Indian J. Genet., 78(2): 242-246 (2018)
DOI: 10.5958/0975-6906.2018.00031.7

Sequence diversification and phylogenetic analysis of self-incompatibility specific determinant genes in Brassicaceae

Xingguo Zhang*, Yuanyuan Li[1], Xingli Ma and Wei Zhu

College of Agronomy, Henan Agricultural University, 95 Wenhua Road, 450002 Zhengzhou, China; [1]Weifang University, 260000 Weifang, China

(Received: June 2017; Revised: March 2018; Accepted: April 2018)

Abstract

It is a long history of theoretical interest in understanding the genetic causes and consequences of shifts between outcrossing and inbreeding in plant mating systems. Self-incompatibility (SI) systems genetically promote outcrossing, and its transitions to inbreeding potentially changes genetic diversity in the evolution of species. SI in the Brassicaceae family is sporophytically controlled by a single locus (*S*-locus), which locates the female determinant gene *SRK* (*S*-locus receptor kinase) and the male determinant gene *SCR* (*S*-locus cysteine rich). Operation of diversifying selection maintains the diversity of *SRK* and *SCR* during the evolutionary course of *S* haplotypes. Here we used deduced amino acid sequences of *SRK* and/or *SCR* publicly available from seven species in three genera, *Brassica*, *Arabidopsis* and *Raphanus*, to investigate the sequence diversification and phylogenetic distribution of *S* haplotypes in the Brassicaceae family. The results reveal that *SCR* shows higher sequence distance than *SRK* within or between genera, and also in the interspecific pairs of *S* haplotypes. Phylogenetic analysis supports the conclusion that SI has a single origin but evolves differently within the Brassicaceae. The genetic mechanisms and consequences of the loss of SI in the Brassicaceae family are discussed.

Key words: Self-incompatibility, *SRK*, *SCR*, phylogenetic analysis, Brassicaceae

Introduction

Many flowering plants have a genetically controlled mechanism such as self-incompatibility (SI) to avoid potentially negative consequences of inbreeding. SI has been extensively studied in the Brassicaceae family and is sporophytically controlled by a single major locus, *S*-locus (Bateman 1955). The *S*-locus is comprised of a number of completely linked genes and segregates as an *S* haplotype (Nasrallah and Nasrallah 1993). The female determinant SRK (*S*-locus receptor kinase) (Takasaki et al. 2000) and the male determinant, SCR/SP11 (*S*-locus cysteine-rich protein/*S*-locus protein 11) (Schopfer et al. 1999; Suzuki et al. 1999), are involved in the SI recognition reaction in Brassicaceae. Many species in Brassicaceae are important resources for SI studies including self-fertile model plant *Arabidopsis thaliana* and its relatives, self-incompatible *A. lyrata*, and major vegetables and crops in *Brassica* and *Raphanus* genera. Breakdown of SI is an important topic which has been greatly studied in Brassicaceae. Two mechanisms have been proposed to account for the loss of SI. The major one is mutations of *SRK* or *SCR*, like *A. thaliana* (Tsuchimatsu et al. 2010) and *B. rapa* (Fujimoto et al. 2006). Another is the interruption of downstream components of the signal transduction pathway like *B. oleracea* (Cabrillac et al. 2001).

Due to the strong negative frequency-dependent selection, the SI system typically maintains a large number of haplotypes (Lawrence 2000). The distribution of *S* haplotypes has been well analyzed in some genera, in which some characteristic differences were found. For example, the *Brassica S* haplotypes can be divided into two dominant classes, class I and class II. Class I *S* haplotypes are dominant over class II *S* haplotypes in the pollen side but co-dominant in the stigma side (Nasrallah et al. 1991). Most of the previous phylogenetic analyses used either two genera or incomplete DNA sequences to dissect the evolution and genetic diversity of *S* haplotypes (Sato et al. 2002). But there are scanty reports using the protein data from both wild species and cultivated species for

*Corresponding author's e-mail: zxhnnd@126.com
Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by indianjournals.com; www.isgpb.org

phylogenetic analysis, in view that mature peptide region is better to discuss on the evolutionary point of SI determinant genes. In our previous study, an *S* haplotype was characterized from a self-compatible *B. rapa* cultivar Dahuangyoucai, which was proved to highly similar to *S-f2* haplotype found in another self-compatible *B. rapa* Yellow sarson, which gave rise to the confusion on the origin of *S* haplotype and genetic evolution (Zhang et al. 2013).

