

Association mapping of drought tolerance in sunflower: phenotypic and genetic characterization of INTA inbred lines

Maria Valeria Moreno¹, Corina M. Fusari², Verónica Nishinakamasu², Daniel Alvarez³, Julio Di Rienzo⁴, Esteban Hopp², Ruth Heinz², Norma Paniego², Verónica V. Lia²

¹Laboratorio de Biotecnología, INTA-EEA Manfredi. Ruta Nac. N° 9. Km 636. (5988) Manfredi, Córdoba, Argentina. E-mail: mvmoreno@manfredi.inta.gov.ar. ²Instituto de Biotecnología, CNIA-INTA Castelar. Los Reseros y Las Cabañas S/N. (1686) Hurlingham, Buenos Aires, Argentina. ³Grupo Mejoramiento de Girasol, INTA-EEA Manfredi. Ruta Nac. N° 9. Km 636. (5988) Manfredi, Córdoba, Argentina. ⁴Cátedra de Estadística y Biometría, FCA-UNC. Ciudad Universitaria, Córdoba, Argentina.

ABSTRACT

- Association mapping is a powerful tool to identify genes or alleles that contribute to variation in complex traits using ensembles of unrelated individuals such as germplasm collections or natural populations. Genome-wide and candidate gene approaches have been applied in plants but, regardless of the genotyping strategy, one of the key aspects of the method is obtaining reliable phenotypic measurements. In candidate gene approaches, both phenotypic evaluation and the selection of regions for genetic analysis are major determinants of the power to identify QTL. Tolerance to water deficit is a highly complex trait for which phenotypic assessment can be difficult. Different quantitative variables may serve to describe a plant response to drought and these may vary along different developmental stages. Here we present a preliminary study aimed to develop an association mapping platform for drought tolerance in sunflower. The goals of this work were: a) to assess the suitability of mannitol assays for the evaluation of drought tolerance during seed germination, and b) to study nucleotide diversity, linkage disequilibrium and evolutionary patterns of seven candidate genes previously reported to be involved in the response to water deficit in sunflower and other species.
- Thirty four sunflower inbred lines were evaluated under two water deficit regimes (200 mM and 400 mM mannitol) during 11 days. The percentage of germinating seeds was scored at the end of the experiment and results were analyzed under a logistic regression model. A set of 7 SSR loci was assayed to infer genetic relationships among the 34 inbred lines studied and seven candidate regions (Hahb4-p, Hahb4, Suntip, FB, HaL1L, HaDhn1 and CK) were screened for variation to identify SNP and InDels.
- Susceptible and resistant genotypes were clearly identified in 200 mM mannitol assays, whereas no such distinction could be achieved under the more stringent water deficit regime. Analysis of nucleotide diversity revealed an average SNP frequency of 1/48.3 bp and an average nucleotide diversity (θ_w) of 0.00501. In agreement with previous reports, a high level of LD was observed for all the regions ($r^2=0.88$ at 900 bp) except for Suntip, for which LD decays to $r^2=0.325$ at 207 bp and HaDhn1, for which LD decays to $r^2=0.249$ at 78 bp. No deviations from neutrality were observed for Suntip and HaDhn1, whereas evidences of purifying selection were found for Hahb4-p, Hahb4, FB and HaL1L.
- The mannitol test proved to be a reliable and potentially useful method for the screening of drought tolerance in sunflower. The assignment of inbred lines to the susceptible and resistant categories was highly concordant with previous field trials. All regions evaluated are interesting candidate genes to be used in association mapping for this trait.
- The results presented here constitute the first steps towards identifying drought tolerance QTL in sunflower via the association mapping approach. Further efforts are currently being made to expand the association mapping population, the number of candidate genes and the phenological stages of the phenotypic evaluation.

