# GENETIC STUDIES ON SUNFLOWER SEED STORAGE PROTEINS

**Irina Anisimova**, N.I.Vavilov Institute of Plant Industry,
42 Bolshaya Morskaya str., St.Petersburg, Russia
Fax: +7 812 311 8762; e-mail: bekryaev@aari.nw.ru

**Alexander Konarev**, All-Russian Institute of Plant Protection,
3 Shosse Podbelskogo, 189620 190000 St.Petersburg-Pushkin, Russia

**Saida Turkav**, All Russia Institute of Oil Crops,
17 Filatova str., Krasnodar, Russia
Fax: +7 861 257 7914

**Vera Gavrilova**, N.I.Vavilov Institute of Plant Industry,
42 Bolshaya Morskaya str., 190000 St.Petersburg, Russia
Fax: +7 812 311 8762; e-mail: vir@glasnet.ru

**Peter Shewry**, IACR-Long Ashton Research Station, University of Bristol,
Bristol BS41 9AF, UK
Fax: +44 01275 394 299; e-mail: pat.baldwin@bbsrc.ac.uk

## Summary

During development sunflower seeds accumulate abundant amounts of storage proteins having biological and ecomonic importance. A polymorphism revealed within major groups of polypeptides, the 11S globulin (helianthinin) and 2S albumin is defined by variability in corresponding multigenic families. By segregation analysis in $F_2$ and backcross hybrid generations the Mendelian loci for helianthinin (*Hel*A - for subunit A, $M_r$ 56,000, *Hel*B - for subunit B, $M_r$ 55,000 and *Hel*C - - for subunit C, $M_r$ 52,000) and the locus for methionine-rich albumin *SFA*8 ($M_r$ 10,000) have been identified. In each locus a number of polymorphic alleles has been revealed. A segregation pattern for helianthinin polypeptide composition in the $F_2$ of a cross wild annual $\times$ cultivated sunflower indicated a single allele difference between the two forms. A distribution of polymorphic helianthinin alleles among Russian varieties and common inbred lines is presented. A high frequency of the "wild type" helianthinin alleles *hel*9 and *hel*11 within certain variety populations and also there occurrence in some inbred lines (for example HA61, RHA273, RHA274) indicate close relationships between domesticated sunflower gene pool and wild annual species. A linkage of the helianthinin loci *Hel*B and *Hel*C was proven by the analysis of three $F_2$ and $F_a$ generations; the recombination frequency values constituted between 19 and 24%. Other marker loci tested (*Hel*A, *SFA*8, *Est*1, *Gpi*1, *PGD*, *TI*m, *Vs*, *Ep*, *P*) segregated independently of the *Hel*B and *Hel*C loci. No linkage was also found between the *SFA*8 locus and the loci *Hel*B, *Hel*C and *Vs*. A family of methionine-rich albumins SFA7/SFA8 was found only in seeds of annual *Helianthus* species but was absent from seed of perennial ones. A high quantitative variation for the levels of SFA7/SFA8 proteins was observed among inbred lines and also in separate seeds in segregating populations.

## 1.Introduction

In addition to lipids, sunflower seeds accumulate significant amounts of storage proteins, accounting for a high proportion of the residue after oil extraction. The major protein components of sunflower seed belong to the two groups, salt-soluble 11S globulins (helianthinin) and water soluble 2S albumins, in relative amounts of about 2:1. Helianthinin is a legumin-like globulin with molecular mass of about 305,000 (Schwenke *et al.* 1979). It is composed of six subunits each consisting of disulfide-linked acidic ($\alpha, \alpha'$) and basic ($\beta, \beta'$) polypeptides. The charge and molecular mass heterogeneity of subunits and polypeptides (Dalgalarrondo *et al.* 1984, Anisimova *et al.* 1991) suggest that helianthinin is the product of multigenic family. Two cDNA clones which probably represent two divergent helianthinin subfamilies have been isolated and characterized (Vonder Haar *et al.* 1988). The 2S albumins of sunflower seed exist as a heterogeneous mixture of single chain polypeptides with $M_r$ 10,000 - 18,000, most, if not all, having intra-chain disulfide bonds (Kortt and Caldwell 1990, Anisimova *et al*. 1995). Two methionine-rich albumins (SFA7 and SFA8) have nearly identical amino acid composition and were established to constitute of about 7% of the total protein in the seeds of the variety Hysun (Kortt and Caldwell 1990). One of the 2S albumin genes which have been so far isolated and sequences represented the single chain methionine-rich albumin SFA8 (Kortt *et al*., 1991). Another known gene, HaG5 (Allen *et al*., 1987, Brunel, 1994) was shown to be a member of small multigenic family consisting of at least two divergent genes. The precursor proteins encoded by these genes are presumed to undergo extensive proteolytic processing giving rise to two different polypeptides which are not disulfide linked but each of which contain disulfide bonds. The undefatted protein extract of seeds also contains major polypeptides of $M_r$ 19,000 and 20,000 representing oil body membrane proteins, or oleosins which are well characterized at the molecular level (Laccy *et al.* 1998).

