S-RNase–mediated Gametophytic Self-Incompatibility is Ancestral in Eudicots

J. E. Steinbachs and K. E. Holsinger†*

*The Field Museum, Department of Zoology, Chicago; and †Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs

The evolutionary relationship between self-incompatibility systems in different families of flowering plants has long been a topic of interest. Physiological differences in the mode of gene action and the enormous sequence differences between genes with different modes of action suggest that many instances of self-incompatibility have arisen independently. In contrast, previous analyses of the *S*-RNase associated with gametophytic self-incompatibility in the eudicot families (Solanaceae, Scrophulariaceae, and Rosaceae) have suggested that sequences within families form well-supported and distinct lineages. In this study we demonstrate that in fact, *S*-RNase–mediated gametophytic self-incompatibility evolved only once in the eudicots.

Introduction

The vast majority of flowering plants produce both pollen and ovules within a single flower. But ''Nature abhors self-fertilization'' (Darwin 1877, p. 293), and a genetically determined self-incompatibility system has evolved to prevent self-fertilization in many flowering plants. In fact, genetically determined self-incompatibility appears to have evolved independently at least 21 times during the evolution of flowering plants. Not surprisingly, the molecular basis of the reaction differs dramatically among lineages. In the sporophytic self-incompatibility system of *Brassica*, the stylar response is mediated by protein receptor kinases (Schopfer, Nasrallah, and Nasrallah 1999). Gametophytic self-incompatibility in Papaveraceae is mediated by cytosolic free calcium ions (Franklin-Tong et al. 1993) and in grasses by *S*-type thioredoxin (Li et al. 1997). In all eudicots (Nandi, Chase, and Endress 1998) where the molecular details of the gametophytic self-incompatibility response are known, however, the stylar response is mediated by a glycoprotein with ribonuclease activity (an *S*-RNase), suggesting that gametophytic self-incompatibility mediated by *S*-RNases arose only once in eudicots. Our phylogenetic analysis of 72 *S*-RNase and *S*-like RNase DNA sequences shows that *S*-RNases have a single origin, that they were derived from *S*-like RNases, and that *Luffa S*-like RNases are derived from *S*-RNase ancestors. Thus, *S*-RNase–mediated gametophytic self-incompatibility appears to be the ancestral condition in eudicots.

The molecular basis of the gametophytic self-incompatibility mechanism has been well studied in the Solanaceae. In this family the most abundant protein in the style during the self-incompatibility response is a glycoprotein (Kehyr-Pour and Pernes 1985) with ribonuclease (RNase) activity (McClure et al. 1989). In fact, the self-incompatibility response will not occur if the molecule lacks RNase activity (Huang et al. 1994). Thus, the term *S*-RNase refers to the characteristic protein associated with the self-incompatibility response.

An *S*-RNase is also involved with the self-incompatibility response in the other three groups of eudicots in which the molecular basis of gametophytic self-incompatibility has been documented, i.e., Rosaceae (Sassa et al. 1996), Scrophulariaceae (Xue et al. 1996), and Campanulaceae (Stephenson et al. 1992), although the sequence of the gene encoding the *S*-RNase protein in the Campanulaceae has yet to be determined. Neither the Papaveraceae nor the Poaceae, the only other families with gametophytic self-incompatibility to have been examined at the molecular level, rely on an *S*-RNase in the self-incompatibility response (Franklin-Tong et al. 1993; Li et al. 1997).

Stigmatic glycoproteins associated with the self-incompatibility response in Papaveraceae have been isolated and cloned (Foote et al. 1994). The sequence shows no similarity to the *S*-RNase protein, but it does show some similarities to the *SLG* and *SRK* genes involved in the self-incompatibility response of *Brassica* (Walker et al. 1996). Furthermore, the self-incompatibility response in *Papaver* involves a highly complex series of events, including changes in calcium ion concentration, phosphorylation of specific proteins, transcription of specific genes within pollen tubes, and DNA fragmentation within nuclei of pollen tubes (Jordan, Franklin, and Franklin-Tong 2000; Snowman et al. 2000).

Unlike the other systems, gametophytic self-incompatibility in grasses is controlled by two unlinked multiallelic loci, *S* and *Z*. Recent molecular work indicates that protein kinases, similar to those in *Brassica*, and a thioredoxin play some role in self-incompatibility. Furthermore, the thioredoxin-like gene appears to be located about 1 cm from the *S* locus. Products of the *Z* locus have yet to be identified (Li et al. 1997; Baumann et al. 2000).