Keywords: candidate genes, nucleotide diversity, SNP, InDel, water deficit

INTRODUCTION

Germplasm banks play a fundamental role in crop breeding programs and their importance as sources of genetic variation has gain renewed attention with the advent of association mapping strategies. Association mapping (AM) is a powerful tool to identify genes or alleles that contribute to variation in complex traits using ensembles of unrelated individuals such as germplasm collections or natural populations. The main goal of AM studies is to detect correlations between genotypes and phenotypes on the basis of linkage disequilibrium (LD) (Breseghello and Sorrells, 2006) and recombination events (Oraguzie et al., 2007). AM can be more efficient in the use of genetic resources than classical biparental mapping since several traits can be studied in a single population using the same genotypic data. Genome-wide and candidate gene approaches for AM have been applied in plants but, regardless of the genotyping strategy, one of the key aspects of the method is obtaining reliable phenotypic measurements. In candidate gene approaches, both phenotypic evaluation and the selection of regions for genetic analysis are major determinants of the power to identify QTL. False positives can arise from unaccounted subdivisions in the sample, like population structure, thus resulting in a systematic bias (Buckler and Thornsberry, 2002; Hirschhorn and Daly, 2005). Association mapping populations (AMP) can include germplasm bank core collections, synthetic outcrossing population or inbred lines, the latter are more stable and adapted to normal growth conditions but they tend to show a more pronounced population structure.

Because of their abundance and wide distribution in the genome, single nucleotide polymorphisms (SNP) are the molecular markers of choice for AM and a large number of SNP discovery projects have been carried out in cultivated species during the last years (Feltus et al., 2006; Fusari, 2010; Giancola et al., 2006; Monna et al., 2006; Rostoks et al., 2005; Schmid et al., 2003).

Tolerance to water deficit is a highly complex trait for which phenotypic assessment can be difficult. Water deficit can produce yield reduction and variations in seed biochemistry composition like oleic/linoleic or protein/oil ratio (Roche et al., 2004). Different quantitative variables may serve to describe a plant response to drought and these may vary along different developmental stages. Reduction of germinative power in most crops is produced by water deficit, therefore the implantation stage is useful for the study of drought effect on seed germination.

The sunflower Active Germplasm Bank of INTA Manfredi (AGB-IM) conducts breeding for diverse agronomically important traits. The purpose of the AGB-IM drought breeding program is to develop inbred lines that are tolerant to water deficit and capable of adapting to a variety of environments. In this context, the goals of the present work were: a) to assess the suitability of mannitol assays for the evaluation of drought tolerance during seed germination, and b) to study nucleotide diversity, linkage disequilibrium and evolutionary patterns of seven candidate genes previously reported to be involved in the response to water deficit in sunflower and other species.

MATERIALS AND METHODS

Thirty four sunflower AGB-IM inbred lines were included in the present study.

Phenotypic characterization: Inbred lines were evaluated under two water deficit regimes (200 mM and 400 mM mannitol, moderate and severe stress, respectively) during 11 days (Bigo, 2005). The experimental design consisted of a complete randomized block design with three blocks, each corresponding to a different date. Sterilized sand was hydrated until 60% yield capacity with mannitol solutions or bidistilled water depending on the treatment. Twenty five disinfected seeds were used for each treatment and inbred line. Capillary irrigation was done at 4, 7 and 10 days after seeding with mannitol solutions or bidistilled water. Plants were grown in a chamber under controlled conditions (25°C and 16 hours of light). Germinative energy (GE) was measured during irrigation on the fourth day, while fresh and dry weight (FW, DW) and the percentage of germinating seeds (PGS) were scored at the end of the experiment. Results were analyzed under a logistic regression model based on predicted values using InfoStat/Profesional v.2008 (Di Rienzo et al., 2008). The dependent variable was PGS, while classifier variables were genotype, treatment and repetition.

Genetic characterization: A set of 7 SSR loci (HA77, HA293, HA928, HA2063, HA2920, HA3239 and HA4103) was assayed to infer genetic relationships among the 34 inbred lines studied. Population structure was analyzed using the Bayesian approach implemented in Structure 2.2 (Pritchard et al., 2000). The parameters used were 1-10 clusters (K), burn-in of 100.000, runs of 500.000 and correlated frequency model (Falush et al., 2003).

Seven candidate regions were screened for variation to identify SNP and InDels and estimate linkage disequilibrium in these regions. The analyzed regions were: Hahb4-p (AF339749.1; Dezar et al., 2005); Hahb4 (AF339748.1, Gago et al., 2002); Suintip (X95950, Poormahammad Kiani et al. 2007); FB (CD850867, Poormahammad Kiani et al. 2007); HaL1L (TC40698, Poormahammad Kiani et al. 2007); HaDhn1 (X92647, Cellier et al., 1998, 2000) and CK (TC54007, Roche et al., 2007). Forward and reverse sequence editing and contig assembly were done with BioEdit Sequence Alignment Editor 1999-2005 (Hall, 1999). Genetic variability in candidate genes was evaluated through nucleotide polymorphism (θ_w ; Watterson, 1975), nucleotide diversity (π ; Tajima, 1983) and haplotypic diversity (HD; Nei, 1987) with DNAsp v.5.10.01 software (Librado and Rozas, 2009). Population bottlenecks were tested using SSR data under the infinitive allele model (IAM) with Bottleneck software (Cornuet and Luikart, 1996).

