*HELIA*, **29**, Nr. 45, p.p. 11-18, (2006) UDC 633.854.78:631.523

# IDENTIFICATION OF WILD SPECIES OF SUNFLOWER BY A SPECIFIC PLASTID DNA SEQUENCE

Vischi, M.*, Arzenton, F., De Paoli, E., Paselli, S., Tomat, E., Olivieri, A.M.

*Dip. Scienze Agrarie ed Ambientali, Università di Udine,*
*Via delle Scienze, 208, I 33100 Udine, Italy*

*Received: October 15, 2006*
*Accepted: December 05, 2006*

SUMMARY

Four sunflower species, *Helianthus annuus*, *H. argophyllus*, *H. debilis* and *H. tuberosus*, were characterized at the molecular level using the plastid trnH-psbA intergenic spacer. The trnH-psbA sequence was selected with the aim of developing a "DNA barcode" system (Kress *et al.*, 2005) as a tool for species and specimen identification. The plastid region was PCR amplified with specific primers and sequenced with an ABI Prism 3730 Automated DNA sequencer. Intraspecific and interspecific sequence variation was evaluated to assess the resolution of the technique. Sequencing of both forward and reverse strands allowed for a high base calling accuracy and overcame the problem of polymerase slippage within microsatellite regions. After sequence editing, a very low (or absent) intraspecific variability was detected, whereas interspecific variability due to SNPs, indels and SSR length was sufficient for an unambiguous identification of each species.

**Key words: barcoding, sunflower, wild species, taxonomy, introgression**

## INTRODUCTION

Methods for identifying species by using short orthologous DNA sequences, known as "DNA barcodes", have been proposed and initiated to facilitate biodiversity studies. A consortium for the Barcoding of Life (http://www.barcodinglife.org) has been set up to build up a digital library of sequences linked to vouchered specimens. The cytochrome *c* oxidase 1 sequence has been found to be widely applicable in animal barcoding but it is characterized by relatively low rates of sequence divergence in land plants (Cho *et al.*, 2004; ref). Only recently, the internal transcribed spacer (ITS) region of the nuclear ribosomal cistron (18S-5.8S-26S) and the *trnH-psbA* plastid intergenic spacer have been proposed by Kress *et al.* (2005) as appro-

* Corresponding author: Phone +39 0432 558609, Fax. +39 0432 558603 ,
e-mail: massimo.vischi@uniud.it

priate for barcoding of flowering plants since they a) have significant species-level genetic variability and divergence, b) are appropriately short so as to facilitate DNA extraction and amplification, and c) are flanked by conserved sites for developing universal primers. The plastid sequence is uniparentally inherited, non-recombining and structurally stable and in initial studies it has proven to be easier to analyze and overcome some technical difficulties in the use of nuclear ITS sequence due to the presence of divergent copies often within single individuals (Muir and Schlotterer, 1999).

The application of barcoding to flowering plants is at the beginning and at the moment there are no such studies in sunflower. The genus *Helianthus* comprises about 50 species and many examples of interspecific hybrids have been reported. To explore the usefulness of the plastid sequence as a barcode, we analyzed some individuals of the cross compatible species *H. annuus*, *H. debilis*, *H. argophyllus* and *H. tuberosus*.

## MATERIALS AND METHODS

### Plant material, DNA extraction and PCR amplification

Seedlings of *H. annuus*, *H. argophyllus*, *H. debilis* and *H. tuberosus* were cultivated at the experimental farm of University of Udine. Young leaves from several plants of each species were collected for total genomic DNA extraction. The collected leaf tissues were immediately ground in liquid nitrogen. Individual DNA extractions were carried out using the CTAB method (Doyle and Doyle, 1987).