The evidence demonstrating the involvement of different molecules in self-incompatibility in the Poaceae, the Papaveraceae, and the families with *S*-RNases argues that gametophytic self-incompatibility likely has multiple origins. It is reasonable to conclude that there are at least three separate origins of gametophytic self-incompatibility in angiosperms. Is it possible, however,

Key words: *S*-RNase, gametophytic self-incompatibility, molecular evolution.

Address for correspondence and reprints: J. E. Steinbachs, The Center for Genomics and Bioinformatics, Indiana University, Jordan Hall, Bloomington, Indiana 47405-3700.
E-mail: jen@compbiology.org.

Mol. Biol. Evol. 19(6):825–829. 2002

that the gametophytic self-incompatibility system of the Rosaceae, Solanaceae, and Scrophulariaceae has only one origin? If so, it would suggest that gametophytic self-incompatibility in all eudicots (Nandi, Chase, and Endress 1998) had a single origin (Holsinger and Steinbachs 1997).

Other studies have considered this question (Sassa et al. 1996; Xue et al. 1996; Richman, Broothaerts, and Kohn 1997; Ushijima et al. 1998; Igic and Kohn 2001), examining the relationships among the *S*-RNases in eudicots and among the structurally related *S*-like RNases. Sassa et al. (1996) first demonstrated that the rosaceous and solanaceous *S*-RNases each formed a monophyletic clade. With the addition of three *S*-RNase sequences from *Antirrhinum*, Xue et al. (1996) suggested that all *S*-RNases formed a monophyletic clade, sharing a common ancestor, but more recent studies found very weak bootstrap support for the nodes uniting all *S*-RNases (Richman, Broothaerts, and Kohn 1997; Ushijima et al. 1998). Previous studies included only a small number of *S*-like RNases (e.g., 7 in Richman, Broothaerts, and Kohn [1997]; 14 in Ushijima et al. [1998]). As a result, these studies had only a limited ability to distinguish between single and multiple origins of *S*-RNase–mediated gametophytic self-incompatibility. Igic and Kohn (2001) improved on previous studies by increasing taxon sampling for the *S*-like group of sequences in addition to using maximum likelihood for tree reconstruction.

As with Igic and Kohn (2001), we independently took advantage of the substantial amount of new sequence information that has become available and assembled the most extensive sample of *S*-RNases and *S*-like RNases yet analyzed. Using more sophisticated alignment tools and methods of phylogenetic analysis, we demonstrate that *S*-RNase–mediated gametophytic self-incompatibility is ancestral in the eudicots.

Materials and Methods

Independent Origins of Genetically Determined Self-Incompatibility

To determine the number of times genetically determined self-incompatibility has evolved independently, we counted the number of instances of independent origins of sporophytic self-incompatibility as determined in a previous study (Holsinger and Steinbachs 1997). This tally accounts for four origins, one in each of Asterid I/II, Asterid III, Rosid I/II/III, and Caryophyllids. We then included the three known different systems of gametophytic self-incompatibility (*S*-RNase–mediated, the system in Papaveraceae, and the system in Poaceae). We also added one for the gametophytic self-incompatibility system found in Magnoliid II. Heterostyly has originated independently throughout the angiosperms; its genetic underpinnings and mode of action differ from those of both sporophytic self-incompatibility and gametophytic self-incompatibility, so we also counted each of these origins (13 in total).

Sequences Analyzed

We collected 72 amino acid sequences from PFAM 5.3 (Bateman et al. 2000), for which there existed a corresponding nucleotide sequence in GenBank (version 117) (Benson et al. 2000). After the selection process, our data set consisted of 49 *S*-RNases from Rosaceae, Solanaceae, and Scrophulariaceae, 21 *S*-like RNases from flowering plants, and two fungal RNases. The final data set includes the following organisms (abbreviated name in alignment: Genbank ID).

Fungal RNases

Aspergillus (rnt2-aspor: g133241), *Rhizopus*(rnrh-rhini: g133233).

S-*like I*

Nicotiana (rn-ngr2: g5902454), *Arabidopsis* (rns2-arath: g289210), *Calystegium* (c-sepium: g7288208).