In the present study, phylogenetic analysis was performed for the first time with mature protein sequences of SRK and SCR alleles publicly available from seven Brassicaceae species belonging to three genera. The results will be helpful to understand the genetic diversity and evolution of *S* haplotypes in Brassicaceae and dissect the genetic causes and consequences of the breakdown of SI in higher plants.

Materials and methods

Sequences collecting

Sequences of *S*-locus genes were collected using genus name and gene name as keywords in Genbank of NCBI website (www.ncbi.nlm.nih.gov). Among all the searched sequences, those with complete CDS were chosen and their amino acid sequences were then collected. The sequence was named based on the species and *S* haplotype, for example, SRK of *S8* in *B. rapa* was described as BraSRK8 but for *A. thaliana*, the *S* haplotype was not denominated, so the ecotype was used as suffix, like AthSRK-Kas2 for SRK of *A. thaliana* ecotype Kas-2.

Phylogenetic analysis

All sequences were multiple aligned using CLUSTAL W program (Thompson et al. 1994) following the default parameters: for pairwise alignments, a gap opening penalty=10 and a gap extend penalty =0.1; for multiple alignments, a gap opening penalty=10 and a gap extend penalty=0.2. The delay divergent sequences = 30%, DNA transition weight =0.5 and use negative matrix OFF. Nucleotide sequence distance between and within each genus was calculated using MEGA3.0 (Kumar et al. 2004) following the parameters, gap/missing data: pairwise deletion; model: nucleotide, *p* distance; pattern among lineages: same (homogeneous); rates among sites: uniform rates. Phylogenetic trees were constructed with deduced amino acid sequences using the minimum-evolution method of MEGA3.0. The default parameters included gap/missing data, pairwise deletion; model, amino: poisson correction; substitutions to include, All; cut-off value for consensus/condensed tree, 50%. The bootstrap probabilities of 1000 trials were used for phylogeny test. No sequences appropriate for rooting the tree were found, so the tree was left unrooted.

Results

Sequence collection of SI determinant genes

A total of 35 SCR and 31 SRK from seven species of three genera were searched. The sequences included the predominant model plant, *A. thaliana*; the major *Brassica* crop, *B. napus* and the major vegetables, *B. rapa*, *B. oleracea* and *R. sativus*. These species showed different *S* phenotype intra-genus, like self-incompatible *A. lyrata* and self-compatible *A. thaliana* and also included the cultivated Brassicaceae species *B. rapa*, *B. napus* and the wild species *A. halleri*.

The length of *SRK* changed from 1,461 bp to 2,583 bp, and the amino acid length was from 795 to 860. *BraSRK733* has eight exons and *BraSRK734* has nine exons, which both were obviously different from other *SRK* genes with seven exons. The length of *SCR* was varied from 225 bp to 288 bp, and the corresponding protein was from 74 to 95 amino acids. In *Brassica*, all the class I *SCR*s were shorter than that of class II in nucleotide length. Inter-specific *S* haplotype pair showed highly similar *SRK* and *SCR*, like *BolS-15*, *BraS-60* and *BnaS-6*. They showed 98% (two bp difference of the 288 bp) and 99% (two bp difference of the 2571 bp) nucleotide sequence identity for the *SCR* and *SRK*, respectively.