Screening the sunflower gene pool has revealed polymorphism of polypeptide composition both in the helianthinin (Anisimova *et al.* 1991) and 2S albumin (Anisimova *et al.* 1995) fractioins. However, the genetic basis of such variability is not well understood. No data is available on inheritance of polypeptide composition. Two objectives were pursued in the present paper: to demonstrate genetic variability in major sunflower seed proteins and to study inheritance of protein variants in different cross combinations.

## 2.Materials and methods

The materials examined comprised 188 inbred lines, 10 interline hybrids (including $F_2$ and backross segregating populations), 23 varieties, 20 ecotypes of wild *Helianthus* species and 13 interspecific hybrid progenies. They originated from the Sunflower World Collection (Vavilov Institute of Plant Industry, St.Petersburg, Russia), All-Russia Institute of Oil Crops (Krasnodar, Russia), Institute of Genetics (Sofia, Bulgaria), University of Birmingham (United Kingdom), University of Giessen (Germany).

To facilitate studies on protein polymorphism and inheritance the modified procedures for isolating helianthinin and 2S albumins from small cotyledon sections (of 2-4 mg weight) have been developed. With the use of this procedure a rest of a seed can be kept for growing plant or analysis of other characters (helianthinin, isozymes etc.). The 11S globulin (helianthinin) fractions were obtained by cryioprecipitation with cold water from salt extracts as described earlier (Anisimova *et al.* 1991). For isolating 2S albumin fractions a modified method of acetone precipitation (Kortt and Caldwell 1990) was used. Electrophoresis of helianthinin (SDS-PAGE) was carried out in a Tris/glycine system following the procedure described in our previous paper (Anisimova *et al*. 1991). The 2S albumins were fractionated in a Tris/tricine

electrophoretic system or separated using high performance liquid chromotography (RP-HPLC) as described by Anisimova *et al*. (1995).

## 3.Results and Discussion

Inheritance studies with using segregating hybrid populations of H.annuus have revealed several major electrophoretic variants of helianthinin polypeptides. The allelism analysis indicated that they are the products of polymorphic alleles in three Mendelian loci (Table 1).

A polymorphism for alleles *hel*9 and *hel*11 was observed within wild *H.annuus* ecotypes. Segregation analysis of crosses between wild *H.annuus* and several inbred lines has demonstrated that the helianthinin gene family of domesticated sunflower differed from that of wild ecotypes by a single allele (*hel*12) at the *Hel*C locus.

TABLE 1. Mendelian loci encoding polymorphic variants of sunflower (*H.annuus*) seed storage protein

| Locus | Polypeptide | Alleles* | |
|---|---|---|---|
| | | Normal | Variant |
| *Hel*A | helianthinin subunit A $M_r$ 56,000 | *hel*33* | *hel*34, *hel*14$^{nul}$ |
| *Hel*B | helianthinin subunit B $M_r$ 55,000 | *Hel*30 | *hel*29 |
| *Hel*C | helianthinin subunit C $M_r$ 52,000 | *Hel*12 | *hel*9, *hel*11, *hel*12$^{nul}$ |
| *SFA*8 | methionin-reach albumin SFA8 $M_r$ 10,000 | *SFA*8$_n$ | *SFA*8$_v$ |

*Helianthinin alleles are designated in accordance with polypeptide indices in electrophoretic banding pattern (Anisimova *et al*. 1996).

Most varieties examined possessed characteristic helianthinin alleles which occured with frequencies 1-68% in random samples of 100 seeds (Table 2). The *hel*34 allele (locus *Hel*A) was found to be characteristic of only the variety Armavirets and showed very limited distribution among various inbred lines. The *hel*29 allele (locus *Hel*B) occurred in the varieties Gigant 549 and Pervenets but was rather common among inbred lines. The *hel*9 and *hel*11 alleles (locus *Hel*C) showed wide distribution within the sunflower gene pool. They were present both in different inbred lines and in most varieties. The varieties VNIIMK1646, Krasnodarets and Start lacked these alleles but could be distinguished on the base of other polymorphic helianthinin variants (data not shown). The varieties Yenisei, Lider, Pervenets, Pochin, Skorospelyi and Yugo-Vostochnyi all possessed both alleles *hel*9 and *hel*11 (with the highest frequency of occurrence being in Yenisei) whereas other varieties had either *hel*9 (Peredovik, Kavkazets, Progress) or *hel*11 (Berezanskii, Voskhod, Zelenka368, Zenit, Lider, Nadeoyzhnyi).

