Theor Appl Genet (1994) 89: 895–899

C. González · N. Henriques-Gil

Genetics of seed storage proteins in the love tree *Cercis siliquastrum* L. (Fabaceae)

Received: 25. November 1993 / Accepted: 17. May 1994

Abstract The patterns of *C. siliquastrum* seed storage proteins ("cercins") are described using SDS-polyacrylamide gel electrophoresis. The polypeptides detected had very different molecular weights (ranging from 168 to 34 KDa) which, together with their high homogeneity, produced a very good resolution of bands. These proteins could be ascribed to five different loci. The analysis of seed sets of individual trees indicated that the love tree is almost completely autogamous with less than 5% of outcrosses. Although this mode of reproduction seems to produce a decrease in heterozygote frequency among the seeds of the population analysed, the levels of variability detected were very high for an autogamous plant: all of the loci were polymorphic, with a mean heterozygosity of 0.327 and a polymorphic index of 0.412. Protein segregation revealed a strong genetic linkage between three cercin loci (*a*, *c* and *d*) while the other two are independent.

Key words *Cercis siliquastrum* · SDS-PAGE
Seed proteins · Polymorphism · Genetic linkage

Introduction

The love tree or Judas tree (*Cercis siliquastrum* L.) originated from the Eastern Mediterranean and Middle East. Since antiquity this tree has been cultivated over a wide range in southern Europe, mainly because its spectacular early spring flowering (Navés 1992). The genus *Cercis* is distributed widely throughout Europe, Asia and North America, and consists of at least five other species (Raulston 1990). In addition to its value as an ornamental plant, many parts of this plant are also edible (Sánchez-Monge 1991), a feature that may be very desirable in these kind of species because of their contribution to wildlife maintenance in urban forestry (Miller 1988).

Little is known about the seed storage proteins in woody material, although some information is available on gymnosperms (Jensen and Berthold 1989; Jensen and Lixue 1991; Gerber et al. 1993) and on genera *Castanea* and *Quercus* (Collada et al. 1991). These proteins and their genes have been more intensively studied in crop species including a number of legumes because of their use in animal and human diets. Major storage proteins of legume seeds have been traditionally referred to as 11S and 7S proteins (or legumins and vicilins), however the use of different methods of purification indicates that their classification may certainly be much more complex (for review, see Casey et al. 1986). While the storage proteins have therefore an importance per se, as a set of genetically controlled products, their analysis may also provide valuable data about the variability (Tucci et al. 1991), reproduction (Porceddu et al. 1980) and evolution (Wolff, 1980, Krochko et al. 1990) of a given plant material. In the study described here, we analysed the variation in seed storage proteins in the love tree by means of SDS-polyacrylamide gel electrophoresis and the genetic segregation relationships of the correspondent loci.

Materials and methods

One single seed per plant was collected from each of 42 love trees from the Retiro Park in Madrid (Spain) for population analysis. An additional variable number of seeds (from 100 to 200) was taken from 17 *C. siliquastrum* plants, 5 of which belonged to the same Retiro Park population while the remainder were collected in the following localities: Albarreal de Tajo, Toledo (3), Alicante (1), Caia, Portugal (1), and, within Madrid province, at the University Campus (3), Casa de Campo (1) and El Escorial (3). Sampling was carried out during the autumn and winter of 1991–1992.

For protein extraction, the seeds were cut in approximately two identical parts, and a piece of cotyledon of about 2–3 mg was crushed inside an eppendorf tube and suspended in 80 µl of a buffer consisting of 2% SDS, 6.25% TRIS HCl pH 6.8, 10% glycerol, 0.01% pyronine Y and 5% 2-mercaptoethanol; this latter substance was not employed in some extractions for the detection of disulfide-linked

Communicated by P. M. A. Tigerstedt

C. González (✉) · N. Henriques-Gil
Laboratorio de Genética,
Centro de Enseñanza Superior San Pablo CEU Montepríncipe,
Boadilla del Monte, 28660 Madrid, Spain

