-978.
- 22. Clark, K. R., Okuley, J. J., Collins, P. D. & Sims, T. L. (1990)
- Plant Cell 2, 815–826.
 23. Nasrallah, J. B., Yu, S.-M. & Nasrallah, M. E. (1988) Proc. Natl. Acad. Sci. USA 85, 5551–5555.
- Ebert, P. R., Anderson, M., Bernatzky, R., Altschuler, M. & Clarke, A. (1989) Cell 56, 255–262.
- 25. Figueroa, F., Eberhardt, G. & Klein, J. (1988) Nature (London) 335, 265-267.
- Lawlor, D. A., Ward, F., Ennig, P., Jackson, A. & Parham, P. (1988) Nature (London) 335, 268-271.
- Klein, J. (1980) in *Immunology 80*, eds. Fougereau, K. & Dausset, J. (Academic, New York), pp. 239-253.
- 28. Hughes, A. L. & Nei, M. (1988) Nature (London) 335, 167-170.