S-*like II*

Arabidopsis (arath: g4262171; rns3-arath: g1173105; i64-arath: g5080798; rns1-arath: g561998; i65-arath: g5080799), *Pyrus* (rnpyrpyr: g1526417), *Cicer* (cicer: g3860325), *Nicotiana* (rne-nical: g532754; rn-ngr3: g5902456), *Solanum* (rnle-lyces: g1710615; rnlx-lyces: g1710616), *Zinnia* (rnze2-zele: g2148018; rnze1-zele: g2148017), *Hordeum* (hvulgare: g7435265), *Nelumbo* (nelumbo: g168740), *Zea* (kin1-zea: g1698670).

Luffa S-*like*

Luffa (lc1-luffa: g976231; lc2-luffa: g976233).

Solanaceae S-*RNases*

Petunia (sb1-pethyb: g4586870; sv-pethyb: g6706722; sxpethyb: g169248; s3-pethyb: g463993; sb2-pethyb: g4586872; s1-pethyb: g463991; sxb-pethyb: g169250; s1-petint: g169242; s3-petint: g169244), *Nicotiana* (rnsb-nical: g2696960; s7-nical: g533129; s6b-nical: g482815; rns2-nical: g133234; sa2-nical: g1184096; rns-nical: g2696958; nicsyl: g2578426), *Solanum* (s12-solch: g5919069; s11-solch: g548222; s14-solch: g7110526; s11-lperu: g1002594; s12-lperu: g1478373; s3-lperu: g2894088; rns2-soltu: g21576).

Scrophulariaceae S-*RNases*

Antirrhinum (rns2-anthi: g2500572; rns5-anthi: g1405428; rns4-anthi: g2500573).

Rosaceae S-*RNases*

Malus/*Pyrus* clade—*Malus* (s26-mdom: g2407178; s9-mdom: g642045; sd-mdom: g7229073; s3-mdom: g643447; s2-mdom: g643445; s27-mdom: g2407180; s24-mdom: g2407182; sh-mdom: g7229075; sg-mdom: g4587109; sf-mdom: g1018987; stmtrans: g7212796), *Pyrus* (s7-pyrpyr: g3434963; s1-pyrpyr: g3434939; s6-pyrpyr: g3434961; s3-pyrpyr: g7384768; s5-pyrpyr:

g1772448; s2-pyrpyr: g4850324). *Prunus* clade—*Prunus* (s6-pavium: g4115488; s1-pavium: g5763515; s4-pavium: g5763517; s3-pavium: g4115490; sb-pdulcis: g3927877; sc-pdulcis: g3927879).

Sequence Analysis

We first used MULTICLUSTAL (Yuan et al. 1999) to align the amino acid sequences. MULTICLUSTAL uses ClustalW (Thompson, Higgins, and Gibson 1994) and Boxshade iteratively to identify a set of alignment parameters that produce a high-scoring alignment. By identifying optimal ClustalW settings, we ensure that our alignment is objective and not dependent on the subjectivity that is associated with the manual alignment process. Furthermore, by reporting our optimal parameter settings, other researchers can independently evaluate our analysis. For this data set the ClustalW parameters are as follows—pairwise gap open penalty: 4; pairwise gap extension penalty: 0.1; pairwise matrix: ID; multiple gap open penalty: 20; multiple gap extension penalty: 0.1; multiple matrix: PAM. We then used mrtrans (Pearson 1992) to align the DNA sequences, codon by codon, to the amino acid alignment.

Previous studies examining the primary and secondary structure of the RNase molecule have demonstrated that there are five highly conserved regions with two hypervariable regions (Green 1994). A signal peptide lies at the N-terminus. We found it difficult to obtain a reliable alignment in the signal peptide region and in the set of residues found after the fifth conserved region. Whereas we aligned the amino acid sequences and corresponding DNA sequences along the entire length of the protein, for the phylogenetic analysis we truncated the sequences at both ends; the part of the sequence used starts at the beginning of the shortest signal peptide (arath sequence) and ends at the end of the shortest protein (sb1_pethyb sequence).