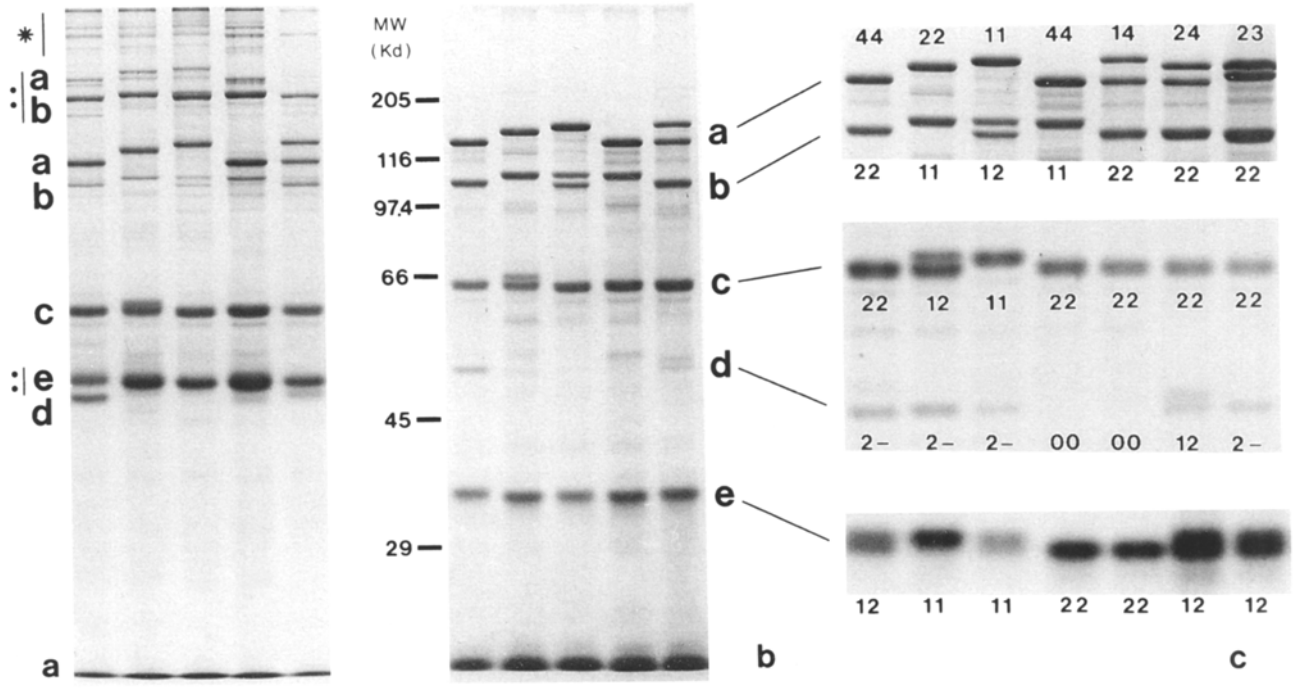

Fig. 1a–c Protein SDS-PAGE patterns obtained from *C. siliquastrum* seeds. **a** Unreduced extracts (the dimers are marked with a *double point* and the polymeric protein forms with an *asterisk*). **b** extracts of the same 5 seeds as in **a** but in the presence of 2-mercaptoethanol (the molecular weights are given in kDa). **c** different genotypes for each of the five loci described (*a–e*) showing all the allelic forms found.

polymers. After 1 hour with occasional shaking, the tubes were heated to 100°C for 1.5 min.

Polypeptide separation was made in 10% polyacrylamide gels plus SDS (SDS-PAGE) as described by Payne et al. (1981) with an electrode buffer of 14% glycine, 3% TRIS and 1% SDS in distilled water. The gels were run at a constant voltage of 60 V for half an hour and then at 50 V for 14 h. Staining was carried out in a solution of 0.02% Coomassie Brilliant Blue R, 5% ethanol and 6% TCA for about 20 h.

Results

The majority of the seed proteins of *Cercis siliquastrum* are salt soluble globulins; consequently, acid or apolar media resulted in a very poor protein extraction (unpublished results). Fig. 1 shows a typical SDS-PAGE of total seed protein extracts of *C. siliquastrum*, which, to coincide with the nomenclature employed for other legume species, we call "cercins". The bands observed in the gels can be ascribed to five loci-*a*, *b*, *c*, *d* and *e* -which codify for polypeptides of 168–142, 104–100, 72–68, 54–52 and 36–34 KDa respectively. This conclusion is supported by the fact that the offspring of completely homozygous plants (that is, those that do not segregate for any seed protein) invariably show these five monomers.

