

## GENETIC DIVERSITY OF SIRE-1 RETROELEMENTS IN ANNUAL AND PERENNIAL GLYCINE SPECIES REVEALED USING SSAP

CATHERINE CHESNAY<sup>1</sup>, AMAR KUMAR<sup>1</sup> and STEPHEN R. PEARCE<sup>2\*</sup>

<sup>1</sup>Scottish Crop Research Institute, Invergowrie, Dundee, DD2 5DA, U.K.

<sup>2</sup>School of Life-Sciences, University of Sussex, Brighton, BN1 9QG, U.K.

**Abstract:** Sequence Specific Amplification Polymorphisms (SSAP) were used to measure the distribution and structure of SIRE-1 retroelement populations in annual and perennial *Glycine* species. For SSAP analysis, primers corresponding to a region immediately upstream of the 3'LTR of the soybean retroelement SIRE-1 were chosen. Analysis reveals that SIRE-1 is present throughout the *Glycine* genus and shows that the annual species have similar SIRE-1 populations whilst the perennial species have much more distinct and diverse populations. The high number of species-specific subgroups suggest that SIRE-1 has been active and evolving independently in each species during the course of *Glycine* evolution.

**Key words:** *Glycine*, Retroelement, SSAP, phylogeny, SIRE-1, *copied*

### INTRODUCTION

*Glycine* is a genus of leguminous plants which include the cultivated soybean (*G. max*) and its wild progenitor (*G. soja*) and a number of perennial species. As a result of the economic significance of soybean, the phylogenetic relationships and evolutionary history of the *Glycine* genus have been a major focus of interest. The phylogenetic composition of *Glycine* is extremely complex with individual species forming nine different genome groups (A-I) [1]. The annual *Glycine* species, which include the cultivated soybean *G. max* and its wild progenitor *G. soja* are mostly self-pollinated [2] and form a single genome group (G) whilst the highly diverse perennial *Glycine* species comprise the remaining eight genome groups. The evolution of the perennial species is highly complex and is probably the result of ancient inter-specific hybridization events between different genome group species. The production of novel hybrids and the

---

\* Author for correspondence; e-mail: S.R.Pearce@sussex.ac.uk

subsequent diversification of the species within each genome group means that major sterility barriers now exist which prevent inter-group hybrids being formed [3].

SIRE-1 is a soybean retroelement which is highly unusual as it contains an envelope gene characteristic of retroviruses but unlike most other envelope containing retroelements, which have an internal structural organization similar to Ty3-gypsy retrotransposons, it has a gene order of Ty1- *copia* retrotransposons [4]. Extensive sequence analysis shows that all the characterized SIRE-1 elements are relatively homogeneous and the dating of insertions by LTR analysis suggests that this element may be a relatively recent feature of the soybean genome [5]. Despite extensive sequence analysis of SIRE-1 insertions in soybean the evolution and distribution of the SIRE-1 retroelement within the perennial species of the *Glycine* genus remains unknown.

SSAP analysis, a method similar to AFLP but based on retroelement polymorphism has proved to be an extremely useful tool in determining the presence and insertional polymorphism of retroelements within crop species and their close species relatives [6-8]. We use SIRE-1 SSAP to investigate the distribution and composition of SIRE-1 populations in annual and perennial *Glycine* species and report that SIRE-1 is present in all *Glycine* species investigated. The high level of population heterogeneity between individual perennial species suggests that although many copies of SIRE-1 in soybean are evolutionarily young the ancestral SIRE-1 element has almost certainly been present throughout the evolution of the *Glycine* genus.

## MATERIALS AND METHODS

### Plant materials

Seeds of *Glycine max* cultivars and *Glycine* species; *soja*, *cyrtoloba*, *tomentella*, *clandestina*, *tabacina*, *canescens* and *falcata* were donated by Prof. R. Shoemaker, Iowa state University USA (Tab. 1). Surface sterilized seeds were germinated and DNA isolated by "qiaquick plant mini kit" (Qiagen). All species are diploid except *G. tabacina* and *G. tomentella* which are tetraploid.

### SSAP analysis

SSAP analysis [9] was carried out using SIRE-1 specific primer labeled with  $\gamma^{33}\text{P}$  ATP, as [10]. The SIRE-1 specific primer was designed to a region just upstream of the 3' LTR terminal sequence (AY205610 8227-2585; 5'-CAGTTATGCAAGTGGGATCAGCA-3'). Along with the labeled SIRE-1 primer, unlabelled Mse-I adaptor primers were used in the labeling reactions with two additional selective bases added (M+GC). The GC couple was chosen to maximize the priming efficiency of the reaction. Labeled products were resolved by denaturing 6% polyacrylamide electrophoresis and the products visualized by autoradiography.

