Research Signpost
37/661 (2), Fort P.O., Trivandrum-695 023, Kerala, India

Molecular Biology of Tropical Plants, 2006: 143-160 ISBN: 81-308-0138-8
Editor: Claudine Franche

7 *Arabidopsis thaliana* as a tool to study the molecular biology of tropical plants

Laplaze, L.[1], Jouannic, S.[1] and Gherbi, H.[2]
[1]UMR 1098 and [2]UMR 5096, Institut de Recherche pour le Développement
911 Avenue Agropolis, F-34394 Montpellier cedex 5, France

Abstract

In recent years the emergence of the model plant Arabidopsis thaliana has completely changed plant molecular biology. Numerous tools and data (including the complete genome sequence) have been developed for this plant. Here, we describe how these resources can be exploited to study the molecular mechanisms of biological processes in tropical plants. We provide some information about Arabidopsis and the molecular biology resources available in this species. In the first part of this paper, we explain how

Correspondence/Reprint request: Dr. Laplaze, L., UMR 1098, Institut de Recherche pour le Développement 911 Avenue Agropolis, F-34394 Montpellier cedex 5, France. E-mail: laplaze@mpl.ird.fr

these resources can be used to identify genes involved in a given biological process, and in a second part, how Arabidopsis can be used to characterize of the function of genes isolated in tropical plants. Finally, we show that Arabidopsis can accelerate map-based cloning of genes in other plants.

Introduction

The last twenty years have seen dramatic changes in plant sciences with the emergence of the model plant *Arabidopsis thaliana*. Although botanists have known this plant for more than four centuries [1], its widespread use as a laboratory organism only started in the 1980s. Adoption of the Arabidopsis system by plant biologists allowed their work to be compared and findings in diverse area to be integrated leading to a very rapid increase in our knowledge of plant biology.

Arabidopsis thaliana (L.) Heynh (2n=10) is a small weed of the *Brassicaceae* family. It is found in many different habitats and has a broad geographical distribution ranging from temperate Europe to high mountains in equatorial Africa. Accordingly, a large collection of wild populations (ecotypes) adapted to different conditions is available. It grows as a small rosette of 2-5 cm in diameter from which a flowering stem of 20-70 cm in height is produced. Flowers are made of four sepals, four petals, six stamens and a single ovary consisting of two fused carpels. Arabidopsis is autogamous and produces several hundred seedpods (siliques) each containing about 50 seeds. In the most common laboratory strains such as Columbia, the entire life cycle can be completed in 2 months.

The emergence of Arabidopsis as a model for plant biology is due to several characteristics. It is small and easy to grow under laboratory conditions. The Arabidopsis life cycle is short (6-8 weeks). It is very fertile (up to 100.000 seeds/plant) and self- or cross-pollination are easy. The Arabidopsis genome is one of the smallest angiosperm genomes with 125 Mbp, roughly 8 and 20 times less than the tomato and maize genome respectively. Moreover, it contains very little dispersed repetitive DNA. The Arabidopsis genome was the first plant genome to be completely sequenced [2] allowing the identification of the complete set of Arabidopsis genes.

As a consequence of its use as a model system, many experimental tools have been developed in Arabidopsis. A very simple and efficient protocol for genetic transformation by *Agrobacterium tumefaciens* is available [3]. The wealth of mutants generated by different methods as well as the many natural populations (ecotypes) provide powerful tools to dissect the genetic bases of plant development and physiology. The availability of insertion alleles (T-DNA, transposons) for the vast majority of genes makes it easy to knock out individual genes in order to test their function. Custom microarrays containing the complete set of Arabidopsis genes are available and can be

used to analyze the effects of physiological or developmental changes in expression in the whole genome. All of these data (microarray experiments, etc) and resources (T-DNA mutants, etc) are easily accessible thanks to a network of databases and stock centers available through a comprehensive online resource called The Arabidopsis Information Resource (TAIR; http://www.arabidopsis.org; [4]).

Arabidopsis is not only a model system but can also be used as a reference system where fundamentals are established and to which other plants can be compared. The most common cellular processes in plants should be conserved while others may have multiple evolutionary origins or correspond to specific processes that have evolved only in some plant families. The extent to which knowledge of cellular processes in Arabidopsis can be generalised has to be tested. However, Arabidopsis represents a powerful tool to understand the molecular bases of biological processes in other plants and even beyond! For instance, some researches conducted in Arabidopsis such as those on the imposition and maintenance of heterochromatic DNA methylation have contributed to related work in animal systems [5].

In this chapter, we will illustrate why the Arabidopsis system is such an efficient tool to study molecular mechanisms underlying different processes of tropical plant biology. We will discuss how this plant can be used to 1) identify candidate genes for interesting traits, 2) characterize genes isolated from other plant species, and 3) accelerate map-based gene cloning in important plant species.

1. Use of Arabidopsis to identify candidate genes

It is sometimes difficult to design an easy and efficient strategy to isolate plants genes responsible for an interesting trait in tropical plants. One possible approach is to identify candidate genes that are involved in a similar biological process in the model plant *Arabidopsis thaliana* and then isolate an orthologue in the plant of interest. Indeed, the wealth of molecular data available on Arabidopsis combined with the development of user-friendly data-mining and analysis tools makes it easy to find Arabidopsis genes associated with a particular physiological or developmental process. Most of this work can be done *in silico* and needs only a limited amount of time.

1.1. Identifying Arabidopsis genes involved in a chosen biological process

Genes involved in a particular physiological or developmental process can be identified by their expression pattern and/or the corresponding mutant phenotype.

