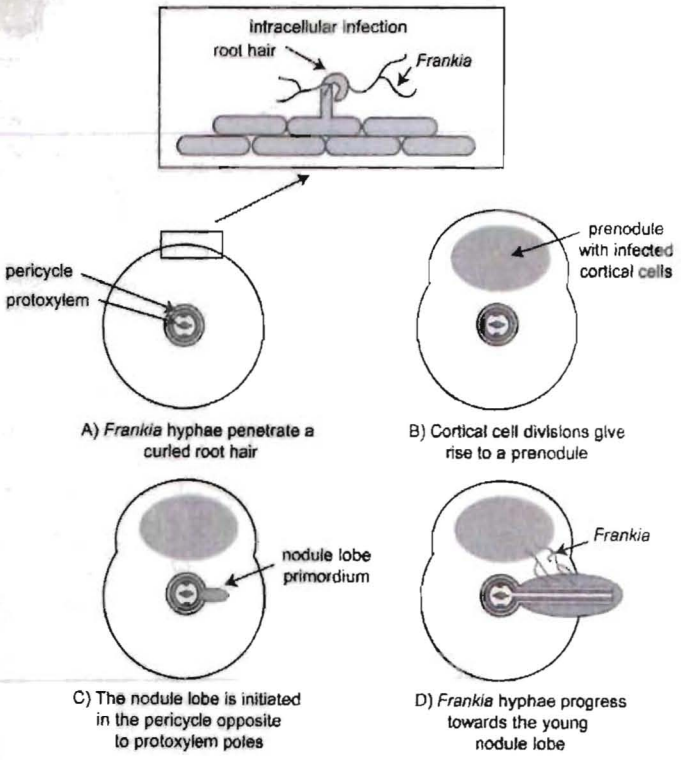
Molecular Biology of Tropical Plants

2006

Editor
Claudine Franche



RESEARCH SIGNPOST

Molecular Biology of Tropical Plants

2006

Editor
Claudine Franche

UMR 1098, Institut de Recherche
pour le Développement (IRD)
BP 64501, 911 Avenue Agropolis
34394 Montpellier cedex 5
France

Molecular Biology of Tropical Plants

2006

Published by Research Signpost

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

Editor
Claudine Franche

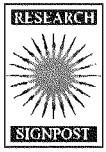
Managing Editor
S.G. Pandalai

Publications Manager
A. Gayathri

ISBN: 81-308-0138-8

Contents

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



Molecular Biology of Tropical Plants, 2006: 1-15 ISBN: 81-308-0138-8
Editor: Claudine Franche

1 Physiology and defence mechanisms to pathogens in tropical woody plants

Lecouls, A.C.[1], Petitot, A-S.[1], Marmey, P.[1], Silva, M-C.[2] Guerra-Guimaraes, L.[2], Fernandez, D.[1] and Nicole, M.[1]
[1]Institut de Recherche pour le Développement (IRD), UMR DGPC Unité « Résistance des Plantes » 911, avenue Agropolis, BP 64501, 34394 Montpellier Cedex 5, France; [2]Centro de Investigação das Ferrugens do Cafeeiro (CIFC); Quinta do Marques, 2784-505 Oeiras, Portugal

Abstract

Resistance of cotton (Gossypium hirsutum) to Xanthomonas campestris pv. malvacearum (Xcm) and of coffee (Coffea arabica) to the orange rust fungus (Hemileia vastatrix) and root-knot nematodes (Meloidogyne sp.) is characterized by a rapid hypersensitive cell death at the infection sites. To elucidate some of the mechanisms underlying these two plants defence reactions, molecular studies were undertaken by different strategies (candidate gene and global approaches). Two gene families, lipoxygenase

Correspondence/Reprint request: Dr. Lecouls, A.C., Institut de Recherche pour le Développement (IRD), UMR DGPC, Unité « Résistance des Plantes » 911, avenue Agropolis, BP 64501, 34394 Montpellier cedex 5, France
E-mail: Anne-claire.Lecouls@mpl.ird.fr

and peroxidase, were studied for their relationships with cotton defence to Xcm and the role they may have in the resistance strategy was physiologically investigated. A catalogue of EST involved in the coffee/rust interaction was also generated from cDNA subtractive libraries. Expression analyses lead to the identification of genes showing enhanced transcript accumulation in the early stages of coffee resistance to rust and nematodes, providing new insights into tropical woody plants responses to pathogens.