Tab. 1. Plant material

#	Name	Species	Cultivar/USDA PI #	Genome size (1C) *
A	Ma1	<i>G. max</i>	Williams	1.13pg
B	Ma2	<i>G. max</i>	Williams 82	“
C	Ma3	<i>G. max</i>	#79593	“
D	Ma4	<i>G. max</i>	#A81-356022	“
E	Ma5	<i>G. max</i>	Harosoy 63	“
F	Ma6	<i>G. max</i>	Kingwa	“
G	So1	<i>G. soja</i>	#10404B	1.15pg
H	So2	<i>G. soja</i>	#440913B	“
I	So3	<i>G. soja</i>	#424004A	“
J	So4	<i>G. soja</i>	#468916	“
K	Cy1	<i>G. cyrtoloba</i>	#373993	1.33pg
L	Cy2	<i>G. cyrtoloba</i>	#373993	“
M	To1	<i>G. tomentella</i>	#321393	“
N	To2	<i>G. tomentella</i>	#321393	“
O	CL1	<i>G. clandestina</i>	#246590	1.20pg
P	CL2	<i>G. clandestina</i>	#246590	“
Q	Ta1	<i>G. tabacina</i>	#339661	2.03pg
R	Ta2	<i>G. tabacina</i>	#193232	“
S	Ca1	<i>G. canescens</i>	#440930	0.95pg
T	Ca2	<i>G. canescens</i>	#440936	“
U	Fa1	<i>G. canescens</i>	#246591	1.65pg
V	Fa2	<i>G. canescens</i>	#246591	“

\* Plant DNA C values database ([www.rbgekew.org.uk/cval/](http://www.rbgekew.org.uk/cval/))

### Data analysis

Banding data was scored manually into binary matrices. Phylogenetic distance was estimated with Nei's genetic distance [11] using powermarker [12]. Tree reconstruction used bootstrapping with a summary consensus tree based on 1000 individual analyses. Bootstrap values were calculated using PHYLIP [13].

## RESULTS

### Analysis of SIRE-1 populations using SSAP.

SSAP involves the restriction of genomic DNA with a frequent cutting restriction enzyme followed by the ligation of adapter sequences to facilitate subsequent PCR. Amplification is then carried out using a primer homologous to a highly represented and polymorphic retrotransposon sequence and also a primer corresponding to the adjacent restriction site adapter. SSAP usually

reveals insertional polymorphism of a given retrotransposon as the products are produced between a retrotransposon terminal sequences and a restriction site in the sequence flanking the transposable element. In this study we modify the technique by designing a primer which primes within a conserved region of SIRE-1 upstream of the 3'LTR terminus and primes toward a variable region of the element [5]. This modification of the technique means that the profiles generated represent the composition of the retrotransposon *population* rather than the insertional polymorphism within the genome as revealed by classic SSAP and are therefore extremely useful for investigating the evolution of the retroelement in different species. The banding profiles corresponding to the composition of the SIRE-1 populations in the collection of perennial and annual *Glycine* species is summarized in Tab. 2 and shown in Fig. 1.

Clear banding profiles have been produced in all plants showing that SIRE-1 is present in all the species studied. As *MseI* is sensitive to DNA methylation it must be noted that not *all* *MseI* restriction sites will be detected using this technique. The SSAP analysis reveals banding patterns in each accession which reflects the variety and relative copy numbers of SIRE-1 element populations. Activity and evolution will change the sequence composition and copy numbers of different sub-classes of SIRE-1 elements resulting in the loss or gain of *MseI* sites or retrotransposon primer sites resulting in changes to the SSAP profiles.

Tab. 2. Summary SSAP polymorphism data for *Glycine* accessions.

	Annual accessions (10 lines)										Combined data		
	A	B	C	D	E	F	G	H	I	J			
# bands scored	12	14	15	14	15	18	17	16	18	17	33		
# polymorphic	3	4	5	5	6	8	8	7	9	8	23		
% polymorphic	25	29	33	36	40	44	47	44	5	47	70		
	Perennial accessions (12 lines)												Combined data
	K	L	M	N	O	P	Q	R	S	T	U	V	
# bands scored	26	25	21	22	19	20	21	29	22	24	17	15	82
# polymorphic	23	22	18	19	16	17	18	26	19	21	14	12	79
% polymorphic	88	88	86	86	84	85	86	90	86	88	82	80	96

Only the most prominent clearly defined bands with sizes ranging from 50bp to approximately 1kb in size were scored as these represent the most significant components of the SIRE-1 populations. The SSAP profiles of SIRE-1 in the annual species, *G. max* and *G. soja* (lanes A-J) produce 33 bands, 70% polymorphic within the annual species with several bands being unique to particular cultivars. Ten of the bands (27%) are non-polymorphic (black arrows in Fig. 1). In contrast the perennial species (lanes K-V) have 82 bands with 79% being polymorphic within the perennial species. Only three are shared between all *perennial* accessions (white arrows) showing that there is much more

diversity of SIRE-1 populations within the perennial species. This shows that SIRE-1 has been evolving independently in each of these species, with the differences being larger for the perennial than the annual species.