RESULTS AND DISCUSSION

Phenotypic characterization

Using PGS as dependent variable, R426-1, R435, R429, B71, B99, RHA274, R431, R421, HA89, R432 and B10 inbred lines were classified as tolerant under moderate water deficit with predicted values ranging from 0.09 to 0.18 (Fig. 1-A). In contrast, R073, R043, R422, R420, B67, 964-C, 1063-C, R307, R423, B59 and A59 were classified as susceptible.

R426-1, R420, R431, R437, R073, R433 and R435 inbred lines were classified as tolerant under severe water deficit with 0.07 and 0.25, being the minimum and maximum predicted values respectively (Fig. 1-B). In this experiment, inbred lines 1063-C, HAR4, 430-C, R417, R422, R049, R043, R421, R413, R307, 962-C, RHA274, HA89, B67, 964-C, B71, R415 and B99 were classified as susceptible. These results were partially consistent with Alemano et al. (2005), probably due to the differences in growth conditions during the assay. The high number of plants with defects under severe water deficit was concordant with results by Hernandez (2005). In agreement with Alemano et al. (2005), no clear category delimitation was possible with GE, FW and DW variables (data not shown), susceptible and tolerant genotypes were clearly identified in 200 mM mannitol assays, whereas no such distinction could be achieved under the more stringent water deficit regime.

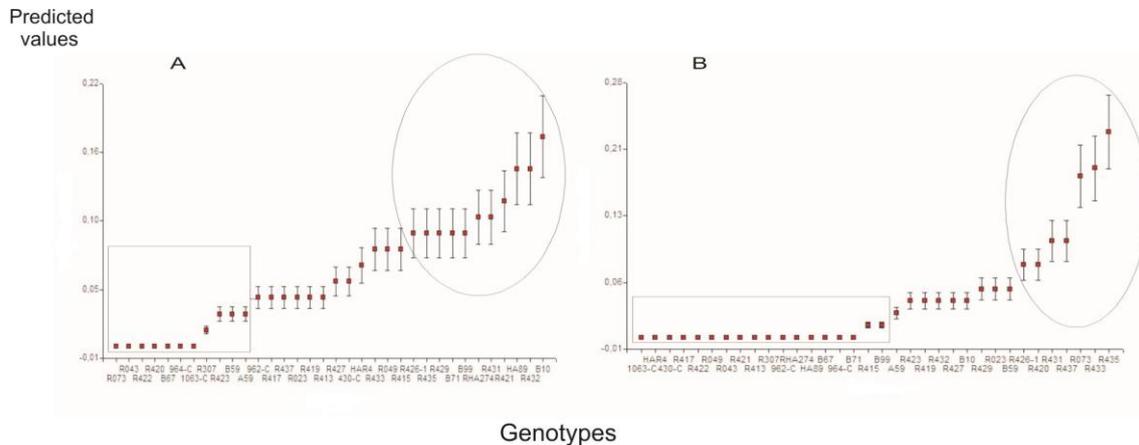


Fig. 1. Behavior of inbred lines under moderate (A) and severe (B) water deficit. Inbred lines within ovals were considered tolerant. Inbred lines within rectangles were considered susceptible.

The assignment of inbred lines to the susceptible and tolerant categories was highly concordant with previous field trials (Alemano et al., 2005). The mannitol test under moderate water deficit proved to be a reliable and potentially useful method for the screening of drought tolerance in sunflower.

Population structure based on SSR data

Bayesian analysis of population structure shows $K=9$ as the more likely number of genetic clusters (Fig. 2), with a slope reduction of $\ln P(D)$ vs. K . Consistent with a nested pattern of population substructuring, three local maximum of ΔK , i.e. $K=2$, $K=6$ and $K=9$ were detected in the Evanno analysis (Evanno et al., 2005) (data not shown).

