

**RECHERCHE DE SYNTENIE ENTRE LES GENOMES D'*HELIANTHUS ANNUUS*,  
*ARABIDOPSIS THALIANA*, ET *BETA VULGARIS* PAR L'EMPLOI D'ESTs  
PROVENANT DE CES ESPECES**

**INVESTIGATION OF SYNTENIC RELATIONSHIPS BETWEEN THE  
*HELIANTHUS ANNUUS*, *ARABIDOPSIS THALIANA* AND *BETA VULGARIS*  
GENOMES REVEALED USING ESTs FROM THESE SPECIES**

*Isabelle Dominguez*<sup>a</sup>, *Abdelali Barakat*<sup>b</sup>, *Stephen R. Barnes*<sup>a</sup>, *Simon T. Berry*<sup>a</sup>

<sup>a</sup> SES-Europe NV/SA, Advanta Biotechnology Laboratory, Industriepark, Soldatenplein Z2,  
nr.15, B-3300 Tienen, Belgium  
Fax : 00-32-(0)16-80-82-63

<sup>b</sup> Laboratoire de Physiologie et Biologie Moléculaire des Plantes - UMR CNRS 5545,  
Université de Perpignan, 52, Av. de Villeneuve, 66860 PERPIGNAN Cedex  
Fax : 00-33-(0)4-68-66-84-99

**Résumé**

Les hybridations Southern d'ESTs d'*Arabidopsis* sur le tournesol conduisent en général à la formation de bandes de faible qualité, difficiles à cartographier et à utiliser dans une étude de synténie. Ces résultats diffèrent de ceux obtenus avec la betterave sucrière, où près de 50% des ESTs d'*Arabidopsis* ont pu être cartographiés. Nous concluons que la taille des génomes, ainsi que les divergences nucléotidiques, incluant des différences dans l'utilisation du code génétique, sont des facteurs déterminants. Afin d'éviter ce problème et tenter d'établir des relations synténiques entre le tournesol, la betterave et *Arabidopsis*, nous avons sélectionné 115 ESTs du tournesol fortement homologues à des séquences publiques d'*Arabidopsis*; 64 séquences hétérologues ont été cartographiées *in silico* sur *Arabidopsis* et 30 par RFLP sur la betterave. Grâce à ces premières données, deux loci étroitement liés sur la betterave et le tournesol ont été déterminés.

**Abstract**

The use of *Arabidopsis* ESTs as probes in Southern hybridisation to sunflower DNA generally yields poor quality signals, insufficient for mapping and the study of synteny between the species. This contrasts sharply with sugar beet, where approximately 50% of the *Arabidopsis* ESTs can be mapped. We conclude that genome size, as well as sequence divergence, including differences in codon usage, are contributory factors. In order to overcome this problem and to try to establish synteny between sunflower, sugar beet and *Arabidopsis*, we have selected 115 sunflower ESTs for their high homology with published *Arabidopsis* sequences; 64 heterologous sequences have been mapped *in silico* in *Arabidopsis*, and 30 by RFLP in sugar beet. From these limited data set, two loci have been identified that show tight linkage in both sunflower and sugar beet.

L'élaboration de cartes génétiques ainsi que la localisation de gènes représentent des outils essentiels dans les programmes de sélection assistée par marqueurs. Trois cartes génétiques du tournesol (*Helianthus annuus*) ont été publiées jusqu'à présent<sup>1,2,3</sup>. Le même effort est porté sur d'autres espèces végétales telles que la betterave<sup>4</sup>, la tomate<sup>5</sup>, le maïs<sup>6</sup>...

La possibilité d'échanger les connaissances acquises sur les différentes espèces végétales permettrait une avancée scientifique remarquable dans le domaine de la génétique des plantes. Les espèces à intérêt agronomique, telles que le tournesol ou la betterave (*Beta vulgaris*), pourraient, en particulier, bénéficier des données cumulées sur les plantes modèles, telle l'*Arabidopsis*. Les études de synténie, menées précédemment chez certaines espèces dicotylédones (dicot) phylogénétiquement proches, ont permis de démontrer la conservation de régions chromosomiques, comme, par exemple, entre la pomme de terre et la tomate<sup>7</sup> ou au sein des *Brassica*<sup>8</sup>. Chez les monocotylédones (monocot), l'observation de nombreuses colinéarités entre les génomes a débouché sur la construction d'une carte commune à cet ordre végétal<sup>9</sup>. Plus récemment, des régions homéologues ont également été reportées entre une dicot (*Arabidopsis*) et une monocot (le riz)<sup>10</sup>.

