Medicago truncatula cDNA and genomic libraries

Pascal Gamas[1], Frédéric Debellé[1], Hélène Berges[2], Laurence Godiard[1], Andreas Niebel[1], Etienne-Pascal Journet[1], Jérôme Gouzy[1].

1. Laboratoire des Interactions Plantes Micro-organismes CNRS-INRA (LIPM), Chemin de Borde-Rouge – Auzeville, BP 52627, 31326 Castanet Tolosan CEDEX, France, http://www.toulouse.inra.fr/centre/lipm/
E-mail: gamas@toulouse.inra.fr

2. Centre National de Ressources Génomiques Végétales INRA (CNRGV), Chemin de Borde Rouge BP 52627, 31326 Castanet Tolosan CEDEX, France, http://cnrgv.toulouse.inra.fr/
E-mail: hberges@toulouse.inra.fr

Table of content

Tables:

1. *M. truncatula* cDNA libraries

1.1. Standard cDNA libraries

1.1.1. EST production

cDNA clone sequencing is certainly the easiest way to identify genes expressed in a defined physiological situation/organ. Only a few *Medicago truncatula (Mt)* cDNA libraries were reported during the "pre-genomics era", all of them being dedicated to a major legume trait, namely symbiotic interactions, either with *Sinorhizobium meliloti* bacteria (Wilson and Cooper, 1994; Gamas et al., 1996) or with *Glomus versiforme* arbuscular mycorrhizal fungi (Liu et al., 1998). These libraries were used to generate a very limited number of cDNA sequences (sometimes now designated ETs, for expressed transcripts), identified from a screen and sequenced from both ends. One of these libraries was generated following standard procedures (see below) but screened with a subtracted probe in order to identify genes expressed specifically in root nodules (Gamas et al., 1996).

With the advent of genomics, the way cDNA libraries were exploited completely changed, with the massive production of ESTs, corresponding to single pass sequencing of randomly sampled cDNAs, allowing a global analysis of transcript populations (transcriptomes). Such an approach was first published in 1998 for *Mt* on a limited scale, with the characterization of root hair-enriched root tip cDNA clones (Covitz et al., 1998), followed in 2000 by the analysis of clones derived from *Mt* nodules (Gyorgyey et al., 2000). *Mt* EST resources were then dramatically and rapidly increased (see Figures 1 & 2, and Tables 1-3), due to several programs sponsored initially by the NSF *Mt* consortium and the Noble Foundation in the USA, and a CNRS-INRA-Genoscope collaboration in France (Frugoli and Harris, 2001). The majority of ESTs were produced in 2000-2002, but it is interesting to note that new cDNA libraries and ESTs were submitted recently (Figure 1 and 2). By March 2006, more than 227,000 *Mt* ESTs were available, of which about 95% were generated from standard cDNA libraries, *Mt* ranking at the seventh position among plants and second to soybean among legumes (http://www.ncbi.nlm.nih.gov/dbEST/dbEST_summary.html).

In total, 41 non-subtracted cDNA libraries were generated using the same *Mt* genotype (Jemalong A17) and two were made from *Mt* R108 (see tables 1-3). Similar cloning procedures were followed in most cases (38 libraries), which certainly made *in silico* library comparisons more reliable (see section 1.3.): cDNA was synthesised from polyA+-enriched RNA, then directionally cloned into a lambda phage vector (Uni-ZAP XR, Stratagene). pBluescript SK^- plasmids containing cDNA inserts were finally mass-excised using Ex-assist helper phage. One difference worth underlining among procedures for EST production is the fact that clones from three libraries (MtBA, MtBB, MtBC, Tables 1-3) were sequenced from both 3' and 5' ends (Journet et al., 2002), while all the other ESTs were generated by 5' sequencing only. Data from 5' ends are generally more informative on the encoded protein, whereas 3' end sequences can be very useful to distinguish between closely related members of multigene families.

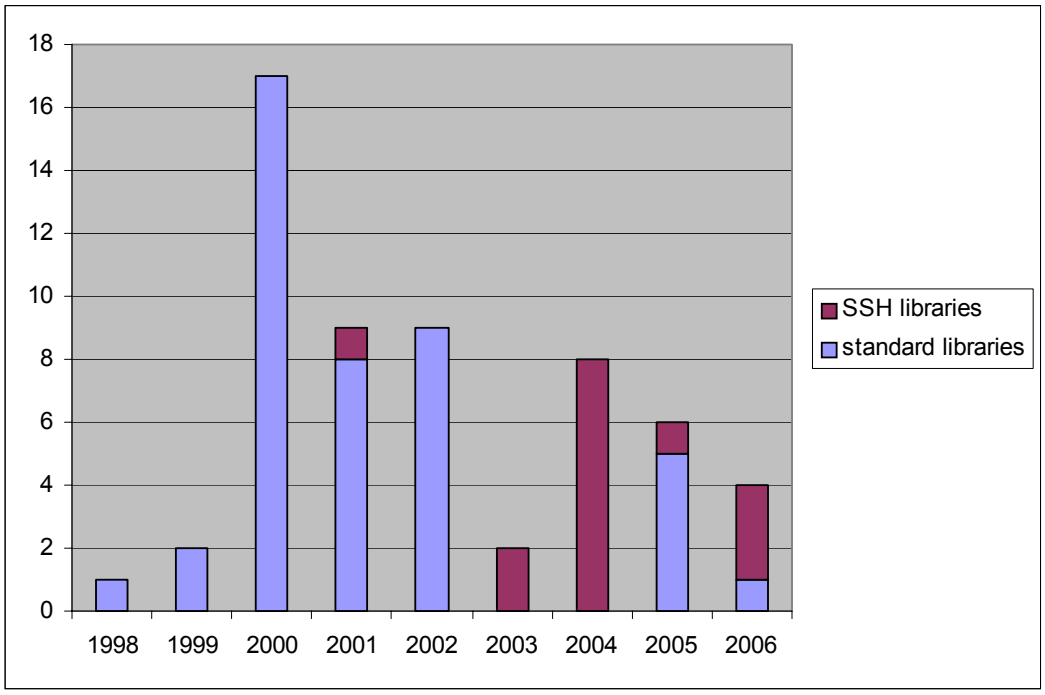

Figure 1: Number of *M. truncatula* cDNA libraries submitted to public databases, corresponding to mid- or large-scale EST programs (as of March 2006) (not considering KVKC and MtUS libraries, corresponding to unigene sets selected from previous ESTs and resequenced)

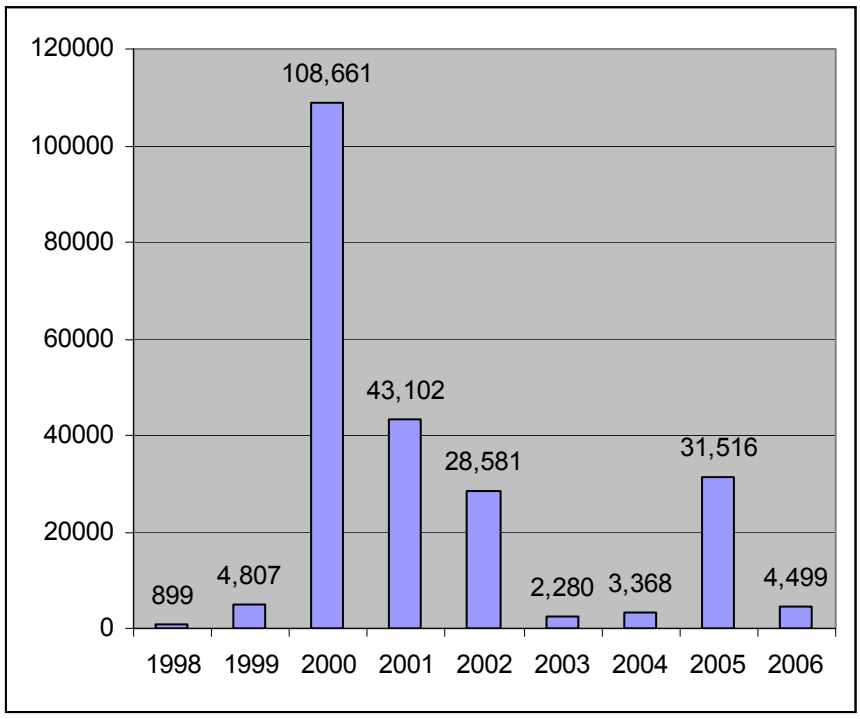

Figure 2: Number of *M. truncatula* submitted ESTs corresponding to mid- or large-scale EST programs (as of March 2006) (not considering KVKC and MtUS libraries, corresponding to unigene sets selected from previous ESTs and resequenced) (not shown).