A number of facts indicate that «wild» allele *hel*9 is probably present in the suppressed state in the cultivated sunflower genome («sleeping» allele) and can be repressed, for example, as a result of interspecific crosses. We examined a number of hybrid progenies between *H.annuus* inbred lines and perennial species which have been produced *via* «embryo rescue» (Krauter *et al*. 1991) or by conventional field technique. All the progenies, regardless of their origin, expressed helianthinin of the maternal («annual») phenotype. In certain progenies, however, the allele *hel*9 was expressed rather than the maternal type allele *hel*12. This fact can be explained by a possible de-suppression of an «ancient» *hel*9 allele present in the genotype of the maternal line in the «sleeping» state.

Eighty two inbred lines out of 188 analysed (~43%) possessed characteristic helianthinin polypeptides (data not shown). It should be noted that the *hel*29 allele was the most frequent among inbred lines originated in VIR whereas in a group of 14 common inbreds the alleles hel9 and *hel*11 were the most frequent that indicates their relation with wild annual gene pool (Table 3).

Up to 12 individual polypeptides have been identified in the 2S albumin fraction using SDS-PAGE and RP-HPLC analyses. According to their N-terminal sequences, the 2S albumin polypeptides of sunflower can be classified into three groups: the methionine-rich albumins

TABLE 2. Distribution of some polymorphic helianthinin alleles among varietal populations

| Variety | Breeder | Frequency of occurencies of alleles** | | | | | |
|---|---|---|---|---|---|---|---|
| | | *hel*9 | *hel*11 | *hel*29 | *hel*34 | *hel*4$^{nul}$ | *hel*12$^{nul}$ |
| Peredovik | VNIIMK | 12 | 0 | 0 | 0 | 0 | 0 |
| Gigant 649 | VIR | 0 | 0 | 68 | 0 | 100 | 0 |
| Armavirets | VNIIMK | 0 | 0 | 0 | 4 | 0 | 0 |
| Berezansky | VNIIMK | 0 | 6 | 0 | 0 | 0 | 0 |
| VNIIMK 1646 | VNIIMK | 0 | 0 | 0 | 0 | 0 | 0 |
| Voskhod | VNIIMK | 0 | 4 | 0 | 0 | 0 | 0 |
| Yenisei | KAI | 9 | 19 | 0 | 0 | 0 | 0 |
| Zarya | VNIIMK | 0 | 0 | 12 | 0 | 0 | 0 |
| Zelenka 368 | VNIIMK | 0 | 4 | 0 | 0 | 2 | 0 |
| Zenit | VNIIMK | 0 | 24 | 0 | 0 | 0 | 0 |
| Kavkazets | VNIIMK | 2 | 0 | 0 | 0 | 0 | 0 |
| Krasnodarets | VNIIMK | 0 | 0 | 0 | 0 | 0 | 0 |
| Lider | VNIIMK | 8 | 3 | 0 | 0 | 0 | 2 |
| Lutch | VNIIMK | 0 | 0 | 0 | 0 | 0 | 2 |
| Nadeozhnyi | VNIIMK | 0 | 22 | 0 | 0 | 0 | 0 |
| Pervenets | VNIIMK | 3 | 1 | 48 | 0 | 4 | 1 |
| Potchin | VNIIMK | 2 | 6 | 0 | 0 | 0 | 0 |
| Progress | VNIIMK | 4 | 0 | 0 | 0 | 0 | 0 |
| Salyut | VNIIMK | 0 | 0 | 0 | 0 | 0 | 1 |
| Skorospelyi | SEAI | 2 | 4 | 0 | 0 | 0 | 0 |
| Start | VNIIMK | 0 | 0 | 0 | 0 | 0 | 0 |
| Yugo-Vostochnyi | SEAI | 4 | 8 | 4 | 0 | 0 | 0 |
| Smena | VNIIMK | 0 | 16 | 0 | 0 | 0 | 0 |