Plastid sequences were amplified using the primers reported by Kress *et al.* (2005). The polymerase chain reaction (PCR) amplifications were performed in 25 $\mu$l total reaction volume containing GeneAmp® 1× PCR Buffer (Applies Biosystem), 2.5 mM $MgCl_2$, 0.3 $\mu$M of each primer and 1.5 U of AmpliTaq Gold™. The PCR cycling conditions were: initial denaturation at 95°C for 3 min followed by 38 cycles of denaturation at 94°C for 30 s, primer annealing at 58°C for 30 s and 72°C for 30 s. A final extension for 7-10 min at 72°C was included to minimize the number of partial strands.

### Sequence analysis

PCR products were purified by the MultiScreen-$PCR_{96}$ Kit (Millipore) according to the manufacturer's instructions. The purified PCR product was used as template in 10 $\mu$l cycle sequencing reaction (for and rev) using the ABIPRISM– BigDye™ Terminator Cycle Sequencing Ready Reaction Kit v.2.0 (Applied Biosystem). The cycle sequencing product was purified by ethanol and sequenced by the Applied Biosystem 3730 DNA sequencer.

Chromatograms were transferred to a Unix workstation, base called with Phred (version 3.01) assembled with Phrap (version 3.01), and the results viewed with

Consed (version 9.0), (Ewin *et al.*, 1998a, 1998b). Sequences were further manually edited by GENEDOC (Nicholas and Nicholas, 1997). Full-length sequences were aligned with CLUSTAL W (Thompson *et al.*, 1994) and BLAST (Altschul *et al.*, 1997) search was carried out to find similarities in GenBank.

## RESULTS AND DUSCUSSION

After PCR amplification and sequencing of the *trnH-psbA* sequence in different DNA samples, we selected a total of 48 sequences, namely 15 sequences of *H. annuus*, 20 of *H. debilis*, 9 of *H. argophyllus* and 4 of *H. tuberosus*. Sequence length ranged between 480 and 500 bp.

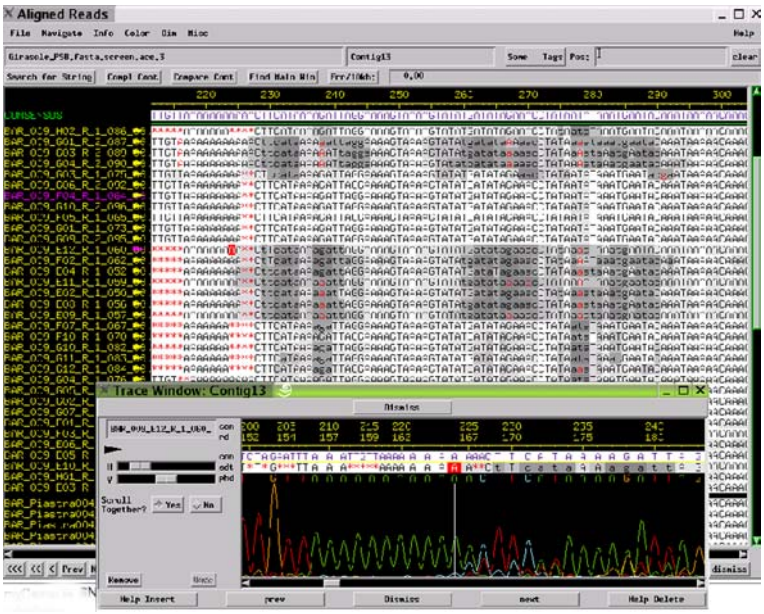


*Figure 1: Electropherogram of trnH-psbA sequence. Note the multiple peaks downstream of the homopolymer (polyA) region.*

The observation of electropherograms emphasized the high quality of forward sequences up to the microsatellite (poly A) region approximately at half length of the sequence. After that point the sequence quality dropped down regarding the presence of double peaks (Figure1). This problem was due to the extent of slippage of the enzyme used in amplification and sequencing as the enzyme keeps falling off and rejoining in random places in the homopolymer region. Therefore, to obtain a complete sequence of high quality, it was necessary to consider the reverse sequence too, since in this case the quality of sequence was complementary. In this way it was possible to obtain complete sequences of high quality for each species before proceeding to the analysis and alignment of sequences.