Figures 3 and 4 indicate the range of conditions represented by *Mt* ESTs. Although some years ago there was a strong bias toward roots and symbiotic organs (nodules and mycorrhizae), this is not as obvious anymore. This reflects the continuous broadening of the *Mt* community and of the biological questions addressed using this model species. Thus, there are now almost as many ESTs representing aerial parts (vegetative plus reproductive organs) as roots, mycorrhizae and nodules. It can also be seen that there are currently more ESTs generated from libraries corresponding to abiotic / biotic stresses (gamma/UV irradiation, drought; methyl jasmonate or elicitor treatment, interaction with pathogenic fungi, oomycetes, viruses, *Spodoptera* caterpillars and aphids) than to symbiotic interactions. It should be borne in mind that part of ESTs from libraries representing *Mt* tissues infected by an eukaryotic organism (contributing to the pool of polyA+ transcripts; e.g. a mycorrhizal fungus or a nematode) may not correspond to *Mt* genes.

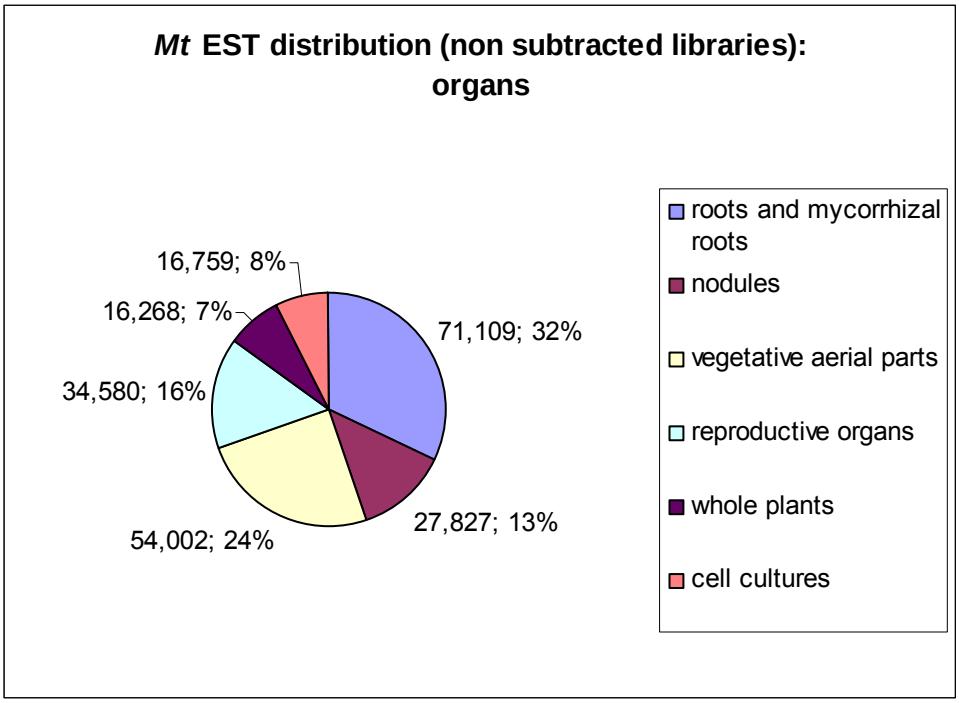

Figure 3: Number of *M. truncatula* ESTs generated from various organs (as of March 2006); a very similar distribution is obtained when considering all *Mt* ESTs (from subtracted and non subtracted libraries).

New directions for cDNA libraries include efforts to generate full-length cDNA clones, which are extremely valuable both for gene identification and annotation and for functional studies of the corresponding proteins. Thus, the TIGR, using the services of Evrogen (www.evrogen.com), recently made such a cDNA library from pooled *Medicago* tissues (MTY library, table 1: flowers, stems, early and late seeds) (C. Town, personal communication). Evrogen combines the full-length Smart technique to produce full-length cDNAs (Zhu et al., 2001) with a proprietary normalization strategy using a novel duplex-specific nuclease (Shagin et al., 2002). The library contains ~5x10^6 clones, from which a total of 5,280 5' sequencing attempts produced 4,420 good reads. These were assembled into 3,619 unique TCs or singletons, of which 1,916 appeared to be full-length. Most importantly, among

these 1,916 full-length clones, 774 sequences were unique when compared to the 9,848 potential full-length sequences from the *Medicago* Gene Index.

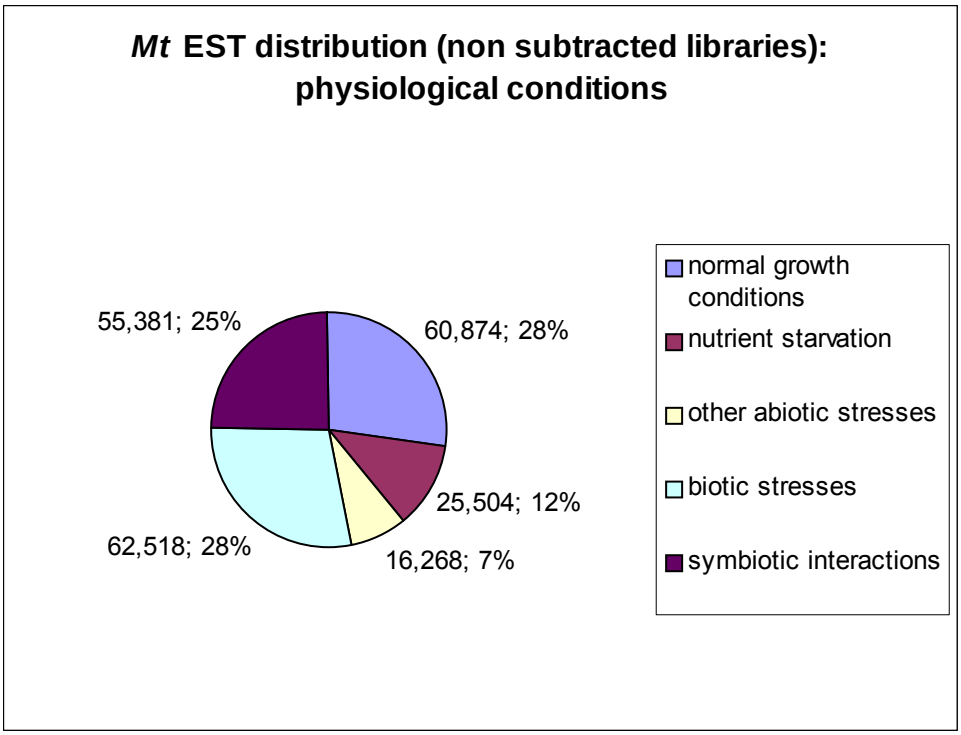

Figure 4: Number of *M. truncatula* ESTs generated from various physiological conditions (as of March 2006).

Another full length cDNA project launched at LIPM, Toulouse (F. de Carvalho-Niebel, personal communication) has recently been funded by the Genoscope (Evry, France) and will lead to the sequencing of 10,000 full-length ESTs generated from isolated, Nod factor-treated, *Medicago* root hairs, using a recently described improved method for fracturing frozen root hairs (Sauviac et al., 2005). In addition to full-length cDNA information this project should yield valuable information about a particular cell type, the root hair that plays a key role during the initial steps of the rhizobium-legume symbiotic interaction.