*VNIIMK - All Russia Institute of Oil Crops; KAI (Russia)- Krasnoyarsk Agricultural Institute (Russia); SEAI - South-East Agricultural Institute (Russia)
**Data were obtained for random samples of 100 seeds

SFA7 and SFA8 ($M_r$~10 000), and proteins derived from the N- and C-terminal parts of the HaG5 precursor protein. Variation in the presence or absence of minor albumin polypeptides was observed among the genotypes analysed. A variant form of SFA8 with slightly slower mobility on SDS-PAGE compared to the normal type was identified. This variant was found in only five of 60 inbred lines examined. The two SFA8 forms clearly differed in a pI with the normal form having pI, of about 6.0 and the variant form having a pI of about 6.5. In $F_1$ seeds of a cross between the inbred lines VIR130 (variant SFA8) and VIR104 (normal SFA8) the two forms were expressed co-dominantly while the ratio of phenotypic classes in $F_2$ fitted 1:2:1 indicated that the normal and variant forms of SFA8 are encoded by allelic loci.

Significant variation in the level of SFA8 in the 2S albumin fractions was observed within the sunflower gene pool. SFA8 was absent from seeds of perennial *Helianthus* species, however, but present in both annual species and cultivated sunflower. The combined level of SFA7 and SFA8 and the relative proportions of these proteins varied significantly among

different inbred lines. A considerable heterotic effect for the level of methionine-rich albumins (~50% prvalence over mean values of parental lines) was observed in $F_1$ hybrid seeds of interline crosses. However, no heterotic effect was noted when the $F_2$ seeds were analysed. Moreover, unexpectedly high variation in the amounts of SFA7 and SFA8 was observed among individual $F_2$ seeds. The data suggest that the levels of methionine-rich albumins in sunflower seed are probably under genetic control.

TABLE 3. Occurrence of some polymorphic helianthinin alleles among common inbred lines

| Line | Variant helianthinin alleles | | | | |
|---|---|---|---|---|---|
| | *hel*9 | *hel*11 | *Hel*29 | *hel*34 | *hel*10$^{nul}$ |
| HA6 | - | - | - | + | - |
| HA60 | - | - | + | - | - |
| HA61 | + | - | - | - | - |
| HA89 | - | - | - | - | - |
| HA113 | - | - | - | - | - |
| HA232 | - | - | - | - | - |
| HA234 | - | - | - | - | - |
| RHA265 | - | - | - | - | + |
| RHA273 | + | - | - | - | - |
| RHA274 | + | - | - | - | - |
| RHA282 | - | - | - | - | - |
| CM44 | - | + | - | - | - |
| CM152 | - | - | - | - | - |
| Do264 | + | - | - | - | - |

TABLE 4. Inbred lines used as parents for the development of segregating populations

| Population | Cross | Segregating loci |
|---|---|---|
| 1 | $F_2$(VIR104 × VIR369) | *Hel*C, $TI_m$ |
| 2 | $F_2$(VIR130 × CM4) | *Hel*B, *Hel*C, *Vs* |
| 3-1 | $F_2$(VIR130 × VIR104) family 1 | *Hel*B, *Hel*C, *SFA*8 |
| 3-2 | $F_2$(VIR130 × VIR104) family 2 | *SFA*8, *Vs* |
| 4 | $F_a$(VIR130 × VIR104) × VIR130 | *Hel*B, *Hel*C, *PGD* |

TABLE 5. Segregation ratio for storage protein, isozyme and inhibitor loci in hybrid generation

| Pair of loci | Population | Frequency of phenotypic classes | $\chi^2_L$ | r+$S_r$(%) |
|---|---|---|---|---|
| *Hel*B, $TI_m$ | 1 | 2:4:2:6:22:7:5:8:2 | 3.63* | independent |
| *Hel*B, *Hel*C | 2 | 1:8:16:9:36:4:10:15:1 | 37.36* | 21.8 ± 4.1 |
| *Hel*B, *Hel*C | 3-1 | 1:3:10:3:22:3:10:7:0 | 34.36 | 24.6 ± 5.5 |
| *Hel*B, *Hel*C | 4 | 0:9:31:0:37:7:0:0:0 | 33.14** | 19.0 ± 4.2 |
| *Hel*B, *SFA*8 | 3-1 | 2:6:7:5:17:8:4:6:5 | 2.92* | independent |
| *Hel*C, *SFA*8 | 3-1 | 5:9:3:3:14:11:2:7:5 | 3.76* | independent |
| *Hel*B, *PGD* | 4 | 0:0:0:21:33:0:21:16:0 | 3.10** | independent |
| *Hel*C, *PGD* | 4 | 0:0:0:19:29:0:20:16:0 | 2.00** | independent |