Notre étude fait partie d'un programme subventionné par la commission européenne (dossier PL96-2170), nommé EuDicotMap. Ce projet a pour objectif majeur d'établir des liens synténiques entre *Arabidopsis* et 7 espèces dicotylédones d'intérêt agronomique. Nous présentons, dans ce document, un premier aperçu de synténie pouvant exister entre les génomes du tournesol, de la betterave et d'*Arabidopsis*.

## **Matériels et méthodes**

**Matériel végétal** Deux populations F2 de betterave, dont les cartes génétiques étaient liées par un groupe de marqueurs moléculaires communs, ont été utilisées. Une population de 160 individus était issue d'un croisement entre deux lignées de betterave sucrière et la deuxième, constituée de 161 individus, provenait d'un croisement entre une betterave et une blette.

**Sélection des Expressed Sequence Tags (ESTs)** Des ESTs d'*Arabidopsis*, de betterave et de tournesol ont été utilisés dans cette étude :

- 180 ESTs d'*Arabidopsis* ont été sélectionnés dans la base de données *Arabidopsis*, sur base de leur forte homologie avec des ESTs du riz.
- 35 ESTs de betterave provenaient de deux banques cDNA (bourgeons et racines) et présentaient une forte homologie avec des protéines d'*Arabidopsis* (BlastX,  $E < 10^{-40}$ ).
- 115 ESTs du tournesol étaient issus du séquençage partiel en 3' de cDNA (banque cDNA de plantules étiolées). Ces ESTs possédaient au moins une séquence nucléotidique hétérologue chez *Arabidopsis* (BlastN,  $E < 0.001$ ) et avaient précédemment été cartographiés sur le tournesol<sup>1</sup>. Ces séquences ont été désignées par le numéro d'accèsion Genbank d'une des séquences hétérologues d'*Arabidopsis*, suivi par les lettres *Ha*.

**Hybridation Southern des ESTs d'*Arabidopsis* et du tournesol** Les hybridations Southern ont été menées comme décrit précédemment<sup>1,4</sup>, en diminuant les conditions de stringence des milieux d'hybridation et de lavage.

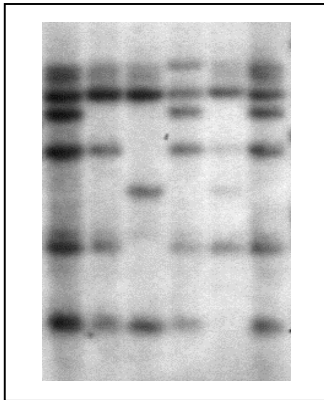
**Cartographie des ESTs du tournesol sur *Arabidopsis* et la betterave** Les séquences d'*Arabidopsis* qui révélaient une homologie significative (BlastN,  $E < 0.001$ ), ont été cartographiées *in silico*, grâce aux données génétiques et physiques disponibles chez cette espèce. La cartographie des ESTs du tournesol sur les populations de betterave a été réalisée par méthode RFLP<sup>4</sup> en utilisant le logiciel MapMaker, version 3.0<sup>11</sup>.

**Analyse de l'utilisation du code génétique** Des ESTs du tournesol (31) et de la betterave (35), présentant une forte homologie avec des ESTs d'*Arabidopsis* (BlastX<math>10^{-40}</math>) ont été sélectionnés. Les séquences protéiques fortement conservées ont été utilisées afin de comparer l'usage des codons entre *Arabidopsis* et les deux autres espèces. La comparaison a été effectuée par un test du  $\chi^2$  grâce à l'analyse de 4000 codons par espèce.