The first step was to evaluate the intraspecific variation of sequences and if there was overlapping with the interspecific divergences. No sequence variation was detected within each of the species *H. annuus, H. debilis* and *H. tuberosus*. On the other hand, within the species *H. argophyllus* we found two groups of sequences with slight differences, namely two SNPs (T>A) and one indel in positions 47, 99 and 146, respectively (Table 1).

Table 1: Intraspecific variation of *trnH-psbA* sequence in *Helianthus argophyllus*. The differences between the two families of sequences are highlighted in yellow/green. Note two SNPs in positions 47 and 99 and one indel in position 146.

```
HELIANTHUS_ARGOPHYLLUS_2     TCTACTATTATCTAGTATTACTATATTTTTCCATTAACATAAAAAAGCAT
HELIANTHUS_ARGOPHYLLUS_78    TCTACTATTATCTAGTATTACTATATTTTTCCATTAACATAAAAAAACAT
                             ********************************************** ***
                             1                                           47

HELIANTHUS_ARGOPHYLLUS_2     ATTCTTTCTCATTTTTATTGAATAAATAAAAGTAATAAATAAGCAAAATT
HELIANTHUS_ARGOPHYLLUS_78    ATTCTTTCTCATTTTTATTGAATAAATAAAAGTAATAAATAAGCAAAAAT
                             ********************************************** *
                             51                                          99

HELIANTHUS_ARGOPHYLLUS_2     TTCATTTTTATCTAGAATTTAAATT------------------GTTAAAA
HELIANTHUS_ARGOPHYLLUS_78    TTCATTTTTATCTAGAATTTAAATT------------------GT-AAAA
                             *************  * *** * **                  ** ****
                             101                                        146
```

The next step was the multiple alignment of sequences of the four species considered in this study. As the first observation, the sequences were highly conserved and the variations were found almost exclusively in the region of approximately 60 bp between the base 100 and 160 (Table 2). In this region three sources of interspecific variation were found: SNPs, indels and the number of repeats of the poly A motif. The divergences in that region were sufficient for an unambiguous identification of each species.

Table 2: Interspecific variation in *H.annuus, H. argophyllus, H. debilis and H. tuberosus*. The differences are highlighted in yellow, green and red (see text for details)

```
HELIANTHUS ANNUUS            TTCATTTTTATCTAGAATTTAAATT------------------GTTAAAA
HELIANTHUS_ARGOPHYLLUS       TTCATTTTTATCTAGAATTTAAATT------------------GTTAAAA
HELIANTHUS_DEBILIS           TTCATTTTTATCTAGAATTTCATTTTTATCTAGAATTTAAATTGTTAAAA
HELIANTHUS_TUBEROSUS         TTCATTTTTATCT--AATTTCAATT-------GAATTTAAATTGT-AAAA
                             *************  * *** * **                  ** ****
                             101

HELIANTHUS ANNUUS            AAAAAAA-CTTCATAAAAGATTAGGAAAAGTAAAAGTATATGATATAGAA
HELIANTHUS_ARGOPHYLLUS       AAAAA---CTTCATAAAAGATTAGGAAAAGTAAAAGTATATGATATAGAA
HELIANTHUS_DEBILIS           AAAAA---CTTCATAAAAGATTAGGAAAAGTAAAAGTATATGATATAGAA
HELIANTHUS_TUBEROSUS         AAAAAAA-CTTCATAAAAGATTAGGAAAAGTAAAAGTATATGATATAGAA
                             *****   ****************************************
                             151
```

*H. tuberosus* was easily recognizable for the species-specific presence of two indels and one SNP in position 114-115, 126-132 and 117, respectively (highlighted in red in Table 2). *H. debilis* was distinguished from *H. annuus* and *H. argophyllus* for one SNP in position 121 and one indel in position 126-143 (highlighted in yellow in Table 2). The sequences of *H. annuus* and *H. argophyllus* exhibited almost complete similarity with the remarkable exception of the number of repeats of the

microsatellite motif. The number of repeats was 11 and 9 for *H. annuus* and *H. argophyllus* respectively (highlighted in green in Table 2), making the two species distinguishable.