1.1.1. Mining microarray experiments databases

The sequencing of the Arabidopsis genome [2] paved the way for analyses of gene expression on a genome scale using microarrays. The availability of Web-based data mining interfaces (Table 1) makes it easy to look for Arabidopsis genes induced or repressed in a given condition. Among them, the Genevestigator interface [6] is designed to analyze expression data from Arabidopsis obtained using the Affimetrix system. The ATH1 Arabidopsis full genome microarray (Affimetrix/The Institute for Genomic Research) is based on *in situ* synthesis of high-density oligonucleotides on glass slides and represents approximately 23,750 genes from Arabidopsis [7]. Thousands of these arrays have been processed to study changes in gene expression in various conditions. Genevestigator enables a search of this reservoir of data for genes expressed in a given developmental stage/environmental condition or for the expression pattern of a chosen gene. It has the advantage of containing a coherent set of data from Arabidopsis obtained using a single hybridisation platform, making it more likely to identify biologically meaningful expression patterns. Validation experiments on selected genes have confirmed the results obtained *in silico* using Genevestigator [6]. Nevertheless, when a set of genes putatively expressed in a condition or developmental stage of interest are identified using Genevestigator, it is very important to interpret the results carefully and to confirm the expression pattern using classical methods (RT-PCR, northern blot).

1.1.2. Gene- and enhancer trap databases

Another way to find genes expressed in a given tissue or in response to given environmental condition is to use gene- or enhancer trap databases (Table 1). These databases can be used to look for genes expressed in particular organs or developmental stages. Instead of looking at mRNA levels in different conditions, gene- and enhancer-trap experiments are used to look for DNA regulatory regions responsible for specific gene expression. For gene traps, a transposon or T-DNA containing a promoterless reporter gene at its border is randomly inserted into the plant genome. Expression of the reporter gene will depend upon the insertion of the mobile DNA element in a transcribed region. The pattern of expression of the reporter gene will depend on the flanking promoter and will mimic the expression of the corresponding gene. The enhancer trap scheme is quite similar: a reporter gene under the control of a minimal promoter is inserted at the border of a mobile DNA (transposon, T-DNA). After random insertion of the mobile DNA into the plant genome, the expression of the reporter gene will depend on the presence of a nearby regulatory sequence or enhancer.

Once a gene- or enhancer trap line presenting an interesting pattern of reporter gene expression has been found, the corresponding gene has to be

identified. PCR techniques such as TAIL-PCR [8] allow amplification and cloning of genomic sequences flanking the mobile DNA (of known sequence). Since the Arabidopsis genome is completely sequenced, it is easy to find the position and orientation of the mobile element insertion and of the nearby Arabidopsis genes from a short genomic DNA sequence. However, enhancers are able to work at great distances (several kbp) and in both orientations and therefore it might be difficult to find the gene corresponding to a given enhancer trap expression profile from the T-DNA insertion site. Some of the gene- or enhancer trap web sites provide information about the putative T-DNA or transposon insertion sites. The expression pattern of the candidate genes has to be tested by studying their mRNA accumulation in different conditions (RT-PCR, northern blot, *in situ* hybridisation) and/or by testing the transcriptional activity of their promoter using transgenic plants containing a promoter-reporter gene fusion.

This approach was used in our group in order to study the molecular bases of lateral root development in the tropical tree *Casuarina glauca*. A collection of Arabidopsis GAL4-GFP enhancer trap lines was screened for lateral root specific expression pattern. The corresponding lines allowed us to identify a few genes specifically expressed during lateral root development such as members of the LOB gene family [9]. Orthologues of these genes in *C. glauca* are therefore good candidates for genes involved in lateral root development in this tropical tree.

1.1.3. Mutant phenotypes

The two strategies described above enable identification of genes whose expression is specific or enhanced during a particular physiological or developmental process. However, this does not prove that these genes are indeed involved in the processes concerned, which requires functional analysis. This can be easily achieved in Arabidopsis thanks to the high number of T-DNA insertion lines available, making it very likely to find a T-DNA insertion line within a gene of interest (about 85% of predicted protein-coding genes have at least one insertion within intron or exon; [10]). Insertion lines can be looked for online at the Arabidopsis insertion database (http://atidb.org) using a gene number or a sequence as a query. Insertion lines are freely available and can be easily ordered from stock centers (Table 1). The phenotype of a knock out mutant can be used to demonstrate the function of a gene specifically expressed in a given developmental or physiological process. However, functional redundancy can prevent the appearance of a clear phenotype. RNAi can be exploited to knock out a whole gene family, or ectopic expression strategies can be used to overcome this problem thanks to the ease of Arabidopsis genetic transformation [3].

Another approach is to look for mutants perturbed during the physiological or developmental process of interest. A large collection of Arabidopsis mutants is available and searchable using the stock centers web sites (Table 1). Using keywords, it is possible to browse a list of mutants with a particular phenotype (short root, resistance to auxin, resistance to salt, etc). For some of these mutants, the corresponding gene has been cloned and characterized. If so, the stock centers web sites provide links to the sequence data. This is a quick and convenient way to identify genes whose mutation perturbs a given biological process in Arabidopsis. This strategy has been used for instance to understand the molecular bases of wood formation. Wood formation is an interesting characteristic of trees. However, many herbaceous species including Arabidopsis form vascular cambium and secondary xylem. Arabidopsis mutants perturbed in secondary xylem development have been used to identify genes involved in this developmental process (for review see [11]). Orthologues of these genes in tree species are candidates for genes involved in wood formation.

Table 1. Arabidopsis online resources.