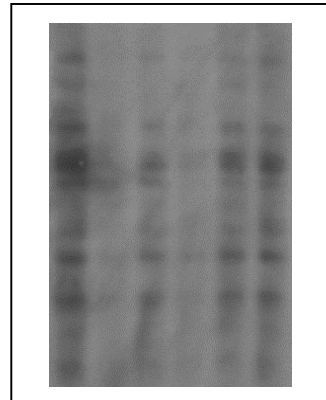
## Résultats et discussion

**Hybridation Southern** Les résultats des hybridations des sondes *Arabidopsis* sur le tournesol et la betterave diffèrent essentiellement par la qualité des bandes obtenues. La quasi moitié de ces sondes, ainsi que celles du tournesol, ont révélé des bandes intenses et distinctes chez la betterave. Sous les mêmes conditions expérimentales, les bandes issues des hybridations interspécifiques, chez le tournesol, étaient trop faibles et diffuses pour permettre leur cartographie. (**Fig.1**).

**Figure 1** : hybridation de l'EST d'*Arabidopsis* 91P17T7 ( $\alpha$ -tubulin) sur la betterave (**Fig 1a**) et sur le tournesol (**Fig 1b**)



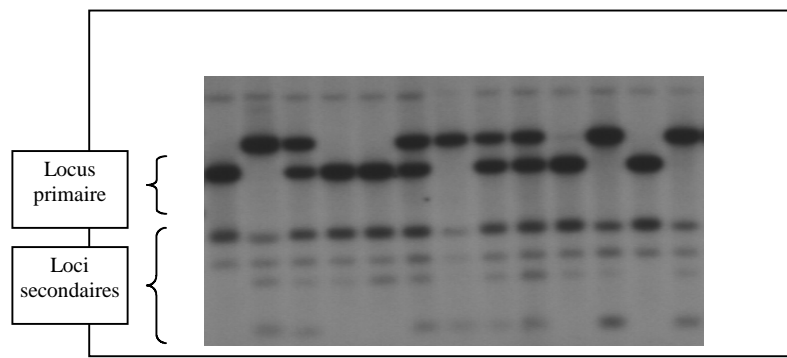
**Figure 1a** : betterave



**Figure 1b** : tournesol

Digestion de l'ADN : DraI ; Migration : 40V, 18 h ; Lavage hybridation : 2XSSC, 0.1% SDS, 60°C

Les différences de qualité des hybridations des sondes d'*Arabidopsis* sur le tournesol et la betterave peuvent être liées à différents facteurs, génomiques et /ou phylogénétiques. Ainsi, la grande taille du génome du tournesol, 20 fois plus grand que celui d'*Arabidopsis* et seulement 4 fois plus que celui de la betterave<sup>12</sup>, perturbe certainement la qualité des hybridations sur le tournesol. Des différences importantes dans le degré d'homologie des séquences nucléotidiques conservées pourraient être une cause supplémentaire des résultats obtenus. Pour la protéine delta-9 désaturase, par exemple, le pourcentage d'homologie entre les gènes U91339 et U91340 du tournesol est de 89%<sup>13</sup> et leur hybridation, en condition standard, sur le tournesol, montrent de faibles bandes pour les loci dupliqués, difficiles à cartographier (**Fig.2**).



**Figure 2 : hybridation de l'EST U91340 du tournesol (protéine delta-9 désaturase) sur le tournesol**

Digestion de l'ADN : DraI ; Migration : 30V, 20h ; Conditions d'hybridation standard

Le degré d'homologie entre ces gènes et leur hétérologue d'*Arabidopsis* (AC002333cds) n'est, elle, que de 75%. L'alignement des séquences homologues et hétérologues semble indiquer, pour ce gène, que les différences d'homologie proviennent notamment des dernières bases des codons (**Fig.3**).