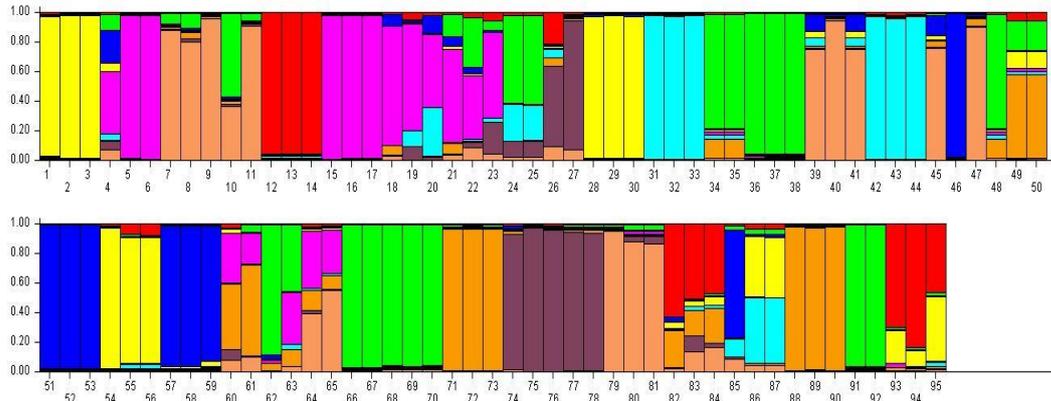


Fig. 2. Bayesian analysis of population structure in inbred lines analyzed.

Genetic variability in candidate genes

Analysis of nucleotide diversity revealed an average SNP frequency of 1/48.3 bp and a mean InDel frequency of 1/332.12 bp. Similar estimates were reported by Kolkman et al. (2007), whereas lower and higher values were obtained by Fusari et al. (2008) and Giordani et al. (2010), respectively. The number of SNP varied between 6 (CK) and 39 (Hahb4-p). The number of InDels ranged from 0 (Hahb4) to 6 (Hahb4-p). Thirty three SNP were detected in coding regions (2976 bp, 1/90.18 bp), compared to 82 SNP found in non-coding regions (2670 bp, 1/32.56 bp). Average nucleotide polymorphism (θ_w) was 0.00501 and nucleotide diversity (π_T) was 0.00294 (Table 1). Nucleotide polymorphism is comparable to the results from Fusari et al. (2008), however nucleotide diversity was markedly lower than the values informed by these authors ($\theta_w=0.0056$, $\pi_T=0.0061$). Higher estimates were found by Kolkman et al. (2007) for both indices ($\theta_w=0.0094$, $\pi_T=0.0107$). Differences between the genetic variability levels detected here and those from previous studies may arise from the nature of the germplasm being considered and/or from the number and identity of the genes selected for analysis.

Table 1. Measures of nucleotide variability in 7 candidates regions for water deficit tolerance.

Regions	SI	θ_w	π	HN	HD	Tajima's D
Hahb4-p	39	0.01018	0.00276	2	0.069	-2.697****
Hahb4	16	0.00664	0.00180	2	0.069	-2.493****
Suntip	23	0.00734	0.00827	6	0.470	0.126
FB	7	0.00164	0.00040	2	0.061	-2.178***
HaL1L	7	0.00265	0.00062	2	0.057	-2.180***
HaDhn1	13	0.00417	0.00562	5	0.692	1.107
CK	6	0.00249	0.00114	2	0.111	-1.494
Average	15.8	0.00501	0.00294	3	0.218	

SI: Number of segregating sites. HN: Number of Haplotypes. HD: Haplotypic diversity.*** p<0.01; **** p<0.001

In agreement with previous results from cultivated sunflower (Fusari et al., 2008; Kolkman et al., 2007; Liu and Burke, 2006), strong LD ($r^2=0.88$ at 900 bp) was observed for Hahb4-p, Hahb4, FB, HaL1L, and CK. In contrast, Suntip LD decays to $r^2=0.325$ at 207 bp and HaDhn1 LD decays to $r^2=0.249$ at 78 bp, as described in the analysis of sequence variability in eight sunflower genes putatively involved in drought response (Giordani et al., 2010). In the context of AM, high levels of within gene LD may allow reducing genotyping costs by using Tag SNP (Zhang et al., 2004).