*tested against expected ratio 1:2:1:2:4:2:1:2:1, DF=4
**tested against expected ratio 1:1:1:1, DF=1

Segregation for various protein, isozyme and morphological characters was tested in $F_2$ and $F_a$ populations derived from crosses mentioned in Table 4. Each locus segregated in accordance with expected ratio 1:2:1 or 3:1 in case of $F_2$ or 1:1 in case of $F_a$. No deviations from the Mendelian ratios were observed. The results of segregation analysis have revealed

linkage of the helianthinin loci *Hel*B and *Hel*C (Table 5). The recombination frequencies varied in a range of 19.0-24.6% in different cross combinations indicating localisation of the two genes in the same linkage group. Locus *SFA*8 showed independent segregation of both loci *Hel*B and *Hel*C. The *Hel*B, *Hel*C and *SFA*8 loci segregated independently of the gene *vs* encoding morphological character "raised leaf veins". In the present (Table 5) and our earlier work (Anisimova *et al*. 1996) the independent inheritance of the *Hel*B and *Hel*C loci of the *Hel*A locus and the loci $TI_m$ (main trypsin inhibitor), *Est*1 (esterase), *Gpi* (glucose phosphate isomerase), *PGD* (6-phosphate gluconate dehydrogenase), *Ep* (striped seeds), *P* (armoured pericarp) has been demonstrated.

## 4.Conclusions

The allelic variation revealed in the genetic loci encoding seed storage proteins is a useful tool for comparison of sunflower varieties, inbred lines and wild samples. The data on linkage relationships among loci examined can be used in studies aimed to construction of genetic linkage map of sunflower.

## 5.References


Allen, R.D., Cohen, E.A., Vonder Haar, R.A., Adams, C.A, Ma, D.P., C.L.Nessler C.L., and Thomas, T.L. (1987) Sequence and expression of a gene encoding an albumin storage protein in sunflower, *Mol.Gen.Genet.* **210**, 211-218.

Anisimova, I.N., Gavrilyuk, I.P., and Konarev, V.G. (1991) Identification of sunflower lines and varieties by helianthinin electrophoresis, *Plant Varieties and Seeds* **4**, 133-141.

Anisimova, I.N., Gavrilova, V.A., Loskutov, A.V., and Tolmachev, V.V. (1996) Identification of gene loci encoding for 11S globulin of sunflower seed, in: Proc.14th Int.Sunfl.Conf. (Beijing/Shenyang, China, 12-20 June 1996). **2**, pp.1167-1170.

Anisimova, I.N., Fido, R.J., Tatham A.S., and Shewry, P.R. (1995) Genotypic variation and polymorphism of 2S albumins of sunflower, *Euphytica* **83**, 15-23.

Brunel, D. (1994) Denaturing gradient gel electrophoresis (DGGE) and direct sequencing of PCR amplified genomic DNA: rapid and reliable identification of *Helianthus annuus* L. cultivars, *Seed Sci.Techn.* **22**, 185-194.

Dalgalarrondo, M., Raymond, J., and Azanza, J.L. (1984) Sunflower seed proteins: characterization and subunit composition of the globulin fraction, *J.Exp.Bot.* **35**, 1618-1628.

Kortt, A.A. and Caldwell, J.B. (1990) Low molecular weight albumins from sunflower seed: identification of a methionine-rich albumin, *Phytochem.* **29**, 2805-2810.

Kortt, A.A., Caldwell, J.B., Lilley, G.G., and Higgins, T.J.V. (1991) Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (*Helianthus annuus* L.), *European Journal of Biochemistry* **195**, 329-334.

Laccy, D.J., Wellner, N., Beadoin, F., Napier, J.A., Shewry, P.R. (1998) Secondary structure of oleosins in oil bodies isolated from seeds of safflower (*Carthamus tinctorius* L.) and sunflower (*Helianthus annuus* L.), *Biochem.J.* **1**, 334 (Pt 2), 469-477.

Schwenke, K.D., Pahtz, W., Linov, K., Raab, B., and Schultz, M. (1979) On seed proteins. Purification, chemical composition and some physico-chemical properties of the 11S globulin (helianthinin) in sunflower seed, *Nahrung* **23**, 241-254.

Vonder Haar, R.A., Allen, R.D., Cohen, E.A., Nessler, C.L. and Thomas, T.L. (1988) Organisation of the sunflower 11S storage protein gene family, *Gene* **74**, 433-443.