**Figure 3 : alignement d'une partie du gène U91339 (protéine delta-9 désaturase) avec son homologue U91340 et son hétérologue d'*Arabidopsis* AC002333cds**

```

U91339      797  ACCGCGCGTACGCAAAGGATTACGGGGATCTAAAACCTGGCTCAAATTTGTGGTACGATCGCC
U91340      772  ACAGCCCGGCACGCAAAGGAGCATGGTGACGTGAAGCTGGCTCAAATGTGCGGGATAATTGCA
AC002333cds 730  ACAGCCCGCCAAGCCAAAGAGCACGGGGACATCAAACCTAGCCCAAATATGTGGCACAATAGCT

U91339      860  GCTGATGAAAAAAGGCACGAAACCGCCTACACTAAAATCGCTGAAAACTATTCGAAATCGAC
U91340      835  GCTGATGAAAAAAGGCACGAAACCGCCTACACAAAAATAGTAGAAAACTCTTCGAAATTGAC
AC002333cds 793  GCAGACGAGAAGCGTCATGAAACAGCATAACCAAGATAGTTGAAAAGCTCTTTGAGATTGAT
  
```

Une utilisation spécifique du code génétique par les espèces aurait donc pu jouer un rôle dans l'évolution des séquences nucléotidiques. Dans ce domaine, notre analyse a montré qu'il existe une différence dans l'utilisation de certains codons entre *Arabidopsis* et les deux autres espèces (**Tab.1**).

**Tableau 1 : différences significatives dans l'utilisation du code génétique entre *Arabidopsis* et les espèces végétales tournesol et betterave pour des protéines fortement conservées**

*différences entre le tournesol et Arabidopsis*

codons	GGA	GGG	GGC	TTG	CCG	CCT	TCT	ACC	ACT	GTC	GTT
acides aminés	Gly	Gly	Gly	Leu	Pro	Pro	Ser	Thr	Thr	Val	Val
Résultats du $\chi^2$	0.001	0.002	0.03	0.034	0.002	0.002	0.006	0.05	0.001	0.004	0.006

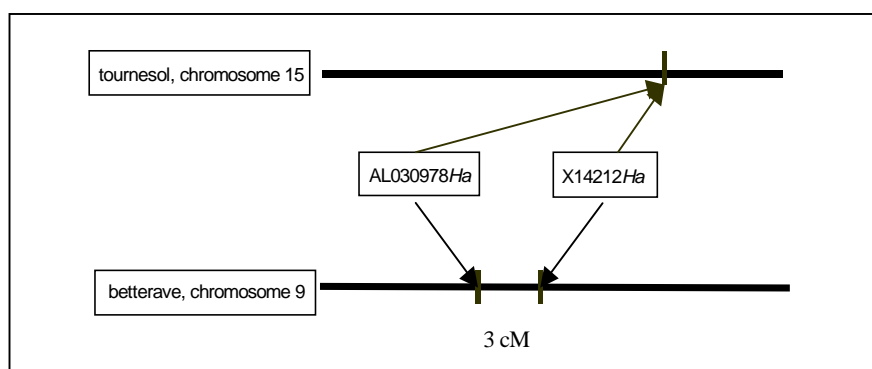
*différences entre la betterave et Arabidopsis*

codons	GCA	GCG	ATC	CTC	TTC	TTT	CCG	ACA	ACG	TAC	TAT
acides aminés	Ala	Ala	Ile	Leu	Phe	Phe	Pro	Thr	Thr	Tyr	Tyr
Résultats du $\chi^2$	0.007	0.021	0.028	0.014	0.001	0.001	0.025	0.034	0.003	0.032	0.032

Résultats du  $\chi^2$  : valeur du test de comparaison (test du chi-2) de l'utilisation du codon par les 2 espèces végétales comparées

**Cartographie des ESTs du tournesol** Pour l'instant, la cartographie *in silico* de 64 séquences a été déterminée chez *Arabidopsis*. La comparaison des génomes du tournesol et d'*Arabidopsis*, grâce à cette première série de marqueurs moléculaires, n'a pas permis d'identifier de région conservée. Cependant, 161 séquences hétérologues supplémentaires ont été déterminées et seront prochainement cartographiées sur *Arabidopsis*. D'autre part, de nouveaux hétérologues, classés suivant leur degré d'homologie avec les ESTs du tournesol, seront certainement révélés, grâce au séquençage complet des ESTs du tournesol utilisés, et à l'achèvement du séquençage du génome d'*Arabidopsis*. L'analyse de ces nouvelles données devrait augmenter les chances d'observer des blocs conservés.