We used the Akaike Information Criterion in Modeltest (Posada and Crandall 1998) to determine the best-fit model of evolution in a likelihood framework. In so doing, we ensure that the model employed in tree reconstruction has some statistical justification. For tree reconstruction the Bayesian methodology we use (Huelsenbeck and Ronquist 2002) incorporates realistic models of DNA sequence evolution, and it allows rapid and accurate assessment of the reliability of the phylogenetic estimates we obtain. We used MrBayes (Huelsenbeck and Ronquist 2002), a Bayesian tree estimation program, on the aligned DNA sequences, with the generalized time reversible (GTR) model of sequence evolution (Tavare 1986) including both among-site rate variation and invariable sites (see online version for MrBayes script file).

Results and Discussion

Figure 1 summarizes the phylogenetic analyses. Because we focus on deep nodes, we collapsed the terminal branches so as to better highlight the larger clades (see online version for a complete, detailed reconstruction). As has been seen in many other studies (Sassa et al. 1996; Richman, Broothaerts, and Kohn 1997; Ushijima et al. 1998; Igic and Kohn 2001), our analysis finds that within-family phylogenies of self-incompatibility alleles are not congruent with organismal phylogenies (results in online version).

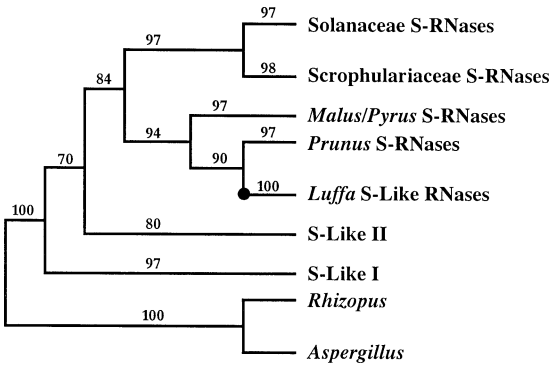

FIG. 1.—Genealogy of *S*-RNases and *S*-like RNases. Numbers above the branch correspond to posterior probabilities. For simplicity, we have collapsed terminal branches to form the designated clades, each of which consists of sequences as specified in *Materials and Methods.* The circle indicates a possible duplication or change in function (or both) giving rise to the *Luffa S*-like sequences.

Some clear trends emerge from these analyses: *S*-RNases within both the Solanaceae and the Scrophulariaceae are monophyletic; *Luffa S*-like RNases appear to be derived from functional *S*-RNases; *S*-like I and *S*-like II RNases are monophyletic, but their relationship to one another is equivocal. Most important, the posterior probability that *S*-RNases in our sample have a single common origin is 84%. Moreover, the *Luffa S*-like RNase sequences appear to be associated with the loss of self-incompatibility: the large clade in which *Luffa* is found has either switched to selfing or found other ways of avoiding selfing (e.g., monoecy, protandry, dioecy).

The similarity between *Luffa S*-like alleles and Rosaceae *S* alleles should not be surprising. Previous studies have demonstrated a decrease in RNase activity in self-compatible plants in members of both the Solanaceae (no RNase activity) (Royo et al. 1994) and the Rosaceae (only moderate activity) (Sassa, Hirano, and Ikehashi 1992). Additionally, the extracellular ribonuclease RNase X2, found in the pistils of *Petunia inflata*, shows similar enzymatic properties to *S* alleles even though the *rxn2* locus is not linked to the *S* locus, nor does the protein play any role in the self-incompatibility response. Furthermore, a genealogical analysis of RNase X2 places it into a clade containing Solanaceae *S* alleles, suggesting that it too is associated with duplication and the acquisition of a new function (Lee, Singh, and Kao 1992).

Although self-incompatibility has evolved independently at least 21 times in angiosperms (see online version for details), our results strongly suggest that *S*-RNase–mediated gametophytic self-incompatibility evolved only once in the ancestor to extant eudicots. At least three additional lines of evidence are needed to provide a strong test of this hypothesis: (1) The molecular details of gametophytic self-incompatibility must be elucidated in other eudicots. Our hypothesis predicts that the stylar

response will be mediated by an *S*-RNase. (2) New *S*-RNase sequences, whether from families already surveyed or from other eudicot families with gametophytic self-incompatibility, must be obtained. Our hypothesis predicts that they will be part of the clade including Solanaceae, Rosaceae, and Scrophulariaceae *S*-RNase sequences. (3) New *S*-like RNases from flowering plants must be obtained. Our hypothesis predicts that they will fall outside the *S*-RNase clade unless they are associated with loss of gametophytic self-incompatibility.