Although the banding patterns indicate many new retroelement subgroups in individual plant species, there is also evidence of loss of element sub-classes. The clearest example of this is the absence of a band in one of the *G. falcata* accessions which is shared between all annual and perennial species studied (asterisk, Fig. 1). Band losses in closely related material is unusual but may be explained by the loss of the particular element sub-class due to deletion possibly as a result of recombination or changes to the restriction site or retrotransposon priming site in this accession.

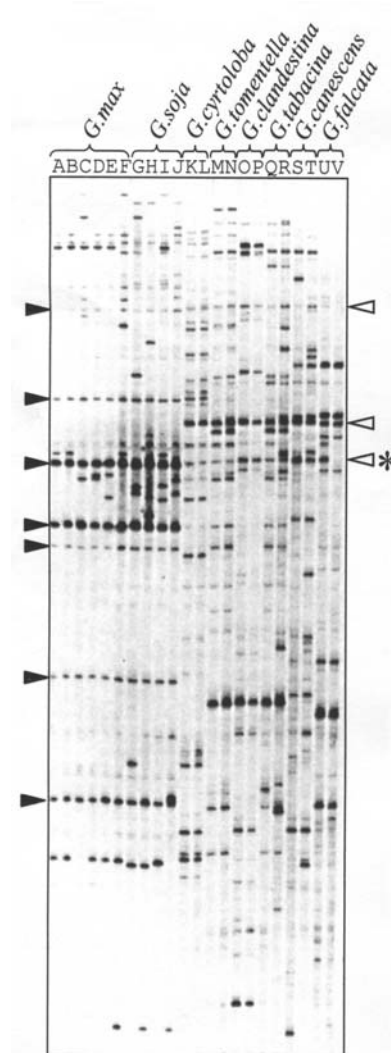


Fig. 1. SIRE-1 SSAP of *Glycine* species and cultivars. Arrows indicate shared bands in annual (black) and perennial (white) species.

### Phylogenetic analysis of *Glycine* species

A comparison of the composition of the SIRE-1 populations in each of the species was produced using banding frequencies using Nei's genetic distance [11] with powermarker [12] and is shown in Fig. 2.

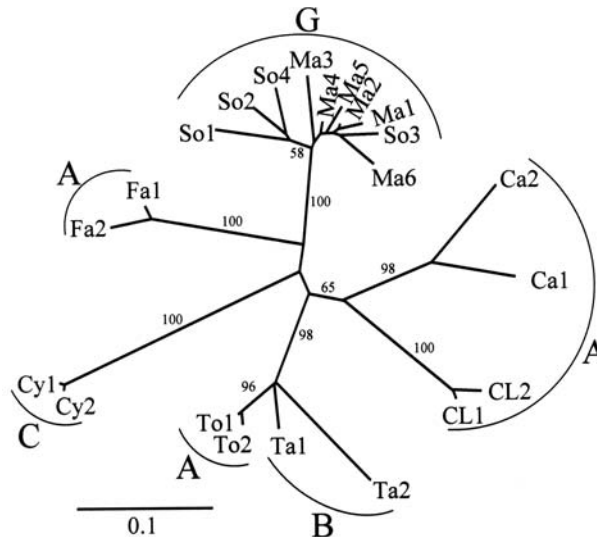


Fig. 2. Phylogenetic bootstrap analysis of *Glycine* SSAP data. Genome groups are shown by capital letters. Scale refers to distance units. Species and cultivar identifiers are listed in Tab. 1.

The annual species (*G. max* and *G. soja*) have very similar SIRE-1 populations which is in contrast to the much more diverse populations among the perennial species. Although only two plants were chosen for each perennial species no species contains identical SIRE-1 population profiles. This indicates that there could be relatively high levels of SIRE-1 sequence evolution in some of the species, with the greatest divergence being noted for *G. canescens*. The brackets and letters in Fig. 2 indicate the genome groups to which each species has been assigned [3] and interestingly there is no correlation between the branching of the tree and the membership of a particular genome group. The closest links identified by SSAP analysis between the perennial species are between *G. tomentella* and *G. tabacina* which are reported to belong to the A and B genome groups. The high levels of confidence with which the individual species are separated on the tree shows that the populations have evolved differently in each of the species since they evolved from a common ancestor and highlights the potential of SIRE-1 SSAP as a method of species identification and diversity analysis in *Glycine*.