Sequence diversification of SI determinant genes

Nucleotide sequence distance between and within the three genera were calculated (Table 1). *SCR* genes showed higher degree of sequence distance within *Brassica* (0.43) than other two genera, *Raphanus* (0.20) and *Arabidopsis* (0.32). It exhibited nearly the same level of polymorphism between genera, *Brassica* and

Table 1. The nucleotide sequences distance between and within the genus

Gene	Genus	*Brassica*	*Raphanus*	*Arabidopsis*	Overall
SCR	*Brassica*	0.43			0.44
	Raphanus	0.46*	0.20		
	Arabidopsis	0.48*	0.60*	0.32	
SRK	*Brassica*	0.15			0.20
	Raphanus	0.14*	–		
	Arabidopsis	0.25*	0.25*	0.17	

*,distance between genus

Raphanus (0.46), *Brassica* and *Arabidopsis* (0.48), but both were lower than that between *Raphanus* and *Arabidopsis* (0.60). *SRK* sequences displayed a moderate polymorphism i.e., the mean sequence distance within each group is 0.15 for *Brassica* and 0.17 for *Arabidopsis*. In view of only one *SRK* of *Raphanus* collected, it was impossible for calculating the distance within *Raphanus*. The distance within *Brassica* (0.15) was comparable to that between it and *Raphanus* (0.14), while both of them showed the same distance from *Arabidopsis* (0.25). For the overall mean distance of the nucleotide sequence, *SCR* displayed striking diversification (0.44) than *SRK* (0.20). Sequence distances of two class *S* haplotypes in *Brassica* were also calculated (Table 2). Class I showed lower sequence distance than class II, like *SRK* within class I and class II were 0.13 and 0.03, respectively, while *SCR* within class I was 0.33 and within class II

Table 2. The nucleotide sequences distance of *S* haplotypes in *Brassica*

Gene	Class	I	II
SCR	I	0.33	
	II	0.12*	0.59
SRK	I	0.13	
	II	0.21*	0.03

*,distance between class

was 0.59. Besides above mentioned sequence distances, *SRK* manifested high polymorphism between classes than *SCR* (distance 0.12 for *SCR* and 0.21 for *SRK*) in view of the length difference.

Phylogenetic trees constructed using SRK sequences

The interspecific pair *S* haplotypes were clustered together with 100% bootstrap support, such as *BolSRK12* and *BnaSRK1*, *BraSRK54* and *BolSRK60*, reflecting the rationality of the phylogenetic tree of SRK (Fig. 1). All the *S* haplotypes from *Brassica* including class I *Brassica S* haplotypes, class II *S* haplotypes and one *RsaSRKa* formed a major group, except two *Brassica S* haplotypes, *BraSRK733* and *BraSRK734*, with longer branches than other haplotypes were clustered with *Arabidopsis S* haplotype clade. *AlySRK6* and *AlySRKb* were seperated as a deviant clade.

Phylogenetic trees constructed using SCR sequences

In the *SCR* tree constructed using 35 SCR protein

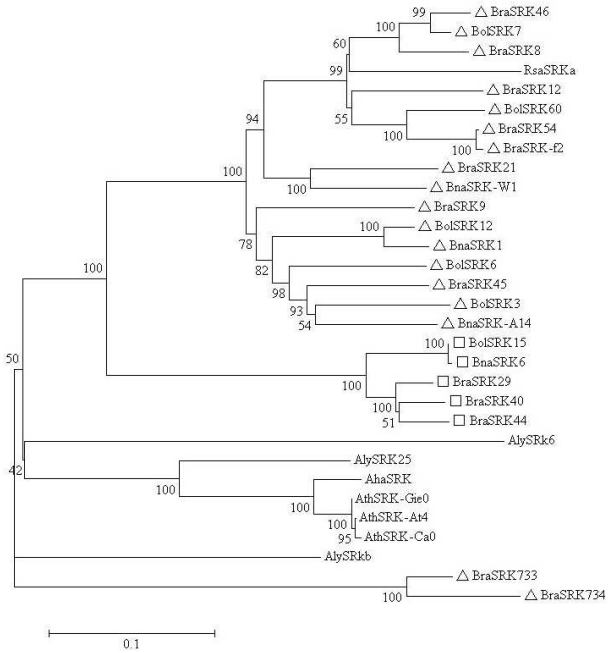

Fig. 1. Minimum-evolution tree constructed with the deduced amino acid of SRK: *Brassica* class I SRK (triangle); *Brassica* class II SRK (square). One thousand bootstrap trials were performed and the bootstrap values are shown beside the branch