GenBank BLAST searches with data returned correct matches only for the sequences of *H. annuus* with more than 95% of homology. That means that the *trnH-psbA* intergenic spacer was not yet sequenced in the other species.

## CONCLUSIONS

The variability we found in the plastid *trnH-psbA* intergenic spacer was sufficient for a correct identification of each of the four sunflower species analyzed in this study. The technique was fast, although some manual editing was necessary before multiple alignments of complete sequences.

We expect that the process could even be faster if specific primers were designed for amplification, sequencing and analysis of shorter sequences, since we found enough variability for species identification in a limited region of about 60 bp, flanked by very conserved sequences. In this way DNA barcoding can be a simple, fast and cost effective molecular tool very useful at the taxonomic level in the genus *Helianthus*.

Wild species of *Helianthus* represent interesting sources of many useful characters (drought and salinity resistance, source of *cms*, disease resistance). The occurrence of interspecific hybrids was reported in several studies (Heiser, 1976, 1978; Reisenberg *et al.,* 1995; Olivieri *et al*., 1999). The morphological identification of hybrids is often difficult due to the slight differences with usually one of the parental species or changes during the developmental stages. For instance, the seedlings of the interspecific cross *H. debilis* × *H. argophyllus* look like the pollinator species (*i.e., H. argophyllus*) at early stage of development regarding leaf shape, color and hairiness look. These traits tend to disappear in next leaves and they are consequently very similar to the female species (*i.e., H. debilis*) although sometime slightly larger in size (Olivieri, A., personal communication). In this context, the plastid, uniparentally inherited, could allow an early identification of the maternal species in the interspecific hybrids.

The internal transcribed spacer (ITS) region of the nuclear ribosomal cistron (18S-5.8S-26S) is the other barcoding sequence proposed by Kress *et al.* (2005). This sequence, bi-parentally inherited, could be useful in hybridization and introgression studies also allowing correct identification of pollen donor species.

The use of the nuclear ITS sequence together with the plastid *trnH-psbA* sequence can increase the power of the resolution of DNA barcoding.

Experiments to test these sequences for these applications are under way.

## REFERENCES

Altschul, S.F., Madden, T.,L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17): 3389-3402.

Doyle, J.J. and Doyle, J.L., 1997. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19: 11-15.

Ewing, B, Green, P., 1998: Basecalling of automated sequencer traces using Phred. II. Error probabilities. Genome Research 8: 186-194.

Ewing, B., Hillier, L., Wendl, M. and Green, P., 1998. Basecalling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Research 8: 175-185.

Heiser, C.B., 1976. Sunflowers. Simmonds NW (ed) Evolution of crop plants. Longman, London, pp. 36-38.

Heiser, C.B., 1978. Taxonomy of *Helianthus* and the origin of domesticated sunflower. Sunflower Science and Technology. Ed. J.F. Carter, pp. 31-54. Madison, WI: Am. Soc. Agron., Crop Sci. Soc. and Soil Sci. Soc. Am.

Kress, J.W., Wurdack, K.J., Zimmer E.A., Weigt, L.A. and Janzen, D.H., 2005. Use of DNA Barcodes to identify flowering plants. Proc. Natl. Acad. Sci. USA 102: 8369-8374.