It can be expected that future new *Mt* libraries might be dedicated to other specific cell types or tissues of interest in order to increase the representation levels of specific transcripts, possibly highly diluted in more complex samples. Such a tissue specific library was generated from alfalfa trichomes, which represent a major site of secondary metabolites synthesis, notably for protection against insect predation (Aziz et al., 2005).

1.1.2. EST editing and analysis

The first steps in EST data analysis consist in clustering ESTs then predicting the sequence of gene transcripts from assembled individual ESTs. Several databases propose such sequences assembled from *Mt* ESTs (and in some cases ETs), after quality assessment and chimera detection. Since different procedures are used for each database, different outcomes can be obtained. Thus the MENS database (Laboratory of Plant Microbe Interactions CNRS-

INRA, Toulouse, http://medicago.toulouse.inra.fr/MENS; Journet et al., 2002) differs from others by the fact that ESTs determined from the 5' and 3' ends of a same cDNA clone were deliberately grouped into the same EST core cluster, even when not overlapping. It is noteworthy that a similar number of clusters was predicted in the three major *Mt* EST databases: for example, starting from approximately 190,000 ESTs, 17,610 clusters (designated TCs, for Tentative Consensus) and 19,341 singletons (unique ESTs) were predicted by the TIGR *Mt* Gene Index (MtGI) Release 7.0 (The Institute for Genomic Research, Rockville, Maryland; www.tigr.org/tdb/tgi.shtml), vs. 17,405 contigs / 17,156 singletons by MtDB 2.0 (Center for Computational Genomics and Bioinformatics, Univ. Minnesota http://www.medicago.org/MtDB2/ (Lamblin et al., 2003)) and 19,552 clusters / 17,468 singletons by MENS. Importantly, a correspondence between clusters found in these three databases is available in MtDB 2.0 (also accessible via MENS). 36,878 unique sequences (still with about ~50% TCs and singletons) were predicted by the TIGR Release 8.0, based on 226,923 ESTs and 317 ETs.

A further step, sequence annotation, has been handled in different ways by these databases. The TIGR MtGI indicates the five best hits found in a non-redundant protein database, and uses the best scoring one to provide a tentative annotation. In addition, MtGI provides identified Gene ontology (GO) assignments as well as tentative orthologues in other species. MENS gives access to detailed results of blastn/x searches against other MtEST clusters (useful to identify gene and protein families) and against DNA and protein public databases, with links to publications. Automatic annotation in MENS is based upon the prediction of encoded proteins by the FrameD software adapted to ESTs (Schiex et al., 2003), followed by InterPro scan [Integrated Resource of Protein Domains and Functional Sites, which integrates the PROSITE, Pfam, ProDom and SMART databases (Apweiler et al., 2001)]. Graphical representations are provided for various pre-calculated searches (protein domains, etc) as well as manual annotation for more than 6,000 clusters, and an interface for such an annotation.

The fact that available *Mt* ESTs originate from a variety of libraries and physiological conditions makes it attractive to infer patterns of gene expression from EST counts. Electronic northerns allowing the visualization of EST distribution in various cDNA libraries are provided by TIGR MtGI and MENS (which also gives a tool to compare several clusters, for example from the same gene family; Figure 5).

Comparison of libraries or of groups of libraries makes it possible to identify lists of genes showing differential EST abundance, which are thus good candidates for being differentially expressed ("*in silico* screening"). Thus MtDB2.0 gives access to a variety of queries, based on EST presence or absence (e.g. only library "A"; library "A" and/or library "B"; library "A" but not library "B"), with options for taxonomy filters (e.g. identification of *Mt* genes specifically expressed in one condition, and with homologies with sequences from defined species) (Lamblin et al., 2003). MtGI and MENS propose tools to identify differentially expressed genes even in situations where ESTs are found in two libraries, owing to a statistical index (likelihood ratio R), that allow evaluation of the rate of false positives due to random fluctuations in cDNA sampling (Stekel et al., 2000; Journet et al., 2002). Such *in silico* studies have been used to identify hundreds of genes predicted to be exclusively or preferentially expressed in root nodules (notably a large family of cystein cluster proteins; Fedorova et al., 2002; Mergaert et al., 2003), or in mycorrhizae (Frenzel et al., 2005).

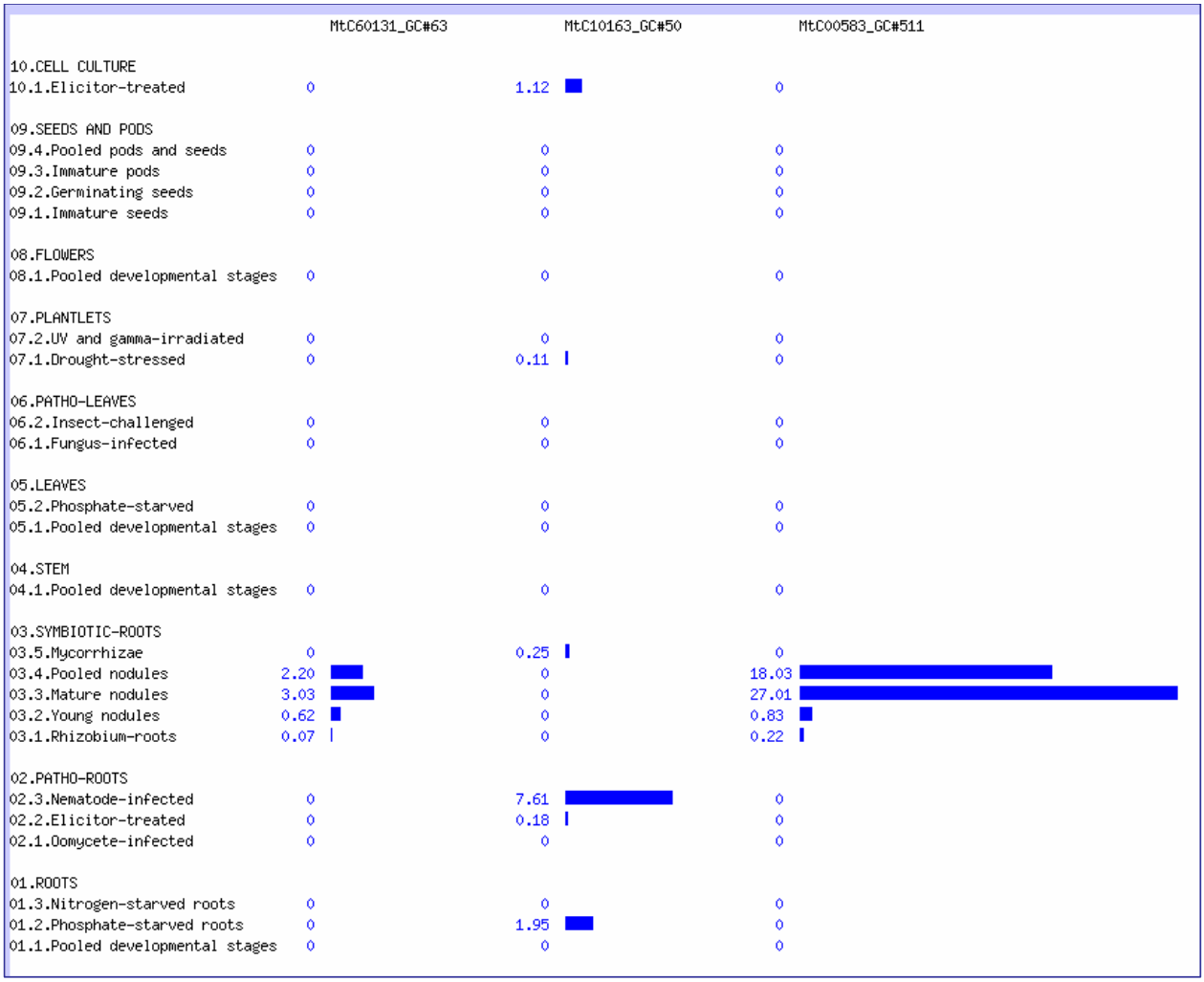

Figure 5: Example of an electronic northern; obtained using a tool provided in the MENS database, with three members of the leghemoglobin family, two of them (MtC60131, MtC00583) being symbiosis-specific, in contrast to the third one (MtC10163). cDNA libraries are pooled in line with broad physiological conditions (left side). The EST frequency observed in each pool, after normalisation with respect to the library size, is presented on the right as a figure (number of ESTs per 1000 sequenced ESTs) and a histogram. An option giving access to EST frequency in each individual library is also available (not shown).

1.2. Subtracted cDNA libraries

Random sampling and sequencing of cDNA clones gives preferential access to strongly expressed genes, whereas weakly expressed genes (or genes expressed in a few cells in a given sample) can be very difficult to identify. Moreover, while it is straightforward to identify novel sequences at the beginning of EST sequencing programs, this becomes more and more difficult when ESTs accumulate for a species, with a correlative increase in costs. For this reason, various methods were developed (i) to normalize libraries in order to have a better representation of low abundance cDNAs and (ii) to enrich libraries in sequences differentially represented between two samples and thereby have a more direct access to genes of interest. The principle of the latter method is to subtract, by hybridization, sequences that are shared between a control (or “driver”) sample and the sample of interest (or “tester”). The first subtracted *Mt* cDNA library was generated by Cook et al. (1995) to enrich for genes

induced during the early phases of the *S. meliloti-Mt* symbiosis. A few years later, a novel technique called Suppression Subtractive Hybridization (SSH) was published, which interestingly combines normalization and subtraction in a single procedure (Diatchenko et al., 1999).

The SSH approach was chosen by several *Mt* groups to produce cDNA libraries (15 in total) complementary to existing ones, with the goal to identify novel markers of symbiotic and pathogenic interactions, abiotic stress responses or seed germination (see tables 4-5). Figure 1 clearly shows that most *Mt* SSH libraries were generated after the peak of standard cDNA libraries. It can also be seen from Tables 4-5 that the number of corresponding ESTs is, not surprisingly, more limited. However a significant proportion of these ESTs actually correspond to novel sequences (as can be verified from MtGI and MENS databases), thereby confirming the complementarity of the SSH approach with regard to standard libraries.

The SSH technique allows two populations of cDNAs to be compared and the differentially expressed ones to be preferentially PCR-amplified and cloned. The subtraction can be done in two reciprocal ways ("forward" and "reverse" in tables 4-5), in order to enrich in sequences either up- or down-regulated between two conditions, although in cases published so far only forward libraries have been characterized. There are now several examples demonstrating that SSH libraries can be a good way to identify differentially expressed genes in *Mt* (Doll et al., 2003; Brechenmacher et al., 2004; Bouton et al., 2005; Frenzel et al., 2005). The enrichment can be quite efficient, with in best cases up to 70% of the clones corresponding to up-regulated genes, but this can vary dramatically depending on the library (Niebel et al., submitted). In any case, independent procedures, such as reverse Northern or macro-array analyses, are required to validate the genes as differentially expressed.

One disadvantage of SSH libraries is that cDNA inserts are generally shorter than in non subtracted libraries. Indeed, several non-overlapping cDNA fragments can be generated and cloned from a single transcript, due to restriction enzyme hydrolysis in the cloning procedure. Consequently, protein prediction and sequence annotation can be difficult or impossible in some cases, unless complementary approaches to obtain full-length clones, such as RACE PCR or PCR cloning from batch cDNA libraries, are carried out. It should also be borne in mind that direct *in silico* comparisons of transcript abundance in SSH libraries is not meaningful, considering that these libraries are highly biased in contrast to standard cDNA libraries.

1.3. cDNA library exploitation

Mt cDNA libraries are of major interest to access the repertoire of expressed genes, using either experimental hybridization screens or computational searches against EST databases (Silverstein et al., 2006). One of the first uses is to mine them for genes of interest, defined from ortho- or paralogues in other organisms or from their EST distribution profile in various cDNA libraries (see above). There are now many examples of *Mt* EST data mining approaches, from simple Blast searches to look for gene homologues (e.g. Liu et al., 2003), to functional motifs searches (e.g. Harrison et al., 2002) or more sophisticated cross-species analyses. These searches allowed putative legume-specific genes to be identified (Graham et al., 2004; Silverstein et al., 2006), or evolutionary events to be analysed in a set of major crop species (Schlueter et al., 2004).

Another obvious application of EST sequencing is the production of macro- and micro-arrays devised for transcript profiling, using sets of appropriate EST cluster representatives (so-called "unigene sets"). Thus, various *Mt* macro- and micro-arrays have been produced and used in the past years. Macro-arrays made from 2,268 clones were used to identify genes accompanying arbuscular mycorrhiza symbiosis (Liu et al., 2003). Macro- and micro-arrays related about 6,000 genes generated exclusively from root, nodule and mycorrhiza cDNAs then allowed the identification of hundreds of genes either up- or down-regulated during nodulation (El Yahyaoui et al., 2004) or in the development of mycorrhizae (Manthey et al., 2004). This micro-array was then extended to a 8k version, incorporating additional flower and pod Mt cDNAs (Firnhaber et al., 2005). Other 6k cDNA micro-arrays representing a broader diversity of libraries were used to analyse earlier nodulation events (Lohar et al., 2006). Macro-arrays focused on a defined biological question were also elaborated: for example macro-arrays were generated based on *in silico* identified *S. meliloti*- and nodule-induced genes as well as cDNAs representing carbon and nitrogen metabolic pathways (Tesfaye et al., 2006).

A new generation of microarrays based on a collection of about 16,000 70-mer oligonucleotides representing MtGi TCs (and in addition in the so-called 16K+ EU version, a few hundreds putative transcription factor genes represented by singletons, identified by functional motifs searches) is now being used in different laboratories worldwide to explore various biological questions (symbiotic or pathogenic interactions, seed formation...) (e.g. Hohnjec et al., 2005).

The Affymetrix technology has also been used, first with DNA chips representing about 10k *Mt* expressed genes used for gene profiling in *S. meliloti*-inoculated roots and in root nodules (Mitra et al., 2004). These Affymetrix chips have in addition allowed the cloning of the *DMI3* symbiotic gene by a method based on transcript abundance in certain mutants (Mitra et al., 2004) and were helpful for the cloning of the *NSP2* symbiotic transcriptional regulator (Kalo et al., 2005). Recently, more comprehensive Affymetrix chips have been set up, with 61,200 probe sets representing *Mt*, *M. sativa* and *S. meliloti* sequences (www.affymetrix.com/products/arrays/specific/medicago.affx).

Finally, *Mt* ESTs and full length cDNAs are obviously very useful for gene annotation and key for *Mt* protein repertoire prediction, which is essential for proteomics approaches (e.g. Mathesius et al., 2001; Watson et al., 2003; Valot et al., 2004; Imin et al., 2005).

2. *M. truncatula* genomic libraries (F. Debellé)

2.1. Genomic libraries production

The first *Mt* genomic libraries were constructed in phage lambda derivatives in the 1990s. They were developed to clone genes of interest together with their regulatory sequences and were generally screened by hybridisation using cDNAs as probes (see for example (Gallusci et al., 1991; Crespi et al., 1994; Burleigh and Harrison, 1998). However, due in particular to their small insert size (< 40 kb) these libraries were not well adapted to positional cloning programs which involve steps of chromosome walking, nor to more extensive genomic programs such as physical mapping. Therefore bacterial artificial chromosome (BAC) vectors which routinely allow the cloning of 100-200 kb inserts were chosen for the construction of the more recent genomic libraries. Yeast artificial chromosome (YAC) libraries were not developed for *Mt* because of their high levels of chimerism and more difficult handling.

The first *Mt* BAC libraries were constructed by D. Cook's group (UC Davis) using the pBeloBAC11 vector (Shizuya et al., 1992) and *Mt* A17 DNA partially digested by HindIII (table 6). The first library, called mth1 (Nam et al., 1999) contained 30,000 clones of 100 kb average size representing approximately 5 haploid genome equivalents. It was later complemented by the larger mth2 library (100K clones of 120 kb average size, 20x coverage) (Choi et al., 2004). These libraries were instrumental in the success of several positional cloning programs leading to the identification of five genes controlling symbioses with rhizobia and/or mycorrhizal fungi (for a review see Stacey et al., 2006). However, in a few instances chromosome walking was blocked by gaps in the HindIII libraries. In order to overcome the cloning bias linked to the use of a single restriction enzyme, other BAC libraries were constructed by groups in France and Korea, using partial digests of *Mt* DNA by the EcoRI (Toulouse-Evry, France, mte1 library; (Chalhoub et al., 2004; Levy et al., 2004) and BamHI (Sogang University, Korea, mtb1 library) restriction enzymes. In addition, a BAC library with larger inserts (up to more than 200 kb) was constructed using HindIII partial digestion (Toulouse-Evry mth4 library). Construction of one additional BAC library (Sau3A fragments) and of a fosmid library (bacterially-propagated phagemid vector system, suitable for cloning genomic inserts of about 40 kb) are in progress (Genoscope, France and Sanger Institute, UK, respectively).

2.2. Genomic libraries exploitation

All *Mt* BAC libraries are currently extensively used for the sequencing of the gene-rich part of the *Mt* genome. Indeed, a strategy of BAC / BAC sequencing, based on a robust physical map anchored to the genetic map, has been chosen by the International Consortium in charge of this program. Thus, 44,000 clones (11x) of the mth2 library have been fingerprinted by HindIII digestion and agarose gel electrophoresis, yielding a physical map with 1370 contigs of 340 kb average length ((Mun et al., 2006) http://www.medicago.org/genome). In parallel, 800 gene-rich (EST containing) seed BACs were fully sequenced (U Oklahoma ; http://www.genome.ou.edu/medicago.html) to provide seed points from which to continue the sequencing effort. This is being done mostly using the STC (sequence tagged connector) strategy: BES (BAC end sequences) have been determined for 55,000 clones of the mth2, 30,000 clones of mte1, and 10,000 clones of the mth4 libraries by the Genoscope (France), TIGR (USA), Sanger (UK), Szeged (Hungary) sequencing centers ; http://www.medicago.org/genome). Alignment of BES to seed BAC sequences, together with fingerprinting data, allow the definition of a minimum tiling path and the selection of the BACs to be sequenced. BAC end sequences (and future fosmid end sequences) then help validating the pseudo-chromosome sequence assembly.

The goal is to sequence 300 Mb (out of 500 Mb of total genomic DNA) by 2007, which should include ~95 % of *Mt* genes. Most genes and flanking sequences will then be easily cloned by PCR amplification, using this sequence data. However BAC libraries will remain a useful resource to subclone genomic regions of interest, and to identify the few genes located outside the planned sequenced regions. This will be made easier by the wealth of resources developed for the physical mapping and sequencing program (BES, FPC map, genetic map). As for experimental screening, although pools of BACs were developed (Nam et al., 1999; Choi et al., 2004) by the UC Davis group for PCR screening, the preferred method for the average user will certainly be screening by hybridisation since high density filters are now easily available from stock centers for most BAC libraries (see below).

3. The CNRGV, an example of international distribution center for cDNA and BAC samples (H. Bergès).

As seen in detail in previous chapters, successive *Mt* genomics projects have generated an abundance of tools for molecular biology (notably cDNA and BAC clones) that should ideally be accessible to the whole *Medicago* community. In the post-genomic era, experiments can be designed to study many genes simultaneously, notably to make progress in functional annotation of the genome, a major challenge for coming years (i.e. formulation and testing of hypotheses that propose function(s) for particular genomic sequences). There is a need for International Biological Resource Centers able to collect, maintain and distribute clones and related information, and to develop activities and services accompanying these evolutions in scientific objectives and experimental approaches.

For example, in France, the CNRGV (Centre National de Ressources en Génomique Végétale; http://toulouse.inra.fr/cnrgv/) has proposed to store and distribute to the international scientific community all *Mt* genomic collections created by various laboratories. It already maintains three *Mt* BAC libraries (mth2, 4, mte1) and is about to collect a set of cDNA libraries made in Europe (MtBA, MtBB, MtBC, MtAMP, MtGIM, MtFLOW, MtPOSE...), more being expected in future. The CNRGV was created at the instigation of INRA in 2004 to ensure that the plant genomic resources, created at public or private research organisations, are maintained in an optimal manner and made available to all of the research community. This non-for-profit centre, certified in a Quality Management System (ISO 9001:2000 quality standards), maintains collections representing a range of model and crop plant species and databases containing molecular, physiological and structural information relevant to these collections. The material and services available at the CNRGV include genomic and cDNA libraries clones, high density macroarrays, genomic pools, screening and robotic services.

In the USA, the Clemson University Genomics Institute (CUGI http://www.genome.clemson.edu/) distributes individual or whole collections of *Mt* BAC clones from mth1 and mth2 libraries. It is hoped that an international network of mirror Biological Resource Centres working in collaboration will be set up in the near future for the benefit of plant science and biotechnology.

Acknowledgments

We are very grateful to Martin Crespi, Fernanda de Carvalho-Niebel, Anis Limami and Chris Town for providing us with unpublished results, and to Julie Cullimore for English reviewing.

References

Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, Sigrist CJ, Zdobnov EM (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res **29:** 37-40

Aziz N, Paiva NL, May GD, Dixon RA (2005) Transcriptome analysis of alfalfa glandular trichomes. Planta **221:** 28-38

Bouton S, Viau L, Lelievre E, Limami AM (2005) A gene encoding a protein with a proline-rich domain (MtPPRD1), revealed by suppressive subtractive hybridization (SSH), is specifically expressed in the Medicago truncatula embryo axis during germination. J Exp Bot **56:** 825-832

Brechenmacher L, Weidmann S, van Tuinen D, Chatagnier O, Gianinazzi S, Franken P, Gianinazzi-Pearson V (2004) Expression profiling of up-regulated plant and fungal genes in early and late stages of Medicago truncatula-Glomus mosseae interactions. Mycorrhiza **14:** 253-262

Burleigh SM, Harrison MJ (1998) Characterization of the Mt4 gene from Medicago truncatula. Gene **216:** 47-53

Chalhoub B, Belcram H, Caboche M (2004) Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size. Plant Biotech. Journal **2:** 181-188

Choi HK, Kim D, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR (2004) A sequence-based genetic map of Medicago truncatula and comparison of marker colinearity with M. sativa. Genetics **166:** 1463-1502

Cook D, Dreyer D, Bonnet D, Howell M, Nony E, VandenBosch K (1995) Transient induction of a peroxidase gene in Medicago truncatula precedes infection by Rhizobium meliloti. Plant Cell **7:** 43-55

Covitz PA, Smith LS, Long SR (1998) Expressed sequence tags from a root-hair-enriched medicago truncatula cDNA library. Plant Physiol **117:** 1325-1332

Crespi MD, Jurkevitch E, Poiret M, d'Aubenton-Carafa Y, Petrovics G, Kondorosi E, Kondorosi A (1994) enod40, a gene expressed during nodule organogenesis, codes for a non-translatable RNA involved in plant growth. Embo J **13:** 5099-5112

Diatchenko L, Lukyanov S, Lau YF, Siebert PD (1999) Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. Methods Enzymol **303:** 349-380

Doll J, Hause B, Demchenko K, Pawlowski K, Krajinski F (2003) A member of the germin-like protein family is a highly conserved mycorrhiza-specific induced gene. Plant Cell Physiol **44:** 1208-1214

El Yahyaoui F, Kuster H, Ben Amor B, Hohnjec N, Puhler A, Becker A, Gouzy J, Vernie T, Gough C, Niebel A, Godiard L, Gamas P (2004) Expression profiling in Medicago truncatula identifies more than 750 genes differentially expressed during nodulation, including many potential regulators of the symbiotic program. Plant Physiol **136:** 3159-3176

Fedorova M, van de Mortel J, Matsumoto PA, Cho J, Town CD, VandenBosch KA, Gantt JS, Vance CP (2002) Genome-wide identification of nodule-specific transcripts in the model legume Medicago truncatula. Plant Physiol **130:** 519-537