Supplementary Material

The accompanying table (see online version) summarizes the occurrences of genetically determined self-incompatibility in angiosperms. The accompanying figure provides more detail in the phylogeny than is seen in figure 1. We have also provided a file containing the aligned DNA sequence data and the commands used in MrBayes.

Acknowledgments

We thank our colleagues at the University of Connecticut, The Field Museum, and the Chicago Botanic Garden for helpful discussions. We were supported by the National Science Foundation (K.E.H.), the Natural Sciences and Engineering Research Council of Canada (J.E.S.), and Sun Microsystems (J.E.S.).

LITERATURE CITED

BATEMAN, A., E. BIRNEY, R. DURBIN, S. R. EDDY, K. L. HOWE, and E. L. L. SONNHAMMER. 2000. The Pfam protein families database. Nucleic Acids Res. **28**:263–266.

BAUMANN, U., J. JUTTNER, X. BIAN, and P. LANGRIDGE. 2000. Self-incompatibility in the grasses. Ann. Bot. **85**(Suppl. A): 203–209.

BENSON, D. A., I. KARSCH-MIZRACHI, D. J. LIPMAN, J. OSTELL, B. A. RAPP, and D. L. WHEELER. 2000. GenBank. Nucleic Acids Res. **28**:15–18.

DARWIN, C. 1877. The various contrivances by which orchids are fertilized by insects. John Murray, London.

FOOTE, H. C. C., J. P. RIDE, V. E. FRANKLIN-TONG, E. A. WALKER, M. J. LAWRENCE, and F. C. H. FRANKLIN. 1994. Cloning and expression of a distinctive class of self-incompatibility *S* gene from *Papaver rhoeas* L. Proc. Natl. Acad. Sci. USA **91**:2265–2269.

FRANKLIN-TONG, V. E., J. P. RIDE, N. D. READ, A. J. TREWAVAS, and F. C. H. FRANKLIN. 1993. The self-incompatibility response in *Papaver rhoeas* is mediated by cytosolic free calcium. Plant J. **4**:163–177.

GREEN, P. J. 1994. The ribonucleases of higher plants. Annu. Rev. Plant Phys. Plant Mol. Biol. **45**:421–445.

HOLSINGER, K. E., and J. E. STEINBACHS. 1997. Mating systems and evolution in flowering plants. Pp. 223–248 *in* K. IWATSUKI and P. H. RAVEN, eds. Evolution and diversification of land plants. Springer-Verlag, Tokyo.

HUANG, S., H. S. LEE, B. KARUNANANDAA, and T. H. KAO. 1994. Ribonuclease activity of *Petunia inflata* S proteins is essential for rejection of self-pollen. Plant Cell **6**:1021–1028.

HUELSENBECK, J. P., and F. R. RONQUIST. 2002. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755.

IGIC, B., and J. R. KOHN. 2001. Evolutionary relationships among self-incompatibility RNases. Proc. Natl. Acad. Sci. USA **98**:13167–13171.

JORDAN, N. D., F. C. H. FRANKLIN, and V. E. FRANKLIN-TONG. 2000. Evidence for DNA fragmentation triggered in the self-incompatibility response in pollen of *Papaver rhoeas*. Plant J. **23**:471–479.

KEHYR-POUR, A., and J. PERNES. 1985. A new *S*-allele and specific *S*-proteins associated with two *S*-alleles in *Nicotiana alata*. Pp. 191–196 *in* D. L. MULCAHY, G. B. MULCAHY, and E. OTTAVIANO, eds. Biotechnology and ecology of pollen. Proceedings of the International Conference on the Biotechnology and Ecology of Pollen; 9–11 July, 1985; University of Massachusetts, Amherst. Springer-Verlag, New York.

LEE, H. S., A. SINGH, and T. H. KAO. 1992. RNase X2, a pistil-specific ribonuclease from *Petunia inflata*, shares sequence similarity with solanaceous S proteins. Plant Mol. Biol. **20**: 1131–1141.

LI, X., N. PAECH, J. NIELD, D. HAYMAN, and P. LANGRIDGE. 1997. Self-incompatibility in the grasses: evolutionary relationship of the *S* gene from *Phalaris coerulescens* to homologous sequences in other grasses. Plant Mol. Biol. **34**: 223–232.