Microarray data mining		
Programme name	Web site	Ref.
ArrayExpress	http://www.ebi.ac.uk/arrayexpress/query/entry	[12]
Gene Expression Omnibus	http://www.ncbi.nlm.nih.gov/geo/	[13]
Genevestigator	http://www.genevestigator.ethz.ch	[6]
NASCArrays	http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl	[14]
TAIR Microarray Experiments Search	http://www.arabidopsis.org/servlets/Search?type=expr&search_action=new_search	
Gene- and enhancer trap databases		
Collection	Web site	Ref.
CSHL Arabidopsis Genetrap	http://genetrap.cshl.org/	[15]
Jim Haseloff's laboratory	http://www.plantsci.cam.ac.uk/Haseloff/gene_expression/geneExpFrameset.html	
Scott Poethig's laboratory	http://enhancertraps.bio.upenn.edu/	
Stock centers		
Stock center	Web site	
ABRC	http://www.arabidopsis.org/abrc/	
NASC	http://nasc.nott.ac.uk/	
SASSC	http://www.brc.riken.jp/lab/epd/Eng/index.html	

1.2. Identification of orthologues of Arabidopsis genes in other plants

The experimental approaches presented above allow the use of the resources and genetic potential of Arabidopsis to identify genes involved in a given biological process. Once candidate genes have been identified in Arabidopsis, three different strategies can be used to identify homologues in the plant of interest. Homologues are genes that have a common evolutionary origin as shown by their similarity in all or part of their sequence. Among homologues, one should distinguish between paralogues, which arise from duplication within a single genome and orthologues, which are equivalent genes in different species that evolved from a common ancestor by speciation.

The first technique is to screen a cDNA library. cDNA prepared from RNA extracted from the tissue and the condition of interest are screened with a labelled probe corresponding to a conserved region of the candidate Arabidopsis gene. Hybridization should be performed under low-stringency conditions. cDNA inserts from positive clones can then be sequenced.

The second approach uses PCR. Degenerated primers corresponding to conserved regions of the Arabidopsis protein can be used to amplify homologous sequences. Amplification can be performed either on genomic DNA (searching for all the homologues in the genome) or on cDNA corresponding to a specific developmental or physiological treatment (identifying homologues expressed in a particular condition). PCR products can be cloned and sequenced. Full-length cDNA are then obtained using RACE-PCR or by screening cDNA libraries. This strategy was used to isolate homologues of the *AUX1* gene in the tropical tree *C. glauca* (our laboratory, unpublished). *AUX1* belongs to a small gene family in Arabidopsis comprising four genes: *AUX1, LAX1, LAX2* and *LAX3* [16]. PCR primers corresponding to conserved regions of the AUX-LAX protein family were designed using sequences from Arabidopsis [16], *Medicago truncatula* [17] and Poplar. Using low annealing temperature (48°C) in the amplification programme, we were able to obtain PCR products from genomic DNA corresponding to two genes called *CgAUX1* and *CgLAX3* according to their sequence homology to Arabidopsis genes. Full-length cDNA clones were obtained by RACE-PCR and the corresponding genomic clones were obtained by PCR on genomic DNA. Further studies suggested that the *AUX1* gene family contains only those two genes in *C. glauca* (our laboratory, unpublished). This strategy was also used in our laboratory to isolate homologues of different families of transcription factors (MADS-box and KNOX for example) or proteases (cystein-proteinase) putatively involved in different aspects of the development of oil palm (*Elaeis guineensis* Jacq.). PCR primers were designed using sequences from Arabidopsis, Antirrhinum, rice and maize. PCR experiments were carried out at low annealing temperature to allow the isolation of

different homologues expressed in the same tissues or in the same physiological condition. The PCR fragments obtained were then sequenced and used for cDNA library screenings to obtain the corresponding full-length cDNAs.

Finally, if ESTs data are available for the tissue and condition of interest, they can be searched *in silico* for homologues using the BLAST programme. For instance, we were able to identify an auxin efflux carrier encoding gene expressed in actinorhizal nodules of *C. glauca* by searching a nodule EST database with an Arabidopsis sequence corresponding to *PIN1*. We also used this strategy to isolate a *SHOOTMERISTEMLESS*-like cDNA, a *JOINTLESS*-like MADS-box cDNA and various cystein-proteinase encoding cDNAs from *E. guineensis* apex, inflorescence and zygotic embryo EST libraries respectively.

Sequences isolated using those techniques represent homologues of the Arabidopsis sequence. If these genes belong to large gene families, it can be difficult to identify the true orthologue of the Arabidopsis candidate gene. In order to do so, some experiments can be conducted. First of all, the pattern of expression of the isolated genes can provide interesting information and can be compared to the Arabidopsis gene in similar conditions. Complementation of the Arabidopsis mutant (if available and showing a phenotype) by expressing the homologue gene from the plant of interest under the control of the endogenous Arabidopsis promoter can be used to show that the proteins have similar functions (see below).

2. Use of Arabidopsis for functional analysis of genes

Before the development of plant transformation techniques, functional analysis of plant genes was limited to the use of heterologous systems such as *Escherichia coli, Saccharomyces cerevisiae* or *Schizosaccharomyces pombe*, which are easily transformable. These systems were used to identify or test the function of plant proteins putatively involved in conserved metabolic or cellular processes such as fatty acid metabolism, ion transport, metal tolerance, and cell cycle control [18,19,20,21] and they are still used for these kinds of studies (see [22,23] for examples). Nevertheless, these systems are limited to processes that are highly conserved between plants and unicellular prokaryotes or eukaryotes.

The availability of rapid transformation techniques using *Agrobacterium* prepared the way for studies of gene function in plant systems. A few years ago, tobacco (*Nicotiana tabacum*) was the model system for this sort of analysis because it was easy to transform and grow. Because of its rapid transformation without tissue culture (floral-dipping, [3]), its biological characteristics and its status as a plant model system, Arabidopsis is now widely used for functional analysis and is often referred as the "green yeast".

In order to analyze the biological function of plant genes, several approaches can be used. Studies of the expression pattern of the gene

concerned using northern-blotting, PCR-derived techniques (RT-PCR or real-time RT-PCR), histological detection (*in situ* hybridization) and promoter analysis in transgenic plants provide some interesting information. Changes in plant gene expression, either down-regulation or ectopic expression in transgenic plants can provide clues to the biological function of the gene under study by characterization of the resulting phenotype. But these approaches are often difficult or impossible in tropical plant species since these species are generally not easily transformable due to their biological characteristics (low *in vitro* culture rates, large, perennial plants, etc.) or the limited number of researchers working on such plants. Arabidopsis provides a way to overcome these limitations, it can be used to study the function of tropical plant genes using 1) overexpression or targeted expression of the native or modified gene in wild-type plants; 2) complementation of mutants, 3) analysis of promoters and 4) analysis of protein localization and protein-protein interactions.

2.1. Overexpression in wild-type plants

A simple method to study gene function is to overexpress the cDNA or genomic sequence-derived coding region under the control of the strong and constitutive 35S promoter from the Cauliflower Mosaïc Virus (CaMV) [24] in transgenic Arabidopsis plants. This method is very easy to develop in a laboratory, as numerous binary vectors carrying this promoter and different selective markers (antibiotics, herbicides e.g. Basta) are available for cloning. Phenotypic analysis of the transgenic plants provides some information about the function of the corresponding gene. Nevertheless, this method is limited by the fact that the phenotype observed may not be directly linked to the native function of the gene as it is expressed in tissues or at developmental stages where it is not normally expressed. It is very important to corroborate the phenotype with the native expression pattern of the gene of interest. An alternative is to use promoters driving a specific expression pattern such as tissue or stress specific promoters or inducible promoters such as ethanol-, glucocorticoid- or heat-inducible ones [25,26,27]. These promoters are very helpful in the case of genes that may be lethal especially during embryogenesis.

This approach has been widely used to study plant genes putatively involved in flower development. This is due to 1) the fact that this developmental process is well characterized in Arabidopsis and 2) the short time needed to obtain flowers in comparison to other species. Genes from varied species such as citrus [28,29], eucalyptus [30], rice [31], orchids [32,33], and gymnosperms [34,35,36] have been tested in Arabidopsis for their involvement in flower development. Most of these studies were performed using overexpression of the cDNA under the control of the 35S promoter. This experimental scheme was used in our laboratory to analyze the function of

different MADS-box transcription factors linked to flower development isolated from oil palm. The lack of an efficient protocol for stable genetic transformation of this species made it impossible to study the function of the genes in oil palm. We consequently expressed the cDNA of these genes under the control of the 35S promoter in transgenic Arabidopsis plants in order to understand the function of the corresponding proteins. In some cases, no modification of the phenotype was observed even if the transgenes were expressed, showing that the oil palm proteins were not able to interact with Arabidopsis DNA or proteins. We sometimes observed phenotypes similar to those obtained with the Arabidopsis homologue suggesting that the oil palm proteins had similar functions to their Arabidopsis counterparts. Finally, we sometimes found new phenotypes suggesting a dominant negative effect of the oil palm protein in the Arabidopsis system. This suggests partial divergence of the protein specificities between the two species.

Other uses of this approach include the functional analysis of genes putatively involved the cold and drought stress responses [37] such as the DREB1/CBF transcription factors from rice, maize and soybean [38,39,40].

2.2. Mutant complementation

Another possible strategy is to complement an Arabidopsis mutant, if available and showing a phenotype, with a homologue gene from the species of interest. This approach will show if a protein from the species of interest can functionally replace an Arabidopsis protein. If this is the case, all the experimental data available on the Arabidopsis protein can be used to understand the function of the heterologous protein. This type of analysis can be conducted only if Arabidopsis homologues exist and the corresponding mutants have already been described. Ideally this experiment is performed by expressing the heterologous gene or cDNA under the control of the endogenous Arabidopsis gene promoter in the mutant background. Alternatively, the 35S promoter can be used. The wealth of Arabidopsis mutants combined with the fact that they can be easily searched and ordered make this approach very attractive.

We used this approach in our laboratory to characterise two *C. glauca* homologues of the *AUX1* gene encoding a putative auxin influx carrier [41]. *aux1* mutant is agravitropic and has a reduced number of lateral roots [41]. The cDNA and genomic sequences corresponding to the *C. glauca* genes were cloned between the *AtAUX1* promoter and the *AtAUX1* terminator. These constructs were introduced into *aux1* mutant plants by floral dip transformation [3]. Analysis of the phenotype of the transgenic plants will indicate whether or not the *C. glauca* genes function as auxin influx carriers.

Many examples of gene characterization using this approach are described in the literature. For instance, different maize genes encoding enzymes of the

flavonoid biosynthesis pathway were analyzed by complementation of Arabidopsis mutants affected in flavonoid biosynthesis such as *ttg1*, *tt3*, *tt4*, *tt5* and *tt7*, restoring the ability of these mutants to accumulate pigments in seed coats and seedlings [42]. Similarly, a cotton gene encoding a LRR receptor-like protein kinase similar to the Arabidopsis brassinosteroid receptor was shown to complement the corresponding mutant named *bri1* [43] thus indicating that it functions as a brassinosteroid receptor gene. A cotton (*Gossypium* sp.) MYB transcription factor gene, *GaMYB2*, expressed in cotton seed trichomes, namely in cotton fibers, is able to complete the Arabidopsis *gl1* mutant affected in trichome formation, when expressed using the GL1 or the 35S promoter [44]. This indicates that *GaMYB2* has the same function as *GL1* and acts as a regulator of trichome development in cotton. Finally, Arabidopsis flower mutants have been used for functional analysis of homologous genes from species such as citrus [29], rice [45], Cycas [36], and Gnetum [34].

2.3. Promoter analysis in Arabidopsis

A powerful tool to study the regulation of the expression of a given gene is to isolate and analyze its promoter region in transgenic plants. Since transformation of most tropical plants is difficult or impossible for various reasons, this kind of study can instead be performed on Arabidopsis. The only limitations are 1) the fact that this approach cannot be used for biological processes specific to certain plant genera and not present in Arabidopsis, and 2) the level of conservation of gene regulation in the heterologous model system. However, various examples in the literature show that regulatory mechanisms are quite often conserved between Arabidopsis and other angiosperms and even gymnosperms. In order to test this, the expression pattern of the studied gene in a homologous environment (as studied by northern blot and *in situ* hybridisation) has to be compared with the expression pattern conferred by the promoter in Arabidopsis (as studied by fusing it to a reporter gene such as *uidA* or *GFP*). If the regulation mechanisms are conserved, Arabidopsis offers many tools for promoter analyses.

Transgenic Arabidopsis plants carrying promoter-reporter gene fusions can be used to study the factors controlling promoter activation (developmental stages, hormones, temperature, salinity, etc). The wealth of genetic resources can be exploited to analyze the signal transduction pathway leading to activation of the promoter using characterised mutants perturbed in a given pathway (hormone transduction pathways for instance).

The promoter can be dissected by 5' and 3' deletion in order to detect the regulatory sequences responsible for promoter activity. Since Arabidopsis transformation is efficient and quick, this can be achieved quite easily. This result can then be used to isolate proteins interacting with the regulatory

sequences. The interaction of a putative regulatory protein and the promoter can also be tested in Arabidopsis by expressing it ectopically and testing the effect of this misexpression on the expression of promoter-reporter genes fusions. For example, the cotton fiber-specific *RDL1* promoter directs trichome-specific expression in Arabidopsis plants [44]. Since cotton fibers are seed trichomes, this indicates that the regulatory mechanisms are conserved between those two plants. Dissection of the cotton promoter in Arabidopsis identified two regulatory domains, a homeobox-binding and a MYB binding sequence. This was used to identify two cotton transcription factors, GaMYB2 and GhHOX3, a MYB and a homeodomain transcription factors respectively, which are able to activate the cotton *RDL1* promoter in Arabidopsis and are highly expressed in cotton fiber cells [44].

Finally, transgenic plants bearing a homozygous promoter-reporter gene fusion can be exploited to genetically dissect the transduction pathway leading to gene expression. These plants can be used in a mutant screen to isolate plants that do not properly express the reporter gene. The characterization of the corresponding mutants can be used to identify plant genes controlling the expression of the gene of interest.

In our laboratory, we use Arabidopsis to analyze the regulation of tropical plant genes. For instance, the promoter region of an oil palm *SHOOTMERISTEMLESS*-like gene is able to drive the expression of the GFP or the GUS encoding genes specifically in the shoot apical meristem of Arabidopsis, showing that the regulatory domains of expression are conserved between the two species. We are also using this strategy to analyze the promoter region of the metallothionein gene *CgMT1* from the tropical tree *C. glauca* [46]. The activation of the *CgMT1* promoter is conserved in other plants such as tobacco, rice [47] and Arabidopsis (our laboratory, unpublished). We showed that the *CgMT1* promoter is activated by oxidative stress but not by metals in Arabidopsis. We are currently dissecting the *CgMT1* promoter in order to identify the DNA motifs responsible for this regulation.

2.4. Protein-protein interactions and cellular localization

Transgenic Arabidopsis can also be used to study protein interactions using FRET or BiFC, which allow the detection of protein-protein interactions *in planta*. These systems are based on fluorescence detection in the plant cells [48,49,50]. They seem to be more specific that the yeast two-hybrid system that is based on the GAL4 activation system [51] and have the advantage to occur in a plant cell allowing plant specific post-translational modifications.

The cellular localization of a studied protein provides important information about its function. A translational fusion between a gene of interest and a gene encoding a fluorescent protein, such as *GFP*, can be expressed either constitutively (using the 35S promoter) or under the control of

its own promoter in transgenic Arabidopsis plants. The localization of the fluorescence indicates where the protein under study is targeted in the cell. Moreover, the cellular localization of some proteins (transcription factors for instance) is dynamically regulated. Arabidopsis transgenics containing the protein-GFP fusion can be used to study the genetic, developmental (tissue specificity for example) or physiological (such as hormones) factors controlling cellular targeting. Another approach is to express a fusion between the protein concerned and a protein tag (c-myc for instance) that can be easily detected using corresponding antibodies.

3. Use of Arabidopsis for positional cloning of genes of interest in tropical plants

Synteny or colinearity can be defined as the study of chromosomal regions among closely or distantly related genomes that have conserved genes and markers. DNA sequence information offers unique opportunities for comparative analysis of genomes. Chromosomal synteny is a very efficient approach for the rapid identification of additional molecular markers in a region of interest and therefore facilitates map-based cloning of genes of interest in targeted species.

Analyses of genome colinearity between distantly related species can be problematic due to reduced gene similarities and recognition of orthologous sequences. One of the key questions is whether map position and order of genes in Arabidopsis and crop species are sufficiently conserved to help the map-based cloning of genes corresponding to interesting traits. Several comparative genetic mapping studies, that include Arabidopsis, indicate that synteny exists across higher plant families, even between Monocots and Dicots [52,53,54,55,56,57]. *Arabidopsis thaliana* is consequently a model of choice to accelerate map-based cloning using comparative genomics, since its genome is completely sequenced [2], and numerous DNA markers and ESTs are available.

Recent studies combined the exploitation of the synteny with the Arabidopsis genome and map-based cloning strategies for rapidly reducing genetic distance to targeted loci, thereby facilitating positional cloning. Below are few examples of how to use Arabidopsis to accelerate map-based gene cloning.

Rossberg *et al.* [56] compared the degree of microcolinearity in the 57 kbp region of the tomato *Lateral Suppressor* gene with Arabidopsis and capsella (a plant closely related to *Arabidopsis*) genomes. These authors were able to find homologous sequences for all five genes of the region. They demonstrated microsynteny between closely and distantly related dicotyledonous species. They concluded that the level of microcolinearity could be exploited to

localize orthologous genes in Arabidopsis and tomato without any ambiguity. Microsynteny between tomato and Arabidopsis was used for positional cloning of the major tomato fruit-shape locus, *ovate* [55]. Colinear segments in tomato chromosome 2 and Arabidopsis chromosome 4 were identified from screened sequences of 2 selected tomato bacterial artificial chromosome (BAC) clones containing the *ovate* locus. Annotated ORFs corresponding to the Arabidopsis syntenic region were compared with a *Solanaceae* EST database leading to the identification of homologous tomato ESTs. These ESTs were then used as markers and mapped. This enabled the construction of a high-resolution map of the *ovate* locus and the isolation of a BAC clone containing the *ovate* gene. In another comparative tomato-Arabidopsis study, Oh *et al.* [58] successfully used microsynteny-based comparative mapping to facilitate the positional cloning of the tomato *DIAGEOTROPICA* (*Dgt*) gene. Tomato RFLP markers from the *Dgt* region were compared to *Arabidopsis* genome and microsyntenic regions were identified in Arabidopsis. Gene sequences from the Arabidopsis syntenic regions were compared to the Tomato Gene Index database to isolate homologous tomato ESTs. These ESTs were converted into co-dominant molecular markers *via* cleaved amplified polymorphic sequence (CAPS). This was used to narrow the genetic distance to *Dgt* locus from 0.8 to 0.15 cM and to localize the *Dgt* gene on 2 overlapping tomato BAC clones.

Comparative sequence analysis of *Arabidopsis* and legume genomes also revealed extensive microcolinearity. Grant *et al.* [59] reported one of the first comparative genetic mapping studies on distantly related families; they demonstrated that significant synteny exists between soybean linkage group A2 and the *Arabidopsis* chromosome 1 over almost their entire lengths. Differences in marker order could be explained by only 2-3 chromosomal rearrangements. Zhu *et al.* [60] also examined syntenic relationships between *Medicago truncatula* and *Arabidopsis*. Using genetic map- and BAC sequence-based approaches, they assessed the level of synteny between the two species. They observed a lack of extended macrosynteny, but a conservation of marker colinearity over small genetic intervals. They also observed that the Arabidopsis genome often shared multiple points of synteny with genetically linked loci in *Medicago*. Recently, Stracke *et al.* [61], combined *Lotus japonicus* AFLP markers with *Arabidopsis* genome sequence information and the pea genetic map for positional cloning of the *Lotus LjSYM2* gene (required for the formation of nitrogen-fixing root nodules) and *PsSYM19*, its orthologue in *Pisum sativum*. They screened the genome of Arabidopsis with *Lotus* markers linked to the *LjSYM2* gene and identified colinear Arabidopsis genomic segments. All the predicted cDNAs within the Arabidopsis syntenic sequences were compared to a *Lotus* ESTs database allowing the identification of homologous cDNAs. New PCR markers were generated from these ESTs and mapped in the vicinity of the *LjSYM2* gene.

This was used to clone the *LjSYM2* gene and subsequently its orthologue in pea, the *PsSYM19*. They demonstrated that it is possible to exploit the *Arabidopsis* genome to generate tightly linked markers in a legume and consequently to map targeted genes.

In another study, Striling *et al.* [62] compared more than 300 kbp of DNA sequence of five *Populus* BAC clones with the genome of Arabidopsis and found significant microsynteny on the scale of BAC-sized DNA fragments. As a consequence, they suggested that DNA sequence and gene position data in Arabidopsis could be used for positional cloning efforts in *Populus* in a large proportion of cases. Georgi *et al.* [63] also compared 3 genomic regions of peach to the Arabidopsis genome. They were able to find short colinear segments (2 to 3 genes in length) located in different positions in the Arabidopsis genome. Nevertheless, they concluded that the *A. thaliana* sequence was extremely useful not only to identify putative coding regions in peach genome but also to determine their intron–exon structure. Yang *et al.* [64] also assessed the degree of synteny between the Arabidopsis genome and a 282 kbp region surrounding the *Citrus tristeza* virus resistance gene (*Ctv*) locus in *Poncirus trifoliata* (a specie closely related to *Citrus*). In this case the microsynteny observed was not sufficient to be useful for positional cloning.

All theses studies demonstrate the tremendous interest of comparative genomics using the Arabidopsis genome. Exploitation of this information will help improve our knowledge on cultivated plants and facilitate the positional cloning of targeted genes in tropical crops.

Conclusion

The entire genome sequence and the wealth of tools developed for functional analysis in *Arabidopsis thaliana* can be used to understand development and physiology in other plants. It can be used for gene identification (homology-based isolation, synteny-based positional cloning, etc), functional analysis (overexpression, mutant complementation, etc), and analysis of promoter regions and associated regulatory processes. Nevertheless, even if *Arabidopsis thaliana* is a powerful system, it is important to bear in mind that it is a heterologous system that cannot reveal specific regulatory networks nor the whole range of biological processes found in the plant kingdom. For instance, about 90% of Angiosperms can enter mycorrhizal symbioses with soil fungi but Arabidopsis does not. Other model plant systems such as rice (for cereals), *Medicago truncatula* (for plant-microbe interactions) or poplar (for trees) are consequently emerging to circumvent this problem. Moreover, it is important to complete all studies of tropical plant genes carried out in Arabidopsis with analyses in the native species (for example the gene expression pattern).

References

1. Rédei, G.P. 1992, Methods in Arabidopsis research, C. Koncz, N.-H. Chua, and J. Schell (Eds.), World scientific, Singapore, 1.
2. Arabidopsis Genome Initiative. 2000, Nature, 408, 796.
3. Clough, S., and Bent, A.F. 1998, Plant J., 16, 735.
4. Rhee, S.Y., Beavis, W., Berardini, T.Z., Chen, G., Dixon, D., Doyle, A., Garcia-Hernandez, M., Huala, E., Lander, G., Montoya, M., Miller, N., Mueller, L.A., Mundodi, S., Reiser, L., Tacklind, J., Weems, D.C., Wu, Y., Xu, I., Yoo, D., Yoon, J., and Zhang, P. 2003, Nucl. Acid Res., 31, 224.
5. Gendrel, A.-V., Lippman, Z., Yordan C., Colot, V., Martienssen, R.A. 2002, Science, 297, 1871.
6. Zimmermann, P., Hirsch-Hoffmann, M., Hennig, L., and Gruissem, W. 2004, Plant Physiol., 136, 2621.
7. Redman, J.C., Haas, B.J., Tanimoto, G., and Town, C.D. 2004, Plant J., 38, 545.
8. Liu, Y.G., Mitsukawa, N., Oosumi, T., and Whittier, R. 1995, Plant J., 8, 457.
9. Shuai, B., Reynaga-Pena, C.G., and Springer, P.S. 2002, Plant Physiol., 129, 747.
10. Bevan, M., and Walsh, S. 2004, Plant Physiol., 135, 602.
11. Nieminen, K.M., Kauppinen, L., and Helariutta, Y. 2004, Plant Physiol., 135, 653.
12. Brazma, A., Parkinson, H., Sarkans, U., Sojatalab, M., Vilo, J., Abeygunawardena, N., Holloway, E., Kapushesky, M., Kemmeren, P., Lara, G.G., Oezcimen, A., Rocca-Serra, P., and Sansone, S.A. 2003, Nucl. Acid Res., 31, 68.
13. Edgar, R., Domrachev, M., and Lash, A.E. 2002, Nucl. Acid Res., 30, 207.
14. Craigon, D.J., James, N., Okyere, J., Higgins, J., Jotham, J., and May, S. 2004, Nucl. Acid Res. (Database issue), 32, D575.
15. Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D.G., Dean, C., Ma, H. and Martienssen, R.A.. 1995, Genes Dev., 9, 1797.
16. Parry, G., Marchant, A., May, S., Swarup, R., Swarup, K., James, N., Graham, N., Allen, T., Martucci, T., Yemm, A., Napier, R., Manning, K., King, G., and Bennett, M. 2001, J. Plant Growth Reg., 20, 217.
17. De Billy, F., Grosjean, C., May, S., Bennett, M., and Cullimore, J. 2001, Mol. Plant Microbe Interact., 14, 267.
18. Hirayama, T., Imajuku, Y., Anai, T., Matsui, M., and Oka, A. 1991, Gene, 105, 159.
19. Kuromori, T., and Yamamoto, M. 1994, Nucl. Acid Res., 22, 5296.
20. Xia, G., Ramachandran, S., Hong, Y., Chan, Y.S., Simanis, V., and Chua, N.H. 1996, Plant J., 10, 761.
21. Cabello-Hurtado, F., Batard, Y., Salaun, J.P., Durst, F., Pinot, F., and Werck-Reichhart, D. 1998, J. Biol. Chem., 273, 7260.
22. Champion, A., Jouannic, S., Guillon, S., Mockaitis, K., Krapp, A., Picaud, A., Simanis, V., Kreis, M., and Henry, Y. 2004, J. Cell Sci., 117, 4265.
23. Zhang, M.Y., Bourbouloux, A., Cagnac, O., Srikanth, C.V., Rentsch, D., Bachhawat, A.K., and Delrot, S. 2004, Plant Physiol., 134, 482.
24. Odell, J.T., Nagy, F., and Chua, N.H. 1985, Nature, 313, 810.
25. Roslan, H.A., Salter, M.G., Wood, C.D., White, M.R., Croft, K.P., Robson, F., Coupland, G., Doonan, J., Laufs, P., Tomsett, A.B., and Caddick, M.X. 2001, Plant J., 28, 225.