Muir, G., Schlotterer, C.,1999. Limitations to the phylogenetic use of ITS sequences in closely related species and populations-a case study in Quercus petraea (Matt.) Liebl. Chapter 11. *In*: Which DNA Marker for Which Purpose? [M].Final Compendium of the Research Project: Development, optimization and validation of molecular tools for assessment of biodiversity in forest trees in the European Union DGXII Biotechnology FW IV Research Program Molecular Tools for Biodiversity. Gillet, EM (ed.).

Nicholas, K.B., Nicholas, H.B.Jr., and Deerfield, D.W.II., 1997. GeneDoc: Analysis and Visualization of Genetic Variation, EMBNEW.NEWS 4: 14.

Olivieri, A.M., Magaia, H.E. and Cagiotti, M.E., 1999. *Helianthus argophyllus* and *H. debilis*: two wild Texas sunflower species present in Mozambique. *In*: Proc. Symp. on Sunflower and Other Oilseed Crops in Developing Countries. Maputo, 9-12 February 1999, pp. 232-238.

Rieseberg, L.H., Randal Linder, C. and Seiler G.J., 1995. Chromosomal and Genic Barriers to Introgression in *Helianthus*. Genetics 141: 1163-1171.

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.

Cho, Y., Mower, J.P., Qiu, Y.-L. and Palmer, J.D. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants Proc. Natl. Acad. Sci. USA 101, 17741-17746.

## IDENTIFICACIÓN DE LAS ESPECIES SILVESTRES DE GIRASOL POR MEDIO DE LA SECUENCIA ESPECÍFICA DE PLÁSTIDO DE DNA

RESUMEN

Fue hecha la caracterización de cuatro especies de girasol, *Helianthus annuus*, *H. argophyllus*, *H. debilis* y *H. tuberosus*, a nivel molecular, por la técnica 'plastid *trnH-psbA* intergenic spacer'. La secuencia *trnH-psbA* fue elegida con el objetivo de crear el sistema de "código de barras DNA" (Kress *et al.*, 2005), que serviría como herramienta para la identificación de las especies vegetales y muestras. La región plastidial está fortificada con la técnica PCR junto con la utilización de los "primers" específicos y secuencionado en el secuenciador automático de DNA "ABI Prism 3730". La calificación de variabilidad de las secuencias de intra- e interespecies fue realizada con el fin de calificar la precisión de la técnica utilizada. El secuencionamiento de las partes de DNA en las dos direcciones, fue satisfactoriamente exacto, y fue superado el problema de distorsión de la imagen dentro de las regiones microsatélites.

Tras la ordenación de las secuencias, la determinada variabilidad de interespecies era baja o no existía. La variabilidad de interespecies SNP, de los fragmentos indel y de longitud de los fragmentos SSR, era suficiente para la identificación irrebatible de cada especie por separado.

## IDENTIFICATION D'ESPÈCES SAUVAGES DE TOURNESOL PAR UNE SÉQUENCE ADN PLASTIDE SPÉCIFIQUE

RÉSUMÉ

Les quatre espèces de tournesol, *Helianthus annuus*, *H. argophyllus*, *H. debilis* et *H. tuberosus* ont été caractérisées au niveau moléculaire par l'espaceur intergénique plastide trnH-psbA. La séquence trnH-psbA a été choisie dans le but de développer un système de „code-barres ADN" (Kress *et al.*, 2005) comme outil d'identification d'espèces et d'échantillons. La région plastide a été amplifiée par la méthode PCR avec des amorces spécifiques et séquencée avec un séquenceur ADN automatique "ABI prisme 3730". La variation de séquence intra et interspécifique a été évaluée pour déterminer la précision de la technique utilisée. Le séquençage des parties d'ADN dans les deux directions a été suffisamment exact et le problème de dérapage polymérase à l'intérieur des régions microsatellites a été maîtrisé. Après organisation de la séquence, une variabilité interspécifique faible ou inexistante a été détectée tandis que la variabilité interspécifique due aux SNP, indels et longueur de SSR suffisait à identifier sans ambiguïté chaque espèce.

Presented at: