**Genoplante: The "winter oilseed rape" program**

Oléagineux, Corps Gras, Lipides. Volume 10, Numéro 3, 212-5, Mai 2003, Colza : enjeux et nouvelles synergies de la recherche, RECHERCHE

**Auteur(s) :** Patrick VINCOURT Michel RENARD , Euralis Semences, avenue Gaston Phoebus, 64230 Lescar INRA, Station de Génétique et d'Amélioration des Plantes, BP 29, 35650 Le Rheu .

**Author(s) :** Patrick VINCOURT Michel RENARD

**Summary :** The French plant genome initiative GENOPLANTE, associating public research and private companies involved in research for agriculture breeding or agrochemicals , supported several projects in oilseed rape genomics. Some of them intensively used Arabidopsis resources. Beside applied projects including gene cloning or QTL identification, GENOPLANTE succeeded in establishing a set of biological and bioinformatics resources available for the genetic improvement of this crop.

ARTICLE

**Auteur(s) :** Patrick VINCOURT[1], Michel RENARD[2]

[1] *Euralis Semences, avenue Gaston Phoebus, 64230 Lescar*
[2] *INRA, Station de Génétique et d'Amélioration des Plantes, BP 29, 35650 Le Rheu*

GENOPLANTE, the French Plant Genome initiative, has been created four years ago to help the scientific community as well as the seed and agrochemical industries to enter this new field of knowledge opened at the end of the twentieth century. Beside long term approaches or fundamental research mainly based on the model plant *Arabidopsis thaliana*, the research activity has been devoted to the crops having an economical importance in Europe, like wheat, maize, rapeseed, sunflower and pea.

The specificity of the GENOPLANTE Oilseed Rape project is that very early in the process, the data obtained and the resources developed on Arabidopsis have been used to successfully translate on rapeseed the knowledge obtained on the model plant. *Chart 1* shows how the research projects have been using the existing resources of Arabidopsis or the new resources developed within GENOPLANTE.

These projects are ordered below in distinguishing resources, QTL mapping, cloning projects, and specific projects mainly developed on Arabidopsis with an aim to acquire fundamental knowledge on the cultivated crop.

**Resources**

***GOP-T1: Rapeseed and Radish BAC libraries***

In order to allow the cloning of some genes of interest, two BAC libraries have been built during the first phase of GENOPLANTE:

*A Rapeseed library of 76000 clones, representing 12 genome equivalents, has been prepared with the cultivar DARMOR Nain* ("dwarf" mutant)

*A Radish library of 120000 clones, with an average size of 100 – 200 kb and representing 23 genome equivalents has been built specially for the cloning of the Rfo* gene (GOP-A1 project)

These resources are made available to the other projects through two- or three-dimensional pools to make their screening by PCR easier.

***GOP-T6: The Genoplante Oil seed rape Chip***

Grain filling is the ultimate phase in yield elaboration, for quantitative as well as for qualitative aspects. Any genetic mechanism during this phase may act as a limiting factor for productivity. cDNA libraries built at different critical stages during grain filling are a useful tool to understand these mechanisms. The objectives of this project are to:

1. Identify and sequence genes that are expressed during seed development and maturation in *B. napus*. This will provide probes for the mapping of seed traits and enable the cataloguing of alleles that may contribute to the agronomic performance of elite lines.

2. Array seed expressed genes. Probing this array will provide information as to the developmental profile and abundance of seed-expressed mRNAs in *B. napus* varieties or mutant lines grown in standard or stressed conditions.Sequence data and arrays will also be useful to identify *B. napus* homologues of genes identified as important for seed development in *Arabidopsis*. Together this data will aid in the identification of *B. napus* genes that contribute to seed yield and quality in normal and adverse conditions. *B. napus* was grown, flowers tagged and seeds collected for mRNA isolation at 5-day intervals. The timing of critical stages such as induction of reserve synthesis and start of maturation were confirmed prior to extraction of RNA by morphological and biochemical analysis. The *B. napus* line chosen for mRNA isolation was Jet Neuf, an older European winter line, as this is relevant to the material used in European research in this species.The project has produced a set of 29 373 seed derived ESTs in the two years of the first phase of GÉNOPLANTE. Included in this population are representatives of entirely new genes, and a large number of genes previously not yet detected in *B. napus*. This resource will be developed further in the next phase of the programme by putting selected representatives for all contigs onto a DNA-chip, which will allow us to study grain filling in more detail as set out in the targets of this project. During the next year, toward the development of an oil seed rape microarray, these sequences as well as the rape another sequences (Rapeseed GOPA1 project) were analyzed by "clustering" and our "unigene set" bioinformatics pipeline to produce a non-redundant set of genes or "unigene set" containing 13,083 genes. This also included the development of the rape Unigene set, massive PCR amplifications of cDNA inserts for use as tethered probes and the perspective uses of this Chip.

***GOP-T4/GOP-T7: Gene homology between arabidopsis and rapeseed: high throughput mapping of SNP's in oilseed rape***

Beyond the knowledge of plant genetic evolution, comparing maps of related species should ease future localization and cloning of genes controlling quantitative and qualitative traits of economic interest for plant breeding purposes.The development of Amplified Consensus Genetic Markers is based on the frequent conservation of peptide sequences within species of the same taxonomic family and on the potential polymorphism within genic sequences. This simple methodology uses a PCR technology and is effective, starting from an *Arabidopsis* sequence, to "sequence without cloning" the homologous genes in the *Brassica* tribe. Further more this tool contributes to a better understanding of the conservation of peptide sequences, permits to study the intra-genic polymorphism within one species, and allows to examine the synteny between species, even in an amphidiploid genome.The identification of Single Nucleotide Polymorphisms within gene sequences from different rapeseed cultivars is another important point as any candidate gene could be potentially mapped, eventually using of high throughput technology.A two-step process is developed: 1. Design of consensus primers to obtain rapeseed gene sequences, alignment of the different homologous sequences. 2. Design of sequence specific primers, sequencing on various rapeseed lines, SNP detection.

Following the GOP-T4 Génoplante I project and based on a strong collaboration with the French Centre National de Génotypage, the mass spec technology is used to map within 12 mapping populations, including the GOP-R3 and GOP-Q5 populations (1750 individuals), the 176 SNP's detected in sequences of 165 *Brassica* genes homologous of *Arabidopsis* genes and with known function. Despite of the quite high genetic diversity within the mapping population, only one half of the sequences are showing polymorphism. The "copy primers" (probes) have been performed for about 80 genes, and the genotyping data have been provided to the teams in charge of mapping.

***GOP-T9: A Physical-functional mapping of the BAC rapeseed library based on the synteny with* Arabidopsis thaliana**

PCR based, Physical – Functional Markers derived from Arabidopsis coding sequences, preferably with a known, agronomicaly interesting function ("entry points") or EST provided by GOP-T6 are used to anchor the Génoplante BAC rapeseed library through 3D-pool screening. It has been shown that 384 PCR reactions are needed to identify a positive reaction for any PFM within the 36 K clones BAC library. Up to now, 800 PFM have been defined, and 70 PFM are screened on the BAC library each week. This should provide a first database to build a physical map, but the final map will not be obtained before the end of this project.The most interesting genes (300 targets) will be also genetically mapped to obtain a consensus genetic map with consistent link to this "physical" database. It appears, from the first set of 200 PFM, that the polymorphism within a set of 16 rapeseed lines is quite low – less than 15% – and that these markers would be only usable as dominant marker.

**QTL mapping and candidate genes**

***GOP-R3: Resistance of* Brassica napus *to Sclerotinia scl***

*Sclerotinia sclerotiorum* is a parasitical fungus causing major damages and today, a fungicide is usually applied, with a clear risk of selection of resistant strains of the fungus. In western Europe, there is no resistant cultivar, and the use of exotic genetic resources is subject to the development of more precise breeding tools,to avoid the introduction of unfavourable traits like high glucosinolate content or susceptibility to *Phoma*. The aim of this project is to provide such tools, through QTL analysis, and possibly to give the first basis for gene cloning or identification.
With measurements of the resistance on stems and on leaves as phenotypic trait, several QTL's explaining each 8 to 11% of the variability have been detected in two segregating populations. In both cases, the favorable alleles are coming from the Asiatic source of germplasm.
The transcripts of infected vs. non infected leaves of one of the *Brassica* resistant parental line have been compared in using the 10 K RHOBIO Arabidopsis microarray. From about 10.000 transcripts, 151 (resp. 136) have been sorted as down- (resp. up -) regulated, including in both cases about 20% of genes involved in stress tolerance, signalization, or disease resistance – becoming a first set of candidate genes – and 40% of non-annotated sequences. The candidate genes will be mapped on the segregating populations and studied in expression experiments.

***GOP-Q5: Mapping QTL for oil content in rapeseed***

Three segregating populations have been or are under evaluation for QTL mapping of oil content, over years and environments, mainly with SSR markers. From the first population, 6 QTL's have been identified, beside of one QTL apparently linked to the “dwarf” (*bzh)* mutant.Some genes involved in the lipid biosynthesis pathway will be mapped by the GOP-T7 project, as well as some other candidates already detected as co-localized with Oil content QTL's from previous studies, and the EST corresponding to 117 genes provided by GOP-T6.

**Cloning genes of interest**

***GOP-A1: Identification and cloning of Rfo***

The OGU-INRA cytoplasm is widely used by the seed companies to develop hybrid varieties, but after ten years of breeding efforts, some linkage between the gene involved in fertility restoration and some unfavourable traits, probably all carried by the Radish introgression, made the breeding of performing male line very difficult.For the purpose of understanding the biological ways of fertility restoration as well as for the eventual purpose of creating a “clean” restorer line through genetic engineering, the project was devoted to the cloning of this *Rfo* locus.Using the radish BAC library (GOP-T1 Génoplante 1 project), a high size recombinant population and some synteny information from Arabidopsis, the identification of 3 PPR proteins as candidate genes for the *Rfo* locus has been obtained, from which one (PPR B) seems to be the most promising (Desloire S. et al, 2003).Some markers of interest for hybrid rapeseed breeding will be provided to the breeders, and the constructs allowing the final demonstration of the role of PPR B will be made available to the research teams.

***GOP-T2: Positional cloning of Clg1 (cleïstogamous mutant)***

Oilseed rape is naturally an allogamous crop, with an ability to disseminate its own pollen as well as to receive the pollen of a quite distant field grown with the same crop. A cleistogamous mutant (*clg*) obtained at INRA by chemical mutagenesis has been made available to this project, with the aim to understand how this ability could be genetically controlled, with the purpose to avoid pollen dissemination or to increase the purity of any "specialty" crop.Five PCR-specific markers surrounding Clg1, the major gene involved in the cleïstogamous phenotype have been mapped in a 2000 DH lines segregating population. On the segment covered by these markers, a subset of 283 plants is showing segregant phenotypes, with some abnormalities revealing the need of a more precise phenotypic evaluation of the trait. The same markers have been used to identify a 710 kb segment on ChIV of *Arabidopsis*, allowing a deeper analysis of the synteny. 82 rapeseed BAC clones have been identified, probably including homeologous regions. Four of them are carrying the gene and two markers; they will be used to sequence the non-mutated gene, and then to identify the mutation in comparing the mutated and non-mutated genotype.

**Specific projects on traits of interest**

***GOP-Q3: Characterization of Arabidopsis t. genes involved in grain filling***

Mature seed composition results from a number of highly regulated metabolic and interacting fluxes. Biosynthetic pathways for carbohydrates and lipids are interconnected as both are synthesised from maternal-derived sucrose and share some common intermediates. The underlying assumption is that altered carbon or storage proteins metabolisms may ultimately result in incomplete filling of the seed hence a wrinkled phenotype. The aim of this project was to identify genes important for seed filling, based on the characterization of T-DNA insertion *Arabidopsis* mutants affected in seed phenotype.

About one hundred T-DNA insertion mutants have been firstly isolated by visual screenings for abnormal shrunken seed phenotypes, and are currently under investigations at several levels including genetic (T-DNA tagging of the mutation), biochemical (lipids, carbohydrate and proteins), cytological and molecular (cloning of the plant genomic T-DNA borders).

The functional analysis of few of the most interesting mutants (search for alleles, functional complementation, gene expression, transcriptome analysis, cytology and biochemical analyses) will be carried out to characterize the function of the tagged genes and to clone oilseed homologs.

Sequences will be provided to the GOP-T7 project, to look for polymorphism within *Brassica* and eventually to be mapped.

***GOP-R1: Generation of gene promoters that are activated in Brassica Napus by Phoma lingam infection, and their use for the genetic engineering of disease resistance***

The imperfect fungus *Phoma lingam* (ascogenous state: *Leptospaeria maculans)* causes *Brassica napus* blackleg (or stem canker) with severe economic losses. This project aims to generate a new resistance trait in *B. napus* that is based on the reactivity of several cultivars of oilseed rape to the proteinaceous elicitor (elicitin) cryptogein from *Phytophthora cryptogea*. We proposed (I) to identify,

analyse, and design gene promoters that are functional in *B. napus* and that are induced rapidly and locally by *Phoma lingam*, and (II) to fuse one of the promoters to the cryptogein gene in order to generate transgenic oilseed rape that produces the elicitin upon pathogen attack. To the end of the project, we aim to obtain plants that develop a hypersensitive response and broad-spectrum disease resistance as a consequence from inducible cryptogein production. A successful application of this strategy requires that the elicitin gene is expressed rapidly, locally and only at the time of pathogen challenge.
The promoters of a 5 members family of genes close to Cci7, induced by elicitin and pathogens in rapeseed, are showing only 50-60% of similarity, and focus is now given on the role of intron in the regulation of expression.

Seven genes showing an overexpression under compatible interaction (*Arabidopsis – Peronospora* model) have been identified after experiments with RHOBIO microarrays, northern blot and RT-PCR experiments. The analysis of their promoters is undertaken.

***GOP-Q7: Identification of factors controlling the expression of lipid related genes***

Despite the long standing importance of vegetable oils in nutrition, little is known about the factors controlling the production of these oils in *Brassica napus*. This project aims at isolating some of these factors.
To achieve this objective, several complementary approaches will be used:

1: Isolation of transcription factors from expression analysis of developing *Brassica* seeds.

2: Isolation of mutants affected in oil accumulation from a population of *Arabidopsis* carrying an oil specific reporter construct.

3: Isolation of trans-acting factors using promoters of genes controlling fatty acid synthesis, modification and utilisation in *Brassica napus* via the yeast one hybrid system and search for cofactors via two hybrid interaction screening.

In all cases this will lead to the isolation of candidate transcription factors, that will then be tested *in vitro* and *in vivo* in *Arabidopsis* and *Brassica napus* to verify their function and to assess their effect on lipid biosynthesis in seeds.

**Conclusion: Not the end, but the beginning of a long story**

As it is often the case when new technologies like bioinformatics, robotics, miniaturization, are entering the discovery process, one should consider that the first result of this coordinated effort is that we are learning to learn ! It was quite obvious, at least for the specialists in plant physiology - and may be for the plant breeders – that biology would not become a simple thing just because of the genomics ! But its also clear that the research teams involved in GENOPLANTE in the public or private sector, built a strong set of biological or bioinformatics resources for the ten next years, and also contributed to establish a new way to question the genetic diversity.

Illustrations

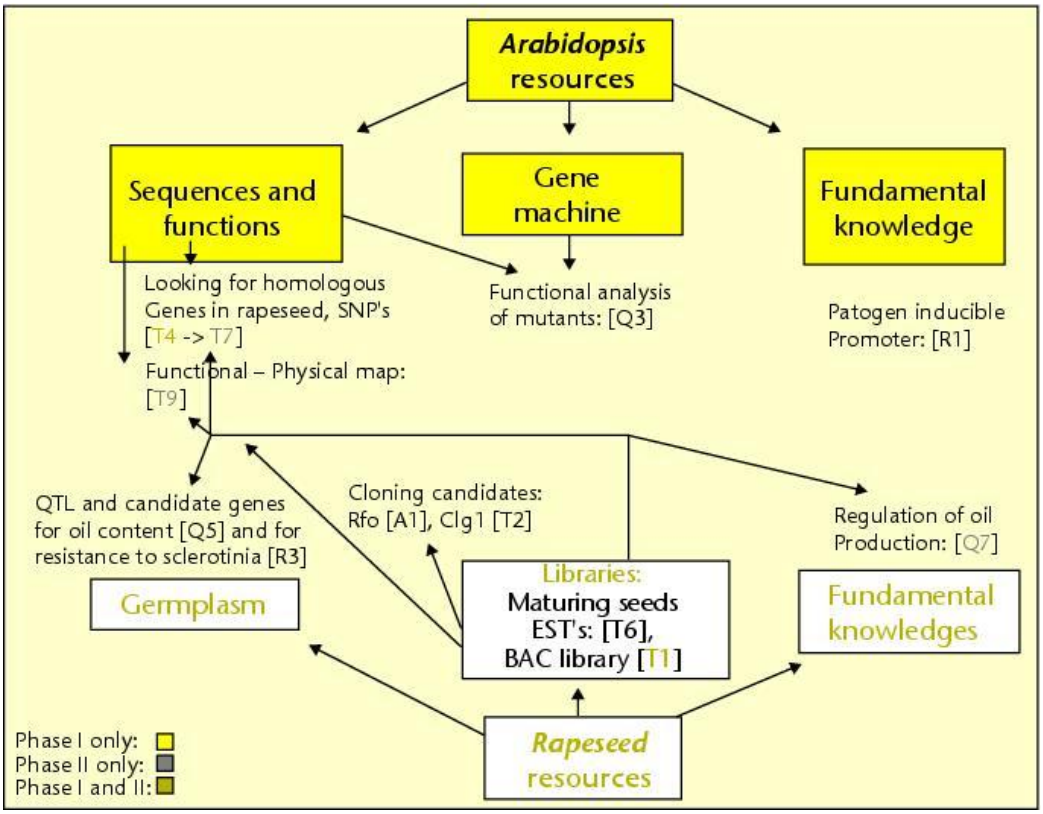


**Chart 1.** *GENOPLANTE. Winter Oilseed Rape General Chart.*