sequences (Fig. 2), class I and class II *S* haplotypes of *Brassica* did not form a monophyletic group, which was obvious in the *SRK* tree. The same gene sequenced by different researchers, *BolSCR13* and *BolSCR13b*, and the interspecific pair consisting of one haplotype from each of *B. oleracea* (*BolSCR12*) and *B. rapa* (*BraSCR47*), were particularly clustered with 100% bootstrap value. Most of the class I haplotypes in *Brassica* were sorted to one clade, except that five *S* haplotypes were more closely related to *Brassica* class II *S* haplotypes and five *Raphanus S* haplotypes. All the class II *S* haplotype from *Brassica* and five *S* haplotype from *Raphanus* were assigned into the same group, indicating that the five *S* haplotypes in *Raphanus* might belong to class II. All *Arabidopsis* SCRs was formed into one group, including AlySCR6 and AlySCRb, even without well-supported bootstrap values, which was different from the profile of the SRK tree.

Genetic evolution of SI determinant genes in Brassicaceae

Combining the two phylogenetic trees, we found the general evolution model in Brassicaceae. *Brassica*/*Raphanus* split from *Arabidopsis* first, following by the

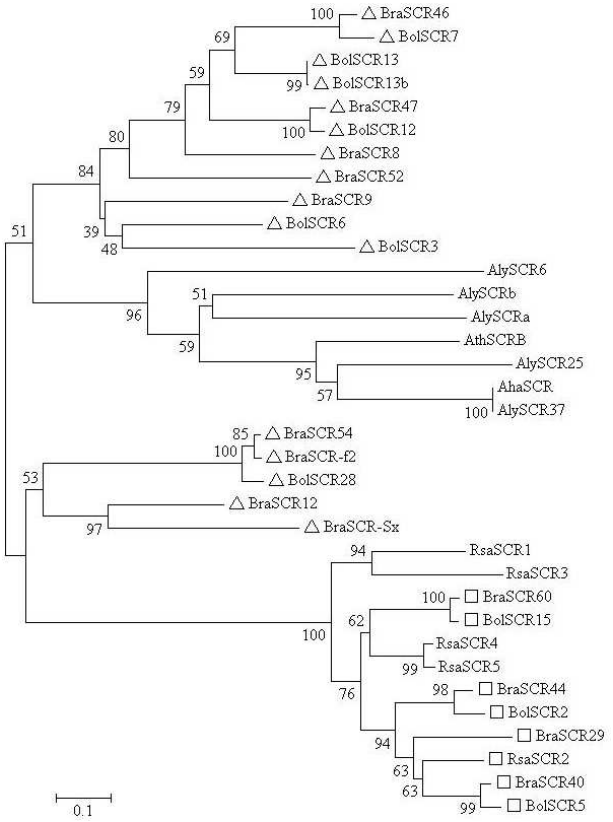

Fig. 2. Minimum-evolution tree constructed with the deduced amino acid of SCR: *Brassica* class I SCR (triangle); *Brassica* class II SCR (square). One thousand bootstrap trials were performed and the bootstrap values are shown beside the branch

divergence of *S* haplotypes before the species differentiation within genus. In *Brassica*, class I and class II *S* haplotypes diverge predate the species differentiation, then *S* alleles within class I diversify more ancient than those within class II.