30 ESTs du tournesol ont également été cartographiés sur la betterave permettant une première analyse de synténie entre le tournesol et la betterave. Celle-ci a révélé la présence de deux marqueurs étroitement liés (AL030978Ha et X14212Ha), respectivement sur les chromosomes 15 du tournesol (0 cM) et 9 de la betterave (3cM) (**Fig.4**).



**Figure 4 : cluster de deux marqueurs chez le tournesol et la betterave**

Dans l'avenir, de nouveaux marqueurs devront être cartographiés autour de cette région pour déterminer la longueur des segments conservés. Cette région de synténie sera également à rechercher chez *Arabidopsis* par la cartographie de l'ensemble des séquences hétérologues à ces deux ESTs.

Des ESTs de betterave présentant une forte homologie protéique avec *Arabidopsis* sont également disponibles. Leur cartographie chez le tournesol et *Arabidopsis* devrait permettre d'augmenter le nombre de marqueurs communs aux 3 espèces végétales.

**Duplications et conséquences sur la recherche de synténie** Les hybridations interspécifiques représentent un élément majeur de l'évolution des plantes<sup>14</sup> et il est donc probable que de nombreuses espèces végétales présentent des régions homéologues. Cela semble être véridique, chez *Arabidopsis* et la betterave, sur lesquelles notre recherche de séquences hétérologues a révélé la présence de nombreux loci dupliqués (**Tab.2**).

**Tableau 2 : répartition des ESTs du tournesol en fonction du niveau de duplication chez *A.thaliana***

loci dupliqués chez <i>A.thaliana</i>	1	2	3	4	5	6	7
ESTs du tournesol	60	26	11	12	5	0	1

Remarque : les séquences d'*Arabidopsis* hétérologues aux ESTs du tournesol ont été déterminées par BlastN, E<0.001

Le tournesol, en tant qu'espèce allopolyploïde, devrait également présenter des duplications<sup>15</sup>. Au sein d'un même génome, la détection de régions homéologues est laborieuse et délicate. Dans ce but, plusieurs méthodes peuvent être utilisées, comme l'élaboration de cartes génétiques fortement saturées<sup>8</sup>, ou l'analyse du déterminisme génétique d'un caractère<sup>16</sup>. Cependant, chez le tournesol, comme chez la betterave, aucun segment homéologue n'a encore été observé<sup>1, 17</sup>. Ce constat est probablement dû, en grande partie, à la difficulté de repérer et cartographier les loci secondaires par hybridation (**Fig.2**) ainsi qu'à l'utilisation de sondes monocopies<sup>1</sup>. Chez le tournesol, 4 marqueurs ont été récemment localisés dans le même ordre, sur des segments de 20-30 cM des chromosomes 16 et 8, ce qui semble indiquer la présence d'une région dupliquée (Advanta, résultats non publiés).

La méconnaissance intrinsèque des génomes à comparer rend les recherches de synténie plus difficiles. Cependant la détection des séquences homologues pourrait être accélérée, dans l'avenir, par l'utilisation de logiciels d'alignement de séquences protéiques et nucléotidiques, grâce à la génération de nouveaux ESTs et l'achèvement du séquençage du génome d'*Arabidopsis*.

## Conclusion générale

Les connaissances actuelles sur l'origine des espèces végétales considérées ainsi que les événements moléculaires intragénomiques survenus au cours de l'évolution restent insuffisants et risquent de rendre les études de synténie difficiles. Les premières études de synténie ont toutefois permis de révéler, d'une part la présence de chromosomes homéologues entre des espèces phylogénétiquement proches, mais également celle plus générale de blocs conservés entre des espèces plus distantes comme la tomate et *Arabidopsis*<sup>18</sup> ou même entre une dicot (*Arabidopsis*) et une monocot (le riz)<sup>10</sup>. Il semble donc probable que, malgré les restructurations moléculaires intervenues au cours de l'évolution d'*Arabidopsis*, du tournesol et de la betterave, des régions homéologues soient encore conservées entre les génomes de ces espèces. Afin d'augmenter les chances d'identifier ces régions de synténie, un effort particulier devra être mis en œuvre pour saturer les cartes génomiques et déterminer l'ensemble des séquences dupliquées au sein d'un même génome.