Firnhaber C, Puhler A, Kuster H (2005) EST sequencing and time course microarray hybridizations identify more than 700 Medicago truncatula genes with developmental expression regulation in flowers and pods. Planta **222:** 269-283

Frenzel A, Manthey K, Perlick AM, Meyer F, Puhler A, Kuster H, Krajinski F (2005) Combined transcriptome profiling reveals a novel family of arbuscular mycorrhizal-specific Medicago truncatula lectin genes. Mol Plant Microbe Interact **18:** 771-782

Frijters ACJ, Zhang Z, van Damma M, Wang G-L, Ronald PC, Michelmore RW (1997) Construction of a bacterial artificial chromosome library containing large EcoRI and HindIII genomic fragments of lettuce. Theor Appl Genet **94:** 90-99

Frugoli J, Harris J (2001) Medicago truncatula on the move! Plant Cell **13:** 458-463

Gallusci P, Dedieu A, Journet EP, Huguet T, Barker DG (1991) Synchronous expression of leghaemoglobin genes in Medicago truncatula during nitrogen-fixing root nodule development and response to exogenously supplied nitrate. Plant Mol Biol **17:** 335-349

Gamas P, de Carvalho-Niebel F, Lescure N, Cullimore J (1996) Use of a subtractive hybridization approach to identify new Medicago truncatula genes induced during root nodule development. Mol Plant Microbe Interact **9:** 233-242

Gomez SK, Cox MM, Bede JC, Inoue K, Alborn HT, Tumlinson JH, Korth KL (2005) Lepidopteran herbivory and oral factors induce transcripts encoding novel terpene synthases in Medicago truncatula. Arch Insect Biochem Physiol **58:** 114-127

Graham MA, Silverstein KA, Cannon SB, VandenBosch KA (2004) Computational identification and characterization of novel genes from legumes. Plant Physiol **135:** 1179-1197

Gyorgyey J, Vaubert D, Jimenez-Zurdo JI, Charon C, Troussard L, Kondorosi A, Kondorosi E (2000) Analysis of Medicago truncatula nodule expressed sequence tags. Mol Plant Microbe Interact **13:** 62-71

Harrison MJ, Dewbre GR, Liu J (2002) A phosphate transporter from Medicago truncatula involved in the acquisition of phosphate released by arbuscular mycorrhizal fungi. Plant Cell **14:** 2413-2429

Hohnjec N, Vieweg MF, Pühler A, Becker A, Küster H (2005) Overlaps in the Transcriptional Profiles of *Medicago truncatula* Roots Inoculated with Two Different Glomus Fungi Provide Insights into the Genetic Program Activated during Arbuscular Mycorrhiza. Plant Physiol **137:** 1283-1301

Imin N, Nizamidin M, Daniher D, Nolan KE, Rose RJ, Rolfe BG (2005) Proteomic analysis of somatic embryogenesis in Medicago truncatula. Explant cultures grown under 6-benzylaminopurine and 1-naphthaleneacetic acid treatments. Plant Physiol **137:** 1250-1260

Journet EP, van Tuinen D, Gouzy J, Crespeau H, Carreau V, Farmer MJ, Niebel A, Schiex T, Jaillon O, Chatagnier O, Godiard L, Micheli F, Kahn D, Gianinazzi-Pearson V, Gamas P (2002) Exploring root symbiotic programs in the model legume Medicago truncatula using EST analysis. Nucleic Acids Res **30:** 5579-5592

Kalo P, Gleason C, Edwards A, Marsh J, Mitra RM, Hirsch S, Jakab J, Sims S, Long SR, Rogers J, Kiss GB, Downie JA, Oldroyd GE (2005) Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. Science **308:** 1786-1789

Lamblin AF, Crow JA, Johnson JE, Silverstein KA, Kunau TM, Kilian A, Benz D, Stromvik M, Endre G, VandenBosch KA, Cook DR, Young ND, Retzel EF (2003) MtDB: a database for personalized data mining of the model legume Medicago truncatula transcriptome. Nucleic Acids Res **31:** 196-201

Levy J, Bres C, Geurts R, Chalhoub B, Kulikova O, Duc G, Journet EP, Ane JM, Lauber E, Bisseling T, Denarie J, Rosenberg C, Debelle F (2004) A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. Science **303:** 1361-1364

Liu CJ, Huhman D, Sumner LW, Dixon RA (2003) Regiospecific hydroxylation of isoflavones by cytochrome p450 81E enzymes from Medicago truncatula. Plant J **36:** 471-484

Liu H, Trieu AT, Blaylock LA, Harrison MJ (1998) Cloning and characterization of two phosphate transporters from Medicago truncatula roots: regulation in response to phosphate and to colonization by arbuscular mycorrhizal (AM) fungi. Mol Plant Microbe Interact **11:** 14-22

Liu J, Blaylock LA, Endre G, Cho J, Town CD, VandenBosch KA, Harrison MJ (2003) Transcript profiling coupled with spatial expression analyses reveals genes involved in distinct developmental stages of an arbuscular mycorrhizal symbiosis. Plant Cell **15:** 2106-2123

Lohar DP, Sharopova N, Endre G, Penuela S, Samac D, Town C, Silverstein KA, VandenBosch KA (2006) Transcript analysis of early nodulation events in Medicago truncatula. Plant Physiol **140:** 221-234

Manthey K, Krajinski F, Hohnjec N, Firnhaber C, Puhler A, Perlick AM, Kuster H (2004) Transcriptome profiling in root nodules and arbuscular mycorrhiza identifies a collection of novel genes induced during Medicago truncatula root endosymbioses. Mol Plant Microbe Interact **17:** 1063-1077

Mathesius U, Keijzers G, Natera SH, Weinman JJ, Djordjevic MA, Rolfe BG (2001) Establishment of a root proteome reference map for the model legume Medicago truncatula using the expressed sequence tag database for peptide mass fingerprinting. Proteomics **1:** 1424-1440

Mergaert P, Nikovics K, Kelemen Z, Maunoury N, Vaubert D, Kondorosi A, Kondorosi E (2003) A novel family in Medicago truncatula consisting of more than 300 nodule-

specific genes coding for small, secreted polypeptides with conserved cysteine motifs. Plant Physiol **132:** 161-173

Mitra RM, Gleason CA, Edwards A, Hadfield J, Downie JA, Oldroyd GE, Long SR (2004) A Ca2+/calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. Proc Natl Acad Sci U S A **101:** 4701-4705

Mitra RM, Shaw SL, Long SR (2004) Six nonnodulating plant mutants defective for Nod factor-induced transcriptional changes associated with the legume-rhizobia symbiosis. Proc Natl Acad Sci U S A **101:** 10217-10222

Mun JH, Kim DJ, Choi HK, Gish J, Debelle F, Mudge J, Denny R, Endre G, Saurat O, Dudez AM, Kiss GB, Roe B, Young ND, Cook DR (2006) Distribution of Microsatellites in the Genome of Medicago truncatula: A Resource of Genetic Markers That Integrate Genetic and Physical Maps. Genetics **172:** 2541-2555

Nam YW, Penmetsa RV, Endre G, Uribe P, Kim D, Cook DR (1999) Construction of a bacterial artificial chromosome library of Medicago truncatula and identification of clones containing ethylene-response genes. Theor Appl Genet **3:** 638-646

Nyamsuren O, Colditz F, Rosendahl S, Tamasloukht M, Bekel T, Meyer F, Kuester H, Franken P, Krajinski F (2003) Transcriptional profiling of Medicago truncatula roots after infection with Aphanomyces euteiches (oomycota) identifies novel genes upregulated during pathogenic interactions. Physiol. Mol.Plant Pathol. **63:** 17-26

Sanchez L, Weidmann S, Arnould C, Bernard AR, Gianinazzi S, Gianinazzi-Pearson V (2005) Pseudomonas fluorescens and Glomus mosseae trigger DMI3-dependent activation of genes related to a signal transduction pathway in roots of Medicago truncatula. Plant Physiol **139:** 1065-1077

Sauviac L, Niebel A, Boisson-Dernier A, Barker DG, de Carvalho-Niebel F (2005) Transcript enrichment of Nod factor-elicited early nodulin genes in purified root hair fractions of the model legume Medicago truncatula. J Exp Bot **56:** 2507-2513

Schiex T, Gouzy J, Moisan A, de Oliveira Y (2003) FrameD: A flexible program for quality check and gene prediction in prokaryotic genomes and noisy matured eukaryotic sequences. Nucleic Acids Res **31:** 3738-3741

Schlueter JA, Dixon P, Granger C, Grant D, Clark L, Doyle JJ, Shoemaker RC (2004) Mining EST databases to resolve evolutionary events in major crop species. Genome **47:** 868-876

Schnabel E, Journet EP, de Carvalho-Niebel F, Duc G, Frugoli J (2005) The Medicago truncatula SUNN gene encodes a CLV1-like leucine-rich repeat receptor kinase that regulates nodule number and root length. Plant Mol Biol. **58**:809-822.