26. Zuo, J., Niu, Q.W., and Chua, N.H. 2000, Plant J., 24, 265.
27. Yoshida, K., Kasai, T., Garcia, M.R., Sawada, S., Shoji, T., Shimizu, S., Yamazaki, K., Komeda, Y., and Shinmyo, A. 1995, Appl. Microbiol. Biotechnol., 44, 466.
28. Pena, L., Martin-Trillo, M., Juarez, J., Pina, J.A., Navarro, L., and Martinez-Zapater, J.M. 2001, Nat. Biotechnol., 19, 263.
29. Pillitteri, L.J., Lovatt, C.J., and Walling, L.L. 2004, Plant Physiol., 135, 1540.
30. Kyozuka, J., Harcourt, R., Peacock, W.J., and Dennis, E.S. 1997, Plant Mol. Biol., 35, 573.
31. Fornara, F., Parenicova, L., Falasca, G., Pelucchi, N., Masiero, S., Ciannamea, S., Lopez-Dee, Z., Altamura, M.M., Colombo, L., and Kater, M.M. 2004, Plant Physiol., 135, 2207.
32. Hsu, H.F., and Yang, C.H. 2002, Plant Cell Physiol., 43, 1198.
33. Hsu, H.F., Huang, C.H., Chou, L.T., and Yang, C.H. 2003, Plant Cell Physiol., 44, 783.
34. Winter, K.U., Saedler, H., and Theissen, G. 2002, Plant J., 31, 457.
35. Carlsbecker, A., Sundstrom, J., Tandre, K., Englund, M., Kvarnheden, A., Johanson, U., and Engstrom, P. 2003, Evol Dev., 5, 551.
36. Zhang, P., Tan, H.T., Pwee, K.H., and Kumar, P.P. 2004, Plant J., 37, 566.
37. Zhang, J.Z., Creelman, R.A., and Zhu, J.K. 2004, Plant Physiol., 135, 615.
38. Dubouzet, J.G., Sakuma, Y., Ito, Y., Kasuga, M., Dubouzet, E.G., Miura, S., Seki, M., Shinozaki, K., and Yamaguchi-Shinozaki, K. 2003, Plant J., 33, 751.
39. Kim, J.C., Lee, S.H., Cheong, Y.H., Yoo, C.M., Lee, S.I., Chun, H.J., Yun, D.J., Hong, J.C., Lee, S.Y., Lim, C.O., and Cho, M.J. 2001, Plant J., 25, 247.
40. Qin, F., Sakuma, Y., Li, J., Liu, Q., Li, Y.Q., Shinozaki, K., and Yamaguchi-Shinozaki, K. 2004, Plant Cell Physiol., 45, 1042.
41. Bennett, M.J., Marchant, A., Green, H.G., May, S.T., Ward, S.P., Millner, P.A., Walker, A.R., Scultz, B., and Feldman, K.A. 1996, Science, 273, 948.
42. Dong, X., Braun, E.L., and Grotewold, E. 2001, Plant Physiol., 127, 46.
43. Sun, Y., Fokar, M., Asami, T., Yoshida, S., Allen, R.D. 2004, Plant Mol. Biol., 54, 221.
44. Wang, S., Wang, J.W., Yu, N., Li, C.H., Luo, B., Gou, J.Y., Wang, L.J., and Chen, X.Y. 2004, Plant Cell, 16, 2323.
45. Chujo, A., Zhang, Z., Kishino, H., Shimamoto, K., and Kyozuka, J. 2003, Plant Cell Physiol., 44, 1311.
46. Laplaze, L., Gherbi, H., Duhoux, E., Pawlowski, K., Auguy, F., Guermache, F., Franche, C., and Bogusz, D. 2002, Plant Mol. Biol., 49, 81.
47. Ahmadi, N., Dellerme, S., Laplaze, L., Guermache, F., Duhoux, E., Bogusz, D., Guiderdoni, E., and Franche, C. 2003, Trans. Res. 12, 271.
48. Immink, R.G., Gadella, T.W. Jr, Ferrario, S., Busscher, M., and Angenent, G.C. 2002, Proc. Natl Acad. Sci. USA, 99, 2416.
49. Brecht, M., Sewald, K., Schiene, K., Keen, G., Fricke, M., Sauer, M., and Niehaus, K. 2004, J. Biotechnol., 112, 151.
50. Walter, M., Chaban, C., Schutze, K., Batistic, O., Weckermann, K., Nake, C., Blazevic, D., Grefen, C., Schumacher, K., Oecking, C., Harter, K., and Kudla, J. 2004, Plant J., 40, 428.
51. Evangelista, C., Lockshon, D., and Fields, S. 1996, Trends Cell Biol., 6, 196.

52. Paterson, A.H., Lan, T.-H., Reischmann, K.P., Chang, C., Lin, Y.-R., Liu, S.-C., Burow, M.D., Kowalski, S.P., Katsar, C.S., DelMonte, T.A., Feldmann, K.A., Schertz, K.F. and Wendel, J.F. 1996, Nature Genet., 14, 380.
53. Devos, K.M., Beales, J., Nagamura, Y., and Sasaki, T. 1999, Genome Res., 9, 825.
54. van Dodeweerd, A.M., Hall, C.R., Bent, E.G., Johnson, S.J., Bevan, M.W., and Bancroft, I. 1999, Genome, 42, 887.
55. Ku, H.M., Liu, J., Doganlar, S., and Tanksley S.D. 2001, Genome, 44, 470.
56. Rossberg, M., Theres, K., Acarkan, A., Herrero, R., Schmitt, T., Schumacher, K., Schmitz, G., and Schmidt, R. 2001, Plant Cell, 13, 979.
57. Salse, J., Piegu, B., Cooke, R., and Delseny, M. 2002, Nucl. Acid Res., 30, 2316.
58. Oh, K.C., Hardeman, K., Ivanchenko, M.G., Ellard-Ivey, M., Nebenfohr, A., White, T.J., and Lomax, T.L. 2002, Genome Biol., 3, 0049.1.
59. Grant, D., Cregan, P., and Shoemaker, R.C. 2000, Proc. Natl Acad. Sci. USA, 97, 4168.
60. Zhu, H., Kim, D.J., Baek, J.M., Choi, H.K., Ellis, L.C., Küester, H., McCombie, W.R., Peng, H.M., and Cook, D.R. 2003, Plant Physiol., 131, 1018.
61. Stracke, S., Sato, S., Sandal, N., Koyama, M., Kaneko, T., Tabata, S., and Parniske, M. 2004, Theor. Appl. Genet., 108, 442.
62. Stirling, B., Yang, Z.K., Gunter, L.E., Tuskan, G.A., and Bradshaw, H.D. 2003, Can. J. For. Res., 33, 2245.
63. Georgi, L.L., Wang, Y., Reighard, G.L., Mao, L., Wing, R.A., and Abbott, A.G. 2003, Genome, 46, 268.
64. Yang, Z.N., Ye, X.R., Molina, J., Roose, M.L., and Mirkov, T.E. 2003, Plant Physiol., 131, 482.

Laplaze Laurent, Jouannic Stefan, Gherbi Hassen (2006)

Arabidopsis thaliana as a tool to study the molecular biology of tropical plants

In : Franche Claudine (ed.). *Molecular biology of tropical plants*

Kerala : Research Signpost, p. 143-160

ISBN 81-308-0138-8.