MCCLURE, B. A., V. HARING, P. R. EBERT, M. A. ANDERSON, R. J. SIMPSON, F. SAKIYAMA, and A. E. CLARKE. 1989. Style self-incompatibility gene products of *Nicotiana alata* are ribonucleases. Nature **342**:955–957.

NANDI, O. I., M. W. CHASE, and P. K. ENDRESS. 1998. A combined cladistic analysis of angiosperms using *rbc*L and non-molecular data sets. Ann. Mo. Bot. Garden **85**:137–212.

PEARSON, W. R. 1992. mrtrans. University of Virginia, Charlottesville.

POSADA, D., and K. A. CRANDALL. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14**:817–818.

RICHMAN, A. D., W. BROOTHAERTS, and J. R. KOHN. 1997. Self-incompatibility RNases from three plant families: homology or convergence? Am. J. Bot. **84**:912–917.

ROYO, J., C. KUNZ, Y. KOWYAMA, M. ANDERSON, A. E. CLARKE, and E. NEWBIGIN. 1994. Loss of a histidine residue at the active site of *S*-locus ribonuclease is associated with self-compatibility in *Lycopersicon peruvianum*. Proc. Natl. Acad. Sci. USA **91**:6511–6514.

SASSA, H., H. HIRANO, and H. IKEHASHI. 1992. Self-incompatibility-related RNases in styles of Japanese Pear (*Pyrus serotina* Rehd.). Plant Cell Physiol. **33**:811–814.

SASSA, H., T. NISHIO, Y. KOWYAMA, H. HIRANO, T. KOBA, and H. IKEHASHI. 1996. Self-incompatibility (*S*) alleles of the Rosaceae encode members of a distinct class of the *T*2/*S* ribonuclease superfamily. Mol. Gen. Genet. **250**:547–557.

SCHOPFER, C. R., M. E. NASRALLAH, and J. B. NASRALLAH. 1999. The male determinant of self-incompatibility in *Brassica*. Science **286**:1697–1700.

SNOWMAN, B. N., A. GEITMANN, S. R. CLARKE, C. J. STAIGER, F. C. H. FRANKLIN, A. M. C. EMONS, and V. E. FRANKLIN-TONG. 2000. Signalling and the cytoskeleton of pollen tubes of *Papaver rhoeas*. Ann. Bot. **85**(Suppl. A):49–57.

STEPHENSON, A. G., J. A. WINSOR, T. E. RICHARDSON, A. SINGH, and T. H. KAO. 1992. Effects of style age on the performance of self and cross pollen in *Campanula rapunculoides*. Pp. 117–121 *in* E. OTTAVIANO, D. L. MULCAHY, M. S. GORLA, and G. B. MULCAHY, eds. Angiosperm pollen and ovules. Springer-Verlag, New York.

TAVARE, S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Pp. 57–86 *in* R. M. MIURA, ed. Some mathematical questions in biology—DNA se-

quence analysis. American Mathematical Society, Providence, RI.

THOMPSON, J. D., D. G. HIGGINS, and T. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choices. Nucleic Acids Res. **22**:4673–4680.

USHIJIMA, K., H. SASSA, R. TAO, H. YAMANE, A. M. DANDEKAR, T. M. GRADZIEL, and H. HIRANO. 1998. Cloning and characterization of cDNAs encoding S-RNases from almond (*Prunus dulcis*): primary structural features and sequence diversity of the S-RNases in Rosaceae. Mol. Gen. Genet. **260**:261–268.

WALKER, E. A., J. P. RIDE, S. KURUP, V. E. FRANKLIN-TONG, M. J. LAWRENCE, and F. C. H. FRANKLIN. 1996. Molecular analysis of two functional homologues of the *S*3 allele of the *Papaver rhoeas* self-incompatibility gene isolated from different populations. Plant Mol. Biol. **30**:983–994.

XUE, Y., R. CARPENTER, H. G. DICKINSON, and E. S. COEN. 1996. Origin of allelic diversity in *Antirrhinum S* locus RNases. Plant Cell **8**:805–814.

YUAN, J., A. AMEND, J. BORKOWSKI, R. DEMARCO, W. BAILEY, Y. LIU, G. XIE, and R. BLEVINS. 1999. MULTICLUSTAL: a systematic method for surveying Clustal W alignment parameters. Bioinformatics **15**:862–863.

ELIZABETH KELLOGG, reviewing editor

Accepted January 15, 2002