Shagin DA, Rebrikov DV, Kozhemyako VB, Altshuler IM, Shcheglov AS, Zhulidov PA, Bogdanova EA, Staroverov DB, Rasskazov VA, Lukyanov S (2002) A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. Genome Res **12:** 1935-1942

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proc Natl Acad Sci U S A **89:** 8794-8797

Silverstein KA, Graham MA, VandenBosch KA (2006) Novel paralogous gene families with potential function in legume nodules and seeds. Current Opinion in Plant Biology **9:** 142-146

Stacey G, Libault M, Brechenmacher L, Wan J, May GD (2006) Genetics and functional genomics of legume nodulation. Curr Opin Plant Biol. **9**:110-21.

Stekel DJ, Git Y, Falciani F (2000) The comparison of gene expression from multiple cDNA libraries. Genome Res **10:** 2055-2061

Tesfaye M, Samac D, Vance CP (2006) Insights into symbiotic nitrogen fixation in *Medicago truncatula*. Mol Plant Microbe Interact **19:** 330-341

Valot B, Gianinazzi S, Eliane DG (2004) Sub-cellular proteomic analysis of a Medicago truncatula root microsomal fraction. Phytochemistry **65:** 1721-1732

Wang K, Boysen C, Shizuya H, Simon MI, Hood L (1997) Complete nucleotide sequence of two generations of a bacterial artificial chromosome cloning vector. Biotechniques **23:** 992-994

Watson BS, Asirvatham VS, Wang L, Sumner LW (2003) Mapping the proteome of barrel medic (Medicago truncatula). Plant Physiol **131:** 1104-1123

Weidmann S, Sanchez L, Descombin J, Chatagnier O, Gianinazzi S, Gianinazzi-Pearson V (2004) Fungal elicitation of signal transduction-related plant genes precedes mycorrhiza establishment and requires the dmi3 gene in Medicago truncatula. Mol Plant Microbe Interact **17:** 1385-1393

Wilson RC, Cooper JB (1994) Characterization of PRP1 and PRP2 from Medicago truncatula. Plant Physiol **105:** 445-446

Wulf A, Manthey K, Doll J, Perlick AM, Linke B, Bekel T, Meyer F, Franken P, Kuster H, Krajinski F (2003) Transcriptional changes in response to arbuscular mycorrhiza development in the model plant Medicago truncatula. Mol Plant Microbe Interact **16:** 306-314

Zhu Y, Machleder E, Chenchik A, Li R, Siebert P (2001) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques **30:** 892-897

Physiological condition	Organ	Library name	Cloning vector	number of ESTs	Submitter	Institution	Reference
Normal growth conditions							
	1. root hairs and root tips	MtRHE	pBK-CMV	899	Long, SR	Stanford Univ., USA	Covitz et al., 1998
	2. roots	Developing root	pBluescript SK(-)	3,054	Paiva, NL	The Noble Foundation, USA	unpublished
	3. leaves and cotyledons	DSLC	pBluescript SK(-)	2,143	Samac, D	Univ. of Minnesota, USA	unpublished
	4. young to senescing leaves	Developing leaf	pBluescript SK(-)	9,415	May, GD	The Noble Foundation, USA	unpublished
	5. internodal stem segments	Developing stem	pBluescript SK(-)	10,783	Dixon, RA	The Noble Foundation, USA	unpublished
	6. flowers	MtFLOW	pGEM-T	1,196	Küster, H	Univ. Bielefeld, Germany	Firnhaber et al., 2005
	7. flowers	Developing flower	pBluescript SK(-)	6,724	May, GD	The Noble Foundation, USA	unpublished
	8. immature pod walls	GPOD	pBluescript SK(-)	3,621	Grusak, MA	USDA/ARS, USA	unpublished
	9. pods including seeds	MtPOSE	pGEM-T	1,162	Küster, H	Univ. Bielefeld, Germany	Firnhaber et al., 2005
	10. immature seeds 11-19 dpp	GESD	pBluescript SK(-)	4,521	Grusak, MA	USDA/ARS, USA	unpublished
	11. immature seeds 25-35 dpp	GLSD	pBluescript SK(-)	4,866	Grusak, MA	USDA/ARS, USA	unpublished
	12. germinating seeds	germinating seed	pBluescript SK(-)	8,070	May, GD	The Noble Foundation, USA	unpublished
	13. stems, flowers, early & late seeds	MTY	pDNR-LIB	4,420	Cheung, F	TIGR, USA	unpublished
Nutrient starvation							
	1. N-starved root tips	MtBA	pBluescript SK(-)	7,939	Journet, EP & Gamas, P	INRA-CNRS, France	Journet et al., 2002
	2. N-starved roots	KV0	pBluescript SK(-)	2,752	VandenBosch, K	Univ. of Minnesota, USA	unpublished
	3. P-starved roots	MHRP-	pBluescript SK(-)	2,658	Harrison, MJ	The Noble Foundation, USA	unpublished
	4. P-starved roots	Rootphos-	pBluescript SK(-)	1,967	Harrison, MJ	The Noble Foundation, USA	unpublished
	5. P-starved leaves	Phosphate starved	pBluescript SK(-)	10,188	Harrison, MJ	The Noble Foundation, USA	unpublished

		leaf					
Other abiotic stresses							
	1. gamma or UV-irradiated seedlings	Irradiated	pBluescript SK(-)	6,748	May, GD	The Noble Foundation, USA	unpublished
	2. drought stressed plantlets	Drought	pBluescript SK(-)	9,520	May, GD	The Noble Foundation, USA	unpublished

N stands for nitrogen, P for phosphate, dpp for days post pollination

Table 1: *M. truncatula* standard (non subtracted) cDNA libraries

Physiological condition	Organ	Library name	Cloning vector	number of ESTs	Submitter	Institution	Reference
Endosymbiotic interactions							
with rhizobium	1. *S.meliloti* infected roots 1 dpi	KV1	pBluescript SK(-)	2,840	VandenBosch, K	Univ. of Minnesota, USA	unpublished
	2. *S.meliloti* infected roots 2 dpi	KV2	pBluescript SK(-)	3,330	VandenBosch, K	Univ. of Minnesota, USA	unpublished
	3. *S.meliloti* infected roots 3 dpi	KV3	pBluescript SK(-)	5,918	VandenBosch, K	Univ. of Minnesota, USA	unpublished
	4. young nodules	MtBB	pBluescript SK(-)	7,807	Journet, EP & Gamas, P	INRA-CNRS, France	Journet et al., 2002
	5. young R108 nodules	R108MtNo	pBluescript SK(-)	438	Mergaert, P	CNRS, France	Gyorgyey et al. 2000
	6. young R108 nodules	NOLLY	pBluescript SK(-)	3,074	Mergaert, P	CNRS, France	unpublished
	7. nodulated roots 4 wpi	Nodulated root	pBluescript SK(-)	3,299	Paiva, NL	The Noble Foundation, USA	unpublished
	8. nitrogen-fixing nodules	GVN	pBluescript SK(-)	6,468	Vance, CP	Univ. of Minnesota, USA	unpublished
	9. senescent nodules	GVSN	pBluescript SK(-)	2,788	Vance, CP	Univ. of Minnesota, USA	unpublished
with arbuscular fungi	1. *G. intraradices* arbuscular mycorrhiza 3 wpi	MtBC	pBluescript SK(-)	8,601	Van Tuinen, D	INRA-CNRS, France	Journet et al., 2002
	2. *G. intraradices* arbuscular mycorrhiza 6 wpi	MtAMP	pGEM-T	3,450	Küster, H	Univ. Bielefeld, Germany	Frenzel et al., 2005
	3. *G. versiforme* arbuscular mycorrhiza 10 to 28 dpi	MtMHAM	pBluescript SK(-)	7,368	Harrison, M	The Noble Foundation, USA	unpublished

Only libraries with more than 20 ESTs are considered

R108 means that *M. truncatula* R108 genotype was used, instead of Jemalong A17

G. stands for *Glomus*, S. for *Sinorhizobium*, dpi for day(s) post inoculation, wpi for week(s) post inoculation

Table 2: *M. truncatula* standard (non subtracted) cDNA libraries (continued)

Physiological condition	Organ	Library name	Cloning vector	number of ESTs	Submitter	Institution	Reference
Biotic stresses							
insects	1. *Spodoptera*-challenged leaves (local and systemic)	Insect herbivory	pBluescript SK(-)	10,309	Korth, K	Univ. of Arkensas, USA	Gomes et al., 2005
	2. Aphid-infected shoots	Aphid-infected Shoots	pBluescript SK(-)	4,955	May, GD	The Noble Foundation, USA	unpublished
microorganisms	1. *Phoma*-infected leaf	Phoma-infected	pBluescript SK(-)	3,281	Paiva, NL	The Noble Foundation, USA	unpublished
	2. *Phytophtora medicaginis* -infected roots 10 dpi	DSIR	pBluescript SK(-)	2,463	Cote, F	Univ. of Georgia, USA	unpublished
	3. *Colletotrichum trifolii*-infected cotyledons and leaves 5 & 8 dpi	DSIL	pBluescript SK(-)	6,003	Samac, D	Univ. of Minnesota, USA	unpublished
virus	1. alfalfa mosaic virus-infected leaves	Virus-infected leaves	pBluescript SK(-)	7,113	May, GD	The Noble Foundation, USA	unpublished
nematodes	1.*Meloidogyne incognita*-infected roots 3 dpi	BNIR	pBluescript SK(-)	3,154	Bird, DM	North Carolina State Univ., USA	unpublished
elicitors and signals	1. oligogalacturonide-treated roots	HOGA	pBluescript SK(-)	5,794	Hahn, MG	Univ. of Georgia, USA	unpublished
	2. β-glucan elicitor-treated roots	MGHG	pBluescript SK(-)	2,687	Hahn, MG	Univ. of Georgia, USA	unpublished
	3. yeast extract-treated cell culture	Elicited cell culture	pBluescript SK(-)	9,859	Dixon, RA	The Noble Foundation, USA	unpublished
	4. MeJA-elicited cell culture	MeJA-elicited root cell suspension culture	pBluescript SK(-)	6,900	May, GD	The Noble Foundation, USA	unpublished

MeJA stands for for Methyl-Jasmonate, dpi for days post inoculation

Table 3: *M. truncatula* standard (non subtracted) cDNA libraries (continued

Physiological condition	Organ	Library name	Subtraction	Cloning vector	number of ESTs	Submitter	Institution	Reference
Normal development								
germination	1. seed embryo axis		forward			Limami, A	INRA-Angers University, France	unpublished
Nutrient starvation								
nitrogen starvation	2. N-starved root	MtC4	forward	pGEM-T	288	Niebel, A	INRA-CNRS, France	Niebel et al., submitted
Other abiotic stresses								
salt stress	1. root		forward	pGEM-T Easy		Crespi, M.	CNRS, France	Merchan et al., submitted
salt stress recovery	2. root		forward	pGEM-T Easy		Crespi, M.	CNRS, France	id
UV-B irradiation	3. leaf		forward	pGEM-T Easy	932	Park, J-S	NIAB, Korea	
Pathogenic & pest interactions								
oomycete interaction	1. *Aphanomyces euteiches* - infected roots, 6 dpi	MTAPHEU	forward	pGEM-T	505	Krajinski, F	LG Molekular genetik, Germany	Nyamsuren et al., 2003

Table 4: *M. truncatula* SSH (subtracted) cDNA libraries

Physiological condition	Organ	Library name	Subtraction	Cloning vector	number of ESTs	Submitter	Institution	Reference
Symbiotic interactions								
response to Nod Factors	1. NF-treated roots	MtSNF	forward	pGEM-T	726	Niebel, A	INRA-CNRS, France	Niebel et al., submitted
	2. control root for NF treatment	MtSCF	reverse	pGEM-T	287	Niebel, A	id	id
***S. meliloti* infection and nodule formation**	1. w.t. *S.meliloti*-infected TR122 roots 3 & 6 dpi	MtSTW	forward	pGEM-T	453	Godiard, L	id	id
	2. *nodA S.meliloti*-infected TR122 roots 3 & 6 dpi	MtSTA	reverse	pGEM-T	436	Godiard, L	id	id
	3. nodules 4 & 10 dpi	MtSN4	forward	pGEM-T	879	Micheli, F	id	id
	4. N-starved control roots	MtSN0	reverse	pGEM-T	270	Micheli, F	id	id
arbuscular mycorrhiza	1. *G. mossae* mycorrhiza 5 dpi	MtGMES	forward	pGEM-T Easy	29	Weidmann, S	id	Weidmann et al., 2004
	2. *G. mossae* mycorrhiza 3 wpi	Mt mycorrhized roots 3 weeks	forward	pGEM-T Easy	37	Brechenmacher, L	id	Brechenmacher et al., 2004
	3. *G. intraradices* mycorrhiza 3 wpi	MtGIM	forward	pGEM-T Easy	1,775	Krajinski, F	LG Molekular genetik, Germany	Wulf et al., 2003
***Pseudomonas fluorescens* endophytic interaction**	1. *P. fluorescens* - inoculated roots 4-8 dpi		forward	pGEM-T	58	Sanchez, L	INRA-CNRS, France	Sanchez et al., 2005

NF stands for Nod factors, *G.* for *Glomus*, *S*. for *Sinorhizobium*, dpi for days post inoculation, wpi for weeks post inoculation.
MtSNF/MtSC, MtSTW/MtSTA and MtSN4/MtSN0 represent 3 couples of libraries generated by reciprocal subtractions ("forward" and "reverse") of driver and tester cDNAs, enriching for sequences more abundant in the induced or control conditions respectively.
TR122 is a supernodulant Jemalong mutant (Sagan et al., 1995), allelic to *sunn (*Schnabel et al., 2005).

Table 5: *M. truncatula* SSH (subtracted) cDNA libraries (continued)

BAC library	Restriction enzyme	Vector	Vector Reference	average insert size	Number of clones	Resources	Stock Center	Submitter	Institution	Reference
mth1	HindIII	pBeloBAC11	Wang et al., 1997	100 kb	30,000		CUGI	Cook, DR	Texas A&M University / UC Davis, USA	Nam et al, 1999
mth2	HindIII	pBeloBAC11	id	120 kb	100,000	FPC, BES, GM, BS	CUGI, CNRGV	id	id	
mth4	HindIII	pIndigoBAC	Chalhoub et al., 2004	up to 200 kb	73,000	BES,GM, BS	CNRGV	Chalhoub, B Debellé, F	INRA-CNRS, France	
mte1	EcoRI	pIndigoBAC	id	120 kb	33,000	BES,GM, BS	CNRGV	id	id	Lévy et al., 2004
mtb1	BamHI	pECBAC1	Frijters et al., 1997	100 kb	24,000			Nam, YW	Sogang Univ., Korea	

kb stands for kilobase, FPC for fingerprinted contigs, BES for Bac end sequences, GM for genetic markers, BS for complete BAC sequences
CUGI stands for Clemson University Genomics Institute, CNRGV for Centre National de Ressources en Génomique Végétale

Table 6: *M. truncatula* BAC libraries