Discussion

The situation that fewer recessive haplotypes at high frequency and a large number of dominant haplotypes at lower frequencies was predicted in *Brassica* (Billiard et al. 2007). It was also reflected in the present study. Greater sequence distance was observed in *SCR* alleles than in *SRK* alleles, but the phylogenetic analyses showed the similar topology, suggesting a positive correlation in divergence time between the female and male determinants. A variety of factors resulted in the high sequence diversification of *S* haplotypes. Little or no recombination occurs between *SRK* and *SCR* promotes the maintenance of extensive nucleotide diversity (Charlesworth et al. 2003). Gene conversion between *SRK* and *SLG* (*S*-locus glycoprotein) also plays a significant role (Sato et al. 2002). Wild and cultivated *Brassicaceae* species share similar levels of *SRK* diversity, indicating that domestication has had but a minor effect on *S*-locus diversity (Edh et al. 2009).

In order to highlight the pending questions, great progresses have been made in Brassicaceae (Tedder et al. 2011). Genetic and phylogenetic features of SI in the tribe Biscutelleae were observed departing from patterns in the reference *Arabidopsis*, while similar to *Brassica* (Leducq et al. 2014). Although the phylogenetic trees constructed using two SI determinants were unrooted, they showed similar pattern that the interspecific *S* haplotype pairs were closely clustered. Phylogenetic results supported the previous conclusion that SI has a single origin but evolves differently in the Brassicaceae family. *Brassica*/*Raphanus* split from *Arabidopsis* first, predating the divergence of *S* haplotypes followed by the species differentiation within genus (Zhang et al. 2011). Five class I *S* haplotypes were closely clustered to the major lineage of class II in the SCR tree, might be due to the synonymous and non-synonymous amino acid substitution rate (Tedder et al. 2011). Previous results found a single polymorphic 24 nucleotide small RNA, named *SP11* methylation inducer 2 (Smi2), controls the linear dominance hierarchy of the four class II *SP11* alleles in *B. rapa* (Yasuda et al. 2016). Multilayered dominance hierarchy was put forward to understand the dominance regulatory mechanism of SI in the Brassicaceae (Fujii and Takayama 2018).

As a longstanding subject, the genetic causes and consequences of loss of SI have received theoretical and empirical attention since the time of Darwin. In Brassicaceae, the model plant *A. thaliana* and cultivated plants in the genus *Brassica*, together with the genus *Capsella* have been extensively studied. Mutation of *S*-locus was generally thought to be the main causes for the breakdown of SI (Rea et al. 2010). Comparisons between multiple populations in selfing and outcrossing species in the Brassicaceae were helpful to evaluate the consequences of mating system differences. Further investigation should combining the sequence data of emerging genomic information, population structure, changes in effective population size and genetic diversity from cultivated plants to wild populations in relation to mating system, together with the individual variations such as maternal effects, background genetics, the history of outcrossing and possibly particular *S* haplotypes.

Authors' contribution

Conceptualization of research (XZ); Designing of the experiments (XZ, YL); Contribution of experimental materials (YL); Execution of field/lab experiments and data collection (XM, WZ); Analysis of data and interpretation (XM, WZ); Preparation of manuscript (XZ).

Declaration

The authors declare no conflict of interest.

Acknowledgements

This work was supported by grants from Natural Science Innovation Fund of Henan Agricultural University (30600963).

References

Bateman A. J. 1955. Self-incompatibility systems in angiosperms.III. *Cruciferae*, Heredity, **9**: 53-68.

Billiard S., Castric V. and Vekemans X. 2007. A general model to explore complex dominance patterns in plant sporophytic self-incompatibility systems. Genetics, **175**: 1351-1369.

Cabrillac D., Cock J. M., Dumas C. and Gaude T. 2001. The *S*-locus receptor kinase is inhibited by thioredoxins and activated by pollen coat proteins. Nature, **410**: 220-223.

Charlesworth D., Bartolomé C., Schierup M. H. and Mable B. K. 2003. Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. Mol. Biol. Evol., **20**: 1741-1753.

Edh K., Widén B. and Ceplitis A. 2009. The evolution and diversification of *S*-locus haplotypes in the Brassicaceae family. Genetics, **181**: 977-984.

Fujii S and Takayama S. 2018. Multilayered dominance hierarchy in plant self-incompatibility. Plant Reprod. **31**(1):15-19.