## DISCUSSION

Previous studies have shown that SIRE-1 is a highly unusual retroelement as although it has a gene order similar to *Ty1-copia* retrotransposons it has an envelope gene characteristic of retroviruses [14]. This, and the relative homogeneity of the SIRE-1 population in soybean has led to the suggestion that SIRE-1 may be a fairly recent acquisition to the soybean genome [5]. The findings presented here show that populations of SIRE-1 sequences are present throughout the species of the *Glycine* genus. SSAP analysis reveals that some element subgroups are shared between both annual and perennial species indicating that these elements are likely to represent ancestral forms of the element which have been present throughout the evolution of the genus although many new forms of SIRE-1 have clearly arisen during the course of evolution.

The relatively high level of shared bands in the annual species confirms the close relationships of these species. The sequence evolution of the SIRE-1 element within this fairly closely related material is shown by several unique SIRE-1 sub-classes and may indicate that this element is evolving rapidly and may therefore be transpositionally active in these plants. There are only very few examples of SIRE-1 elements which are shared between all perennial and annual species, and there are a lower number of shared bands generally within the perennial compared to the annual species. Although it is possible that a few universally shared bands may have arisen independently in the species through size homoplasy, this is unlikely to have happened independently in all accessions studied. The absence of large numbers of universally shared bands across the genus provides evidence for the antiquity and the wide diversity represented within the *Glycine* genus and the relatively rapid evolution of retrotransposon sequences.

The high proportion of polymorphic bands within the study population as well as highlighting the diversity of the perennial species also suggest that SIRE-1 rather than being evolutionarily young has been present, active and rapidly evolving in all species throughout the evolution of the *Glycine* genus. Further study of this transposable element should therefore prove a useful tool in the study of this complex and productive genus.

## REFERENCES

1. Doyle, J.J., Doyle, J.L., Rauscher, J.T. and Brown, A.H.D. Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus *Glycine*) **New Phytol.** 161 (2003) 121-132.
2. Ahrent, D.K. and Caviness, C.E. Natural cross-pollination of twelve soybean cultures in Arkansas. **Crop Sci.** 34 (1994) 376-378.
3. Hymowitz, T., Singh, R.J. and Kollipara, K.P. The genomes of *Glycine*. **Plant Breed. Rev.** 16 (1998) 289-317.

4. Laten, H.M. Phylogenetic evidence for Ty1-*copia* -like endogenous retroviruses in plant genomes. **Genetica** 107 (1999) 87-93.
5. Laten, H.M., Havecker, E.R. and Voytas, D.F. SIRE-1, an endogenous retrovirus family from *Glycine max*, is highly homogeneous and evolutionarily young. **Mol. Biol. Evol.** 20 (2003) 1222-1230.
6. Gribbon, B.M., Pearce, S.R., Kalendar, R., Schulman, A., Paulin, L., Jack, P., Kumar, A. and Flavell, A.J. Phylogeny and transpositional activity of Ty1-*copia* group retrotransposons in cereal genomes. **Mol. Gen. Genet.** 261 (1999) 883-891.
7. Pearce, S.R., Knox, M., Ellis, T.H.N., Flavell, A.J. and Kumar, A. Pea Ty1-*copia* group retrotransposons: Transpositional activity and use as molecular markers to study genetic diversity in *Pisum*. **Mol. Gen. Genet.** 263 (2000) 898-907.
8. Tam, S.M., Mhiri, C., Vogelaar, A., Kerkveld, M., Pearce, S.R. and Grandbastien, M-A. Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. **Theor. Appl. Genet.** 110 (2005) 819-831.
9. Waugh, R., McLean, K., Flavell, A.J., Pearce, S.R., Kumar, A., Thomas, B.B.T. and Powell, W. Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence specific amplification polymorphisms (SSAP). **Mol. Gen. Genet.** 253 (1997) 687-694.
10. Pearce, S.R., Stuart-Rogers, C., Kumar, A. and Flavell, A.J. Rapid isolation of plant Ty1-*copia* group retrotransposons LTR sequences for molecular marker studies. **Plant J.** 19 (1999) 1-7.
11. Nei, M. and Takezaki, N. Estimation of genetic distances and phylogenetic trees from DNA analysis. **Proc. 5<sup>th</sup> World Cong. Genet. Appl. Livestock Prod.** 21 (1983) 405-412.
12. Liu, K. and Muse, S.V. PowerMarker: Integrated analysis environment for genetic marker data. **Bioinformatics** 21 (2005) 2128-2129.
13. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. **Evolution** 39 (1985) 783-791.
14. Peterson-Burch, B.D., Wright, D.A., Laten, H.M. and Voytas, D.F. Retroviruses in plants? **Trends Genet.** 16 (2000) 151-152.