No deviations from neutrality were observed for Suntip and HaDhn1, whereas evidences of purifying selection were found for Hahb4-p, Hahb4, FB and HaL1L (Tajima's D, Table 1). Negative and statistically significant Tajima's D values were also found by Giordani et al. (2010) for genes related with drought tolerance in sunflower. Although both demographic processes and selective forces may explain the observed polymorphism profiles (González-Martínez et al., 2006), no significant differences were detected when comparing the observed expected heterozygosity derived from the SSR data to the one expected for a population under drift-mutation equilibrium, thus excluding severe bottlenecks as responsible for the excess of low frequency variants in candidate genes.

In sum, the results presented here constitute the first steps towards identifying drought tolerance QTL in sunflower via the association mapping approach. Further efforts are currently being made to expand the association mapping population, the number of candidate genes and the phenological stages of the phenotypic evaluation.

ACKNOWLEDGMENTS

Several grants from CONICET and INTA are gratefully acknowledged.

REFERENCES

- Aleman, S.G.; Álvarez, D.; Bigo, V.; Ferragutti, S.; Vigliocco, A.E.; Abdala, G. 2005. Caracterización de Germoplasma de girasol en condiciones de estrés hídrico. Buenos Aires (AR): ASAGIR.
- Bigo, V. 2005. Caracterización de germoplasma de girasol (*Helianthus annuus*), en condiciones de estrés hídrico. Tesis de grado. Río Cuarto, Córdoba (AR). Universidad Nacional de Río Cuarto. Facultad de Ciencias Exactas, Físico-Químicas y Naturales. Laboratorio de Fisiología Vegetal. 19 p.
- Braverman, J.M.; Hudson, R.R.; Kaplan, N.L.; Langley, C.H.; Stephan, W. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* 140: 783-796.
- Breseghele, F.; Sorrells, M.E. 2006. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172: 1165-1177.
- Buckler, E.S.; Thornsberry, J.M. 2002. Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology* 5: 107-111.
- Cellier, F.; Conejero, G.; Breitler, J.C.; Casse, F. 1998. Molecular and physiological responses to water deficit in drought-tolerant and drought-sensitive lines of sunflower. Accumulation of dehydrin transcripts correlates with tolerance. *Plant Physiology* 116: 319-328.
- Cellier, F.; Conejero, G.; Casse, F. 2000. Dehydrin transcript fluctuations during a day/night cycle in drought stressed sunflower. *Journal of Experimental Botany* 51: 299-304.
- Cornuet, J.M.; Luikart, G. 1996. Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144: 2001-2014.
- Dezar, C.A.; Gago, G.M.; González, D.H.; Chan, R.L. 2005. *Hahb-4*, a sunflower homeobox-leucine zipper gene, is a developmental regulator and confers drought tolerance to *Arabidopsis thaliana* plants. *Transgenic Research* 14: 429-440.
- Di Rienzo, J.A.; Casanoves, F.; Balzarini, M.G.; González, L.; Tablada, M.; Robledo, C.W. 2008. InfoStat, Versión 2008. Córdoba (AR). Universidad Nacional de Córdoba. Facultad de Ciencias Agropecuarias. Grupo InfoStat.
- Evanno, G.; Regnaut, S.; Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* 14: 2611-2620.
- Falush, D.; Stephens, M.; Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.
- Feltus, F.A.; Singh, H.P.; Lohithaswa, H.C.; Schulze, S.R.; Silva, T.D.; Paterson, A.H. 2006. A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved non-coding sequences in orphan crops. *Plant Physiology* 140: 1183-1191.
- Fusari, C.M.; Lia, V.V.; Hopp, H.E.; Heinz, R.A.; Paniago, N.B. 2008. Identification of single nucleotide polymorphisms and analysis of linkage disequilibrium in sunflower elite inbred lines using the candidate gene approach. *BMC Plant Biology* 8:7.
- Fusari, C.M. 2010. Mapeo por asociación en girasol: diversidad nucleotídica, desequilibrio de ligamiento e identificación de genes involucrados en la resistencia a la podredumbre húmeda del capítulo. Ph.D. diss. (AR). Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. 251 p.
- Gago, G.M.; Almoguera, C.; Jordano, J.; González, D.H.; Chan, R.L. 2002. *Hahb4*, a homeobox-leucine zipper gene potentially involved in abscisic acid-dependent responses to water stress in sunflower. *Plant Cell and Environment* 25: 633-640.
- Giancola, S.; McKhann, H.I.; Berard, A.; Camilleri, C.; Durand, S.; Libeau, P.; Roux, F.; Reboud, X.; Gut, I.G.; Brunel, D. 2006. Utilization of the three high-throughput SNP genotyping methods, the GOOD assay, Amplifluor and TaqMan, in diploid and polyploid plants. *Theoretical and Applied Genetics* 112: 1115-1124.
- Giordani, T.; Buti, M.; Natali, L.; Pugliesi, C.; Cattonaro, F.; Morgante, M.; Cavallini, A. 2010. An analysis of sequence variability in eight genes putatively involved in drought response in sunflower (*Helianthus annuus* L). *Theoretical and Applied Genetics*. DOI 10.1007/s00122-010-1509-0.