1. Berry, S.T., A.J. Leon, P. Challis, C. Livini, R. Jones, C.C. Hanfrey, S. Griffiths, A. Roberts, 1996, Construction of a high density, composite RFLP linkage map for cultivated sunflower (*Helianthus annuus* L.), *Proceedings of 14<sup>th</sup> International Sunflower Conference*
2. Gentzbittel, L., E. Mestries, S. Mouzeyar, F. Mazeyrat, S. Badaoui, F. Vear, D. Tourvieille de Labrouhe, P. Nicolas, 1999, A composite map of expressed sequences and phenotypic traits of the sunflower (*Helianthus annuus* L) genome, *Theor Appl Genet* **99**:218-234
3. Jan, C.C., B.A. Vick, J.F. Miller, A.L. Kahler, E.T.I. Butler, 1998, Construction of an RFLP linkage map of cultivated sunflower, *Theor Appl Genet* **96** : 15-22
4. Barnes, S., G. Massaro, M. Lefebvre, M. Kuiper, E. Verstege, 1996, A combined RFLP and AFLP genetic map for sugar beet, *Proceeding of the 59<sup>th</sup> IIRB Congres* (cl.II)
5. Bernatzky, R., Tanksley S.D., 1986, Towards a saturated linkage map in tomato based on isozymes and random cDNA sequences, *Genetics* **112** : 887-898
6. Helentjaris, T., 1987, A genetic linkage map for maize based on RFLPs, *Trends Genet* **3** :217-221
7. Bonierbale, M.W., R.L. Plaisted, and S.D. Tanksley, 1988, RFLP Maps Based on a Common Set of Clones Reveal Modes of Chromosomal Evolution in Potato and Tomato, *Genetics* **120**:1093-1103
8. Teutonico, R.A., Osborn T.C., 1994, Mapping of RFLP and qualitative trait loci in Brassica rapa and comparison to the linkage maps of B.napus, B. oleracea, and *Arabidopsis thaliana*, *Theor. Appl. Genet.* **89**:885-894
9. Moore, G., K.M. Devos, Z. Wang, and M.D. Gale. 1995a, Grasses, line up and form a circle. *Curr. Biol.* **5**: 737-739
10. Dodeweerd, A, C.R. Hall, E.G. Bent, S.J. Johnson, M.W. Bevan, I. Bancroft, 1999, Identification and analysis of homoeologous segments of the genomes of rice and *Arabidopsis thaliana*, *Genome* **42**: 887-892
11. Lander, E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E., Newburg L., 1987, MAPMAKER : an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations, *genomics* **1**, 174-181
12. Bennett, M.D., J.B. Smith, J.S. Heslop-Harrison, 1982, Nuclear DNA amounts in angiosperms, *Proc. R. Soc. Lond.* **B 216**:179-199
13. Hongtrakul, V., M.B. Slabaugh & S.J. Knapp, 1998, DFLP, SSCP, SSR markers for delta9-stearoyl-acyl carrier protein desaturases strongly expressed in developing seeds of sunflower: intron lengths are polymorphic among elite inbred lines, *Molecular Breeding* **4**: 195-203
14. Rieseberg, L.H., 1997, Hybrid origins of plant species, *Annu. Rev. Ecol. Syst.* **28**: 359-389
15. Heiser C.B., Smith D.M., 1955, New chromosome numbers in *Helianthus* and related species, *Proc. Ind. Acad. Sci* **64** : 250-253
16. Lee, J.M, A.L Bush, J.E. Specht, R.C. Shoemaker, 1999, Mapping of duplicate genes in soybean, *Genome* **42** : 829-836
17. Halldén, C., D. Ahrén, A. Hjerdin, T.Säll, N.-O. Nilsson, 1998, No Conserved Homeologous Regions Found in the Sugar Beet Genome, *Journal of Sugar Beet research* **35** : Nos 1-2
18. Matern, A.L., T.M. Fulton, D. Zamir, S. Grandillo, D.J. Schneider, S.D. Tanksley, 1999, A search for tomato-*Arabidopsis* syntenies, *PAG VII Conference*, San Diego, poster P208