In the material analysed, variability was detected for all of the five loci described (Fig. 1c). The highest levels of variation were found in locus *a* for which 4 different alleles were determind: *a1*, *a2*, *a3*, and *a4*, in a descending polypeptide molecular weight, (from 168 to 142 KDa; that is, a difference of about 200 amino acids). Loci *b*, *c* and *e* showed two alleles (*b1-b2*, *c1-c2*, *e1-e2*), while locus *d* had an additional third null allele (*d0*). The frequencies of the different phenotypes among the 42 seeds from the population sample are given in Table 1. All those variations cannot depend on additional loci. If, for instance, *c1* and *c2* were codified by two different loci, then we should expect some cases of homozygous trees for both of them, which in turn would give rise to seeds always showing the two proteins. This was never observed and, in fact, the existence of several *c1c2* seeds was seen to be associated with both *c1c1* and *c2c2* full-sibs (see Fig. 2), thereby supporting their allelic nature. The same applies to the other proteins. Moreover, since one or two bands may appear in a given seed when the polypeptides of higher molecular weight are considered, *a1–a4* must be allelic variants (if more than one locus existed then we could find three or more bands in the same seed).

Figs. 1a and b correspond to electrophoregrams of proteins of the same seeds extracted in the absence and in the presence, respectively, of 2-mercaptoethanol. The variability described above enabled identification of the different genotypes using reduced or unreduced extracts. The relative mobility of a given protein in both cases will indicate the existence of disulphide bonds. For example, polypeptides *c* and *d* appeared in monomeric form and *e* formed a disulphide-linked dimer with an apparent double molecular weight . In vivo, the products of loci *a* and *b* seemed to adopt a less strict status as the major portion of the *a* proteins were found as monomers although dimers also existed; in contrast, the *b* dimer was more frequent than its corresponding monomer. A higher mobility of *c* and *d* pro-

Table 1 Number of seeds showing each possible genotype for the different storage protein loci among the total sample of 42 seeds from the population of Retiro

Locus	*Phenotype*									
	11	*12*	*13*	*14*	*22*	*23*	*24*	*33*	*34*	*44*
a	6	3	0	2	14	6	4	2	1	4
		11		*12*		*22*				
b		3		13		26				
c		1		8		33				
e		14		18		10				
		00		*1–*		*2–*		*12*		
d		30		1		9		2		

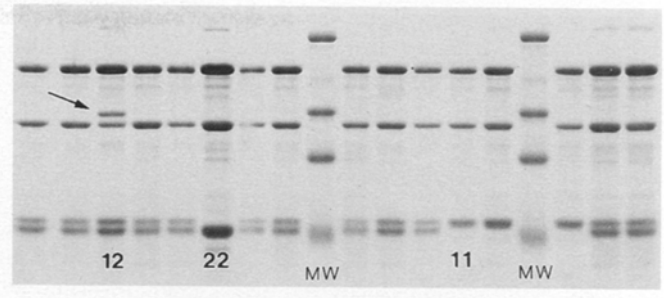

Fig. 2 Protein patterns of 16 seeds from the offspring of a *a2a2 b2b2 c1c2* tree (for locus *c* an example for each genotype is indicated). One seed showed the *b1* polypeptide (*arrowed*), indicating an outcross origin. [*MW* Molecular weight markers of 205, 116, 97.4 and 66 kDa (*top to bottom*)]

teins was obtained in unreduced extracts, most probably a result of an intrachain disulphide bond. Additionally, a number of different polymers where observed, but their correspondence to one or more particular loci was not clear.

The maternal genotypes of each of the 17 families collected were assessed in an electrophoretic screening of a variable number of seeds per family (10–20). When all the offspring contained a given allele the mother was considered to be homozygous; in contrast, from a heterozygous 12 mother genotype 11, 12 and 22 seeds may be found (see Fig. 2).

Mating system

This study was restricted to the 5 offspring obtained in the Retiro Park because this population of *C.siliquastrum* is quite large (more than 100 trees). The variability observed may be used to estimate the selfing and outcrossing rates of the love trees: the presence in a seed of a polypeptidic variant whose allele is not present in the mother plant necessarily identifies an outcrossing origin (see Fig. 2). From a total of 50 seeds per tree, only 2 were produced by cross-fertilization among the offspring of each of 3 trees; in the remainder no alien alleles could be detected. Using the method of Shaw et al. (1981) we obtained outcrossing rates of 4.1and 4.2% for the two former cases and, evidently, 0% for the other 2.

Segregation analysis

Four plants were diheterozygous (*a1a2 b1b2*, *c1c2 e1e2* and 2 *a1a2 c1c2*) and an additional one was heterozygous for three seed protein loci (*a1a2 c1c2 d1d2*). Their offsprings were the subject of a further detailed electrophoretic screening in order to detect a possible linkage relationship between loci. The segregations obtained and the genetic distances between loci, estimated following the maximum likelyhood method, are given in Table 2. It was evident that loci *a*, *c* and *d* show a strong linkage with a very low number of recombinants, while pairs *a-b* and *c-e* segregate independently.

Table 2 Segregations obtained from diheterozygous plants. χ^2 values were calculated from a (1:2:1)×(1:2:1) contingency table; when significant ($P< 0.05$; 4 *df*), the genetic distances were estimated following the maximum likelyhood method

Seed *genotype*		Plant and maternal genotype				
		CS15	CS1,2,3	CS2	CS2	CS12
		a1b1 a2b2	*a1c2* a2c1	*a1d1* a2d2	*c2d1* c1d2	*c1e1* c2e2
11	11	7	0	16	0	6
11	12	14	0	4	0	14
11	22	6	39	0	33	9
12	11	14	0	0	0	16
12	12	25	91	45	45	36
12	22	20	1	9	9	15
22	11	7	55	0	16	4
22	12	13	2	0	4	6
22	22	6	0	33	0	3
Independence χ^2		1.78 ns	358.85 $p< 10^{-6}$	150.13 $p< 10^{-6}$	150.13 $p< 10^{-6}$	1.21 ns
Distance		–	0.8 cM ± 0.5	6.3 cM ± 1.8	6.3 cM ± 1.8	–

Discussion

The seed proteins of *C. siliquastrum*, with respect to solubility are typical of those of legume species. Legume globulins, however, usually show a high heterogeneity of polypeptide forms, even for the same genotype (Casey et al. 1986). In contrast, there are a comparatively low number of protein variants in *Cercis,* which results in a good resolution of bands in the proteinograms (Fig. 1). The only heterogeneity that we found was related to disulphide-linked polymers, which can show a considerable range in the number or types of aggregated units, given the different bands of slow migration from unreduced extracts (Fig. 1b). More precise knowledge of the structure of the native proteins obviously requires further biochemical analysis. In *Vicia faba* up to 29 disulphide-linked subunit pairs have

been detected (Tucci et al. 1991): it seems that the rules for assembly of native proteins are not well defined and that random association might occur in many instances (Casey et al. 1986). It is important to note that the data available on seed proteins in Fabaceae species comes from crops, included in the subfamily Papilionoideae, while the genus *Cercis* belongs to the subfamily Cæsalpinioideae (often considered an independent family from that of true legumes).

From Table 2 it is evident that loci *a*, *c* and *d* form one linkage group, with *b* and *e* being independent of it (*a-b* and *c-e* segregations tested). The order of those three loci cannot be precisely determined because no *a-c* recombinants appeared in family CS2 (where *a*, *c*, and *d* segregated), and so an identical genetic distance to locus *d* was obtained (6.3 ± 1.8 cM). This strong linkage between three out of the five seed storage protein loci is remarkable and could suggest either a common origin or regulation. In other legume species like *Pisum* (Casey 1979), the french bean (Brown et al. 1981) and the soybean (Schuler et al. 1982), the existing multigene families for seed proteins indicate an evolution by a series of gene duplications (Higgins 1984). Unfortunately, this could not be analysed in the present study since we had no other data than the molecular weight of the gene protein products.

Due to a relatively slow petal maturation with respect to that of both the stamen and ovary in the love tree, this species has been considered to have an autogamous mode of reproduction. The offsrings that we analysed indicate that these trees are almost exclusively self-fertilizing. Obviously, trees isolated several kilometers from any other *Cercis* specimen have an extremely low opportunity for cross-fertilization and, in fact, segregations adjusted to self-fertilization-expected proportions, with an apparently normal fertility. Murawski and Hamrick (1991) in an isoenzymatic analysis of several tree species, including two Fabaceae, proved that the rates of autogamy were inversely related to the population density. However, in the love tree, the high selfing rate seems to have only a physiological basis: the Retiro Park *Cercis* population consists of more than 100 trees (within 2 km^2), and the autogamy rates are very high.

The fact that *C. siliquastrum* is a mainly autogamous species contrasts strongly with the levels of variability found, since self-fertilization should lead to a decrease in heterozygote frequency. Although our sample was quite small (42 seeds), the variability affected each of the five seed protein loci, including different alleles (2.6 per locus), with a polymorphic index (as defined by Hamrick 1979) of PI=0.412. According to Loveless and Hamrick (1984), long-lived species have an increased genetic variation within populations because the loss of variability is retarded while the population size and mating opportunities are increased. Nevertheless, the PI found for *C. siliquastrum* was even higher that the mean PI obtained for woody plants (0.354; Hamrick 1979). The high autogamy of *Cercis* could be the reason why that only 16 seeds were heterozygous for locus *a*, where 4 different alleles exist (Table 1), in contrast to with the 28 expected on the basis of the observed allelic frequencies and random mating. (For the remainder loci, the expected numbers, though always higher, were very similar to the observed ones.) The variability was still high among the seed sample (the mean heterozygosity was H=0.327) because the tree population was very polymorphic.

Our data are based on only five loci, all of which have a precise and similar function, that is, codifying for seed storage proteins. The significance of the above variables will depend on the representation of such genes on the total genome. It should be pointed out that the variability measured is based on polypeptide molecular weight differences so most probably it is only an infra-estimation of the true levels of variation in protein sequences.

Although *C. siliquastrum* has reproduced in Spain for many centuries we must keep in mind that it is not an autoctonous species in the Iberian Peninsula (Navés 1992), and it has never been subject of selective breeding programmes. It would be very interesting to know if, in its places of origin, the ecological features would allow higher rates of cross-fertilization. The a high amount of genetic variation within this autogamous tree remains an open question.

Acknowledgements The authors are indebted to Ms. A. Miner and Ms. V. González for their excellent assistance in the collection of the material and to Ms. C. Garma in the laboratory analyses.

References

Brown JWS, Ma Y, Bliss FA, Hall TC (1981) Genetic variation in the subunits of globulin-1 storage protein of the french bean. Theor Appl Genet 59:83–88

Casey R (1979) Genetic variability in the structure of the a-subunits of legumin from *Pisum*. A two-dimensional gel electrophoresis. Heredity 43:265–272

Casey R, Domoney C, Ellis N (1986) Legume storage proteins and their genes. Oxford Surv Plant Mol Cell Biol 3:1–95

Collada C, Caballero RG, Casado R, Aragoncillo C (1991) Seed storage proteins in Fagaceae: similarity between *Castanea* globulins and *Quercus* glutelins. Plant Sci 75:145–154

Gerber S, Rodolphe F, Bahrman N, Baradat Ph (1993) Seed-protein variation in maritime pine (*Pinus pinaster* Ait.) revealed by two-dimensional electrophoresis: genetic determinism and construction of a linkage map. Theor Appl Genet 85:521–528

Hamrick JL (1979) Genetic variation and longevity. In: Solbrig OT, Jain S, Jonhson Gb and Raven PH (eds) Topics in plant population biology. Columbia University Press, New York pp 84–113

Higgins TJV (1984) Synthesis and regulation of major proteins in seeds. Annu Rev Plant Physiol 35:191–221

Jensen U, Berthold H (1989) Legumin-like proteins in gymnosperms. Phytochemistry 28:1389–1394

Jensen U, Lixue C. (1991) *Abies* seed protein profile divergent from other Pinaceae. Taxon 40:435–440

Krochko JE, Charbonneau MR, Coulter KM, Bowley SR, Bewley JD (1990) A comparison of seed storage proteins in subspecies and cultivars of *Medicago sativa*. Can J Bot 68:940–948

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. Annu Rev Ecol Syst 15:65–95

Miller RW (1988) Urban Forestry. Prentice-Hall, N.J.

Murawski DA, Hamrick JL (1991) The effect of the density of flowering individuals on the mating systems of nine tropical tree species. Heredity 67:167–174

Navés F (1992) El árbol en jardinería y paisajismo. Omega, Madrid

Payne PI, Corfield KG, Holt LM, Blackman JA (1981) High-molecular-weight subunits of glutenin and bread-making quality in progeny of six crosses of bread wheat. J Sci Food Agric 32:51–60

Porceddu E, Monti L, Frusciante L, Volpe N (1980) Analysis of cross pollination in *Vicia faba*. Z Pflanzenzucht 84:313–322

Raulston JC (1990) The redbud. Am Nur 171:39–51

Sánchez-Monge E (1991) Flora Agrícola I. M. Agricultura, Pesca y Alimentación. Madrid

Schuler MA, Schmitt ES, Beachy RN (1982) Closely related families of genes code for the a and a' subunits of the 7S storage protein complex. Nucleic Acids Res 10:8225–8244

Shaw DV, Kahler AL, Allard RW (1981) A multilocus estimator of mating system parameters in plant populations. Proc Natl Acad Sci USA 78:1298–1302

Tucci M, Capparelli R, Costa A, Rao R (1991) Molecular heterogeneity and genetics of *Vicia faba* seed storage proteins. Theor Appl Genet 81:50–58

Wolff G (1980) Investigations on the relations within the family papilionaceae on the basis of electrophoretic banding patterns. Theor Appl Genet 57:225–232