- González-Martínez, S.C.; Ersoz, E.; Brown, G.R.; Wheeler, N.C.; Neale, D.B. 2006. DNA Sequence variation and selection of Tag Single-Nucleotide polymorphisms at candidate genes for drought stress response in *Pinus taeda* L. *Genetics* 172: 1915-1926.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.
- Hernández, C.R. 2005. Acondicionamiento osmótico de simientes de girasol (*Helianthus annuus* L.) para el avance de la germinación en siembras precoces para zonas áridas. Ph.D. diss. Madrid (ES). Universidad Politécnica de Madrid. 210 p.
- Hirschhorn, J.N.; Daly, M.J. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews: Genetics* 6: 95-108.
- Kolkman, J.M.; Berry, S.T.; Leon, A.; Slabaugh, M.B.; Tang, S.; Gao, W.; Shintani, D.K.; Burke, J.M.; Knapp, S.J. 2007. Single nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics* 177 (1): 457-468.
- Librado, P.; Rozas, J. 2009. DNASP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.
- Liu, A.; Burke, J.M. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* 173: 321-330.
- Monna, L.; Ohta, R.; Masuda, H.; Koike, A.; Minobe, Y. 2006. Genome-wide searching of single-nucleotide polymorphisms among eight distantly and closely related rice cultivars (*Oryza sativa* L.) and a wild accession (*Oryza rufipogon* Griff.). *DNA Research* 13: 43-51.
- Nei, M. 1987. *Molecular evolutionary genetics*. New York (USA). Columbia University Press.
- Oraguzie, N.C.; Rikkerink, E.H.A.; Gardiner, S.E.; De Silva, H.N. 2007. *Association mapping in plants*. Oraguzie, N.C.; Rikkerink, E.H.A.; Gardiner, S.E.; De Silva, H.N. New York (USA). Springer Science+Business Media.
- Poormohammad Kiani, S.; Grieu, P Maury, P.; Hewezi, T.; Gentzbittel, L.; Sarrafi, A. 2007. Genetic variability for physiological traits under drought conditions and differential expression of water stress-associated genes in sunflower (*Helianthus annuus* L.). *Theoretical and Applied Genetics* 114: 193-207.
- Pritchard, J.K.; Stephens, M.; Donnelly, P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Roche, J.; Essahat, A.; Bouniols, A.; El Asri, M.; Mouloungui, Z.; Mondières, M.; Alghoum, M. 2004. Diversified composition of sunflower (*Helianthus annuus* L.) seeds within cultural practices and genotypes (hybrids and populations). *Helia* 27:73-98.
- Roche, J.; Hewezi, T.; Bouniols, A.; Gentzbittel, L. 2007. Transcriptional profiles of primary metabolism and signal transduction-related genes in response to water stress in field-grown sunflower genotypes using a thematic cDNA microarray. *Planta* 226: 601-617.
- Rostoks, N.; Mudie, S.; Cardle, L.; Russell, J.; Ramsay, L.; Booth, A.; Svensson, J.T.; Wanamaker, S.I.; Walia, H.; Rodriguez, E.M.; Hedley, P.E.; Liu, H.; Morris, J.; Close, T.J.; Marshall, D.F.; Waugh, R. 2005. Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Molecular and General Genetics* 274: 515-527.
- Schmid, K.J.; Sorensen, T.R.; Stracke, R.; Torjek, O.; Altmann, T.; Mitchell-Olds, T.; Weisshaar, B. 2003. Large-scale identification and analysis of genome-wide single nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. *Genome Research* 13(6A): 1250-1257.
- Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105(2): 437-460.
- Watterson, G.A. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7: 256-276.
- Zhang, K.; Qin, Z.S.; Liu, J.S.; Chen, T.; Waterman, M.S.; Sun, F. 2004. Haplotype block partitioning and Tag SNP selection using genotype data and their applications to association studies. *Genome Research* 14: 908-916.