Fujimoto R., Sugimura T. and Nishio T. 2006. Gene conversion from *SLG* to *SRK* resulting in self-compatibility in *Brassica rapa*. FEBS Letters, **580**(2): 425-430.

Kumar S., Tamura K. and Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Bioinformatics, **5**: 150-163.

Lawrence M. J. 2000. Population genetics of the homomorphic self-incompatibility polymorphisms in flowering plants. Ann. Bot., **85**: 221-226.

Leducq J-B, Gosset C-C, Calin K, Schmitt É, Castric V and Vekemans X. 2014. Self-Incompatibility in Brassicaceae: Identification and Characterization of SRK-Like Sequences Linked to the S-Locus in the Tribe Biscutelleae. Genes Genomes Genetics, **4**: 983-992.

Nasrallah J. B. and Nasrallah M. E. 1993. Pollen-stigma signaling in the sporophytic self-incompatibility response. Plant Cell, **5**: 1325-133.

Nasrallah J. B., Nishio T. and Nasrallah M. E. 1991. The self-incompatibility genes of *Brassica*: expression and use in genetic ablation of floral tissues. Ann. Rev. Plant Physiol. Plant Mol. Biol., **42**: 393-422.

Rea A. C., Liu P. and Nasrallah J. B. 2010. A transgenic self-incompatible *Arabidopsis thaliana* model for evolutionary and mechanistic studies crucifer self-incompatibility. J. Exp. Bot., **61**(7): 1897-1906.

Sato K., Nishio T., Kimura R., Kusaba M., Suzuki T., Hatakeyama K., Ockendon D. and Satta Y. 2002. Co-evolution of the *S*-locus genes, *SRK*, *SLG* and *SP11*/*SCR*, in *Brassica oleracea* and *Brassica rapa*. Genetics, **162**: 931-940.

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

Suzuki G., Kai N., Hirose T., Fukai K., Nishio T., Takayama S., Isogai A., Watanabe M. and Hinata K. 1999. Genomic organization of the *S* locus: Identification and characterization of genes in *SLG*/*SRK* region of *S*(9) haplotype of *Brassica campestris* (syn. *rapa*). Genetics, **153**(1): 391-400.

Takasaki T., Hatakeyama K., Suzuki G., Watanabe M., Isogai A. and Hinata K. 2000. The *S* receptor kinase determines self-incompatibility in *Brassica* stigma. Nature, **403**: 913-916.

Tedder A., Ansell S. W., Lao X., Vogel J. C. and Mable B. K. 2011. Sporophytic self-incompatibility genes and mating system variation in *Arabis alpine*. Ann. Bot., **157**: 1-15.

Thompson J. D., Higgins D. G. and Gibson T. J. 1994. CLUSTAL W, improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., **22**(22): 4673-4680.

Tsuchimatsu T., Suwabe K., Shimizu-Inatsugi R., Isokawa S., Pavlidis P., Städler T., Suzuki G., Takayama S., Watanabe M. and Shimizu K. K. 2010. Evolution of self-compatibility in *Arabidopsis* by a mutation in the male specificity gene. Nature, **464**: 1342-1346.

Yasuda S., Wada Y., Kakizaki T., Tarutani Y., Miura-Uno E., Murase KFujii S., Hioki T., Shimoda T., Takada Y., Shiba H., Takasaki-Yasuda T., Suzuki G., Watanabe M. and Takayama S. 2016. A complex dominance hierarchy is controlled by polymorphism of small RNAs and their targets. Nat. Plants, **3**: 16206.

Zhang X. G., Yin D. M., Ma C. Z. and Fu T. D. 2011. Phylogenetic analysis of *S*-locus genes reveals the complicated evolution relationship of S haplotypes in *Brassica*. Plant Mol. Biol. Rep., **29**: 481-488.

Zhang X. G., Ma C. Z., Yin D. M., Zhu W., Gao C. B. and Fu T. D. 2013. Characterization of *S* haplotype in a new self-compatible *Brassica rapa* cultivar Dahuangyoucai. Czech J. Genet. Plant Breed., **49**(4): 157-163.