Introduction

Among tropical plants of agronomic interest, coffee (*Coffea* sp.) and cotton (*Gossypium hirsutum* L.) are two of the most valuable world's traded commodities, contributing to several billions US dollars annually. Production is distributed out of more than 70 tropical and subtropical developing countries where these cash crops are considered as an important currency source, but also as a favourable social stability factor [1]. The main cultivated varieties of these two crops are highly susceptible to several diseases and pathogens. For instance, cotton bacterial blight caused by *Xanthomonas campestris* pv. *malvacearum* (*Xcm*) is an important and potentially destructive disease affecting cotton yield losses in excess of 20%. Similarly, the orange rust fungus (*Hemileia vastatrix*) and root-knot nematodes (*Meloidogyne sp.*) are considered as the more important agronomic constraints in the major coffee-growing areas (Brazil, Latin America and Asia), regarding the substantial losses of vigour and yield they can cause in *C. arabica* plantations (70% of the production worldwide) [2].

Limitation of these pests and diseases in woody crops has generally been based on toxic, costly, and poorly effective chemical treatments, now widely restricted due to increasing environmental concerns. As an alternative, resistant varieties have been developed by traditional breeding approaches based on the exploitation of natural resistance sources. Coffee and cotton resistance to *H. vastatrix*, *M. exigua* and *Xcm*, respectively, are conferred by major genes [3, 4, 5, 6] whose phenotypic expression is the so-called hypersensitive response (HR), mediated by the gene-for-gene model [7].

The HR, which is activated early during the infection process, is characterized by the formation of necrotic lesions in the region of pathogen attack resulting from a programmed cell death directly responsible for the pathogen confinement and growth limitation. It is associated with activation of defence mechanisms in the dying area as well as in the surrounding tissues such as changes in protein phosphorylation, generation of reactive oxygen species (the oxidative burst), modification of ion fluxes, cell wall reinforcement by deposition of lignin and callose, lipid peroxidation, synthesis of antimicrobial molecules (phytoalexines), production of signalling hormones, and activation of pathogenesis-related (PR) genes [8, 9, 10]. Understanding of the cellular and

molecular mechanisms involved in plant defence reactions provides others possibilities to develop new management strategies, in addition to the exploitation of resistance genes in field. To this purpose, studies have been initiated to characterize physiological events triggered by the plant/pathogen recognition and identify genes specifically involved in the defence response of cotton to bacterial blight and of coffee to rust and nematodes.

So far, two different strategies have been used: a candidate gene approach for the cotton/*Xcm* interaction and a global strategy concerning the resistance of coffee to rust and nematodes. The candidate gene strategy focus on genes which have already been identified in other plant/pathogen interactions, since defence mechanisms and molecular processes underlying the HR have been extensively investigated in model species (*Arabidopsis thaliana*, rice, tobacco ...), but poorly in tropical woody plants [11, 12, 13]. Thus, the choice of a suitable candidate gene is based on the *a priori* belief that it could play a relevant role in the studied pathosystem and depends on the degree of homology with known-function sequences. The global strategy, (also called without *a priori* approach), consists in the establishment of a catalogue of genes implicated in resistance mechanisms. Construction of cDNA libraries corresponding to diverse interactions, systematic ESTs sequencing, and differential screening of the libraries foster the isolation of clones directly involved in the plant defence responses.

Candidate gene strategy: Cotton resistance to *Xanthomonas campestris*

Resistance of cotton plants (*Gossypium hirsutum* L.) challenged by the bacterial pathogen *Xcm* responds hypersensitively [14]. The Réba B50 cultivar carrying the B_2B_3 blight resistance genes which confer immunity to the avirulent race 18 of *Xcm* [15] was used to characterize early events involved in the establishment of cotton HR to the parasite [16]. Two gene families, lipoxygenase and peroxidase, were studied for their relationships with plant defence. The role these proteins may have in the resistance strategy of cotton was physiologically investigated.

1. Lipoxygenases

Dramatic damage undergone by membranes during HR was correlated with production of poly-unsaturated fatty acid (PUFA) hydroperoxides and associated with active oxygen species (AOS) generation. Induction of alteration of membrane structure by lipoxygenases (LOX) during HR was suggested as an alternative hypothesis to the AOS role for lipid peroxidation [17]. Activation of LOX was demonstrated in several plants both during compatible and incompatible interactions.

Accumulation of hydroperoxydes correlated with HR symptoms

Microscopical investigations of infected tissues revealed irreversible membrane disorganization during HR of the cotton cultivar Reba B50 to *Xcm* race 18 [18, 19], suggesting the existence of a causal link between membrane lipid peroxidation and hypersensitive cell death. Based on a detailed biochemical and molecular description [20], investigations on cotton cotyledon tissues undergoing HR discriminated between a free radical-mediated process, and a LOX pathway, i.e. non-specific *vs.* specific peroxidation, respectively. Hypersensitive death induced by the avirulent race 18 from *Xcm* is correlated with a massive LOX-mediated production of 9S-hydroperoxides, tissue dehydration and apparition of HR lesions on infected leaves at 24 hours post-infection (hpi). Regiospecificity and enantioselectivity of LOX activity characterized in cotyledon extracts at 24 hpi was in accordance with its involvement in the production of the 9S-hydroperoxides. Upstream from this LOX-mediated lipid peroxidation, a narrow peak of intense LOX activity was observed at 9 hpi (Figure 1-A). Analysis of the corresponding extracts by electrophoresis on IEF gels revealed the presence of different LOX isoforms. The activity of the acidic isoforms (particularly the pI 4.6 one) was correlated with induction of the HR, although it was slightly induced at 9 hpi also. The increase

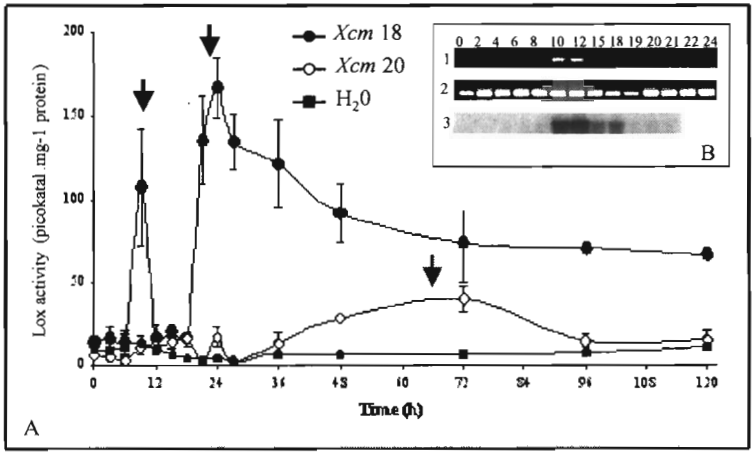


Figure 1. Activation of cotton lipoxygenase during the HR to *Xcm*. A: Time course of lipoxygenase activity in *Xcm*-infected cotton cotyledons showing two peaks in resistant plants (black arrows) and one peak, later and weaker, in susceptible plants (grey arrow); controls were performed on H_2O-infiltrated cotyledons. **B**: Transcriptional activity of cotton *GhLox1* gene during the HR (1: RT-PCR with *GhLox1* specific primers; 2: RT-PCR with actin primers; 3: Northern blot with *GhLox1* as probe).

in pl 4.6 LOX isoform activity and the appearance of a pl 7.4 band paralleled the increase in 9S-LOX activity at 9hpi, suggesting this band to likely have also a 9S-LOX specificity.

Transcription of LOX genes

A cotton LOX gene (*GhLox1*) was cloned from cotyledon tissues. Its expression during HR was studied by semi-quantitative RT-PCR and Northern blot (Figure 1-B). *GhLox1* transcripts were detected only during the incompatible reaction, between 6 and 18 hpi by RT-PCR, with a higher accumulation at 12 hpi, and between 10 and 18 hpi by Northern blot. The molecular determinants of the specificity of *GhLox1* were identified in the sequence and characterized this LOX gene as a 9-LOX. These results showed the correlation between LOX gene transcription and LOX activity.

Taken together, the present work gives evidence for a crucial role of 9S-LOX-mediated lipid peroxidation in the execution of HR cell death in cotton. Through which signalling pathway this LOX-dependent mechanism operates still remains debatable, but salicylic acid, methyl-jasmonate (MeJA), and hydrogen peroxide (H_2O_2) are putative candidates to be involved in LOX induction. In addition, the early and narrow production of MeJA 2 hpi during cotton HR indicated that other LOX genes (i.e. 13-LOX) could be associated with resistance.

2. Peroxidases

Peroxidases (POD; EC 1.11.1.1.7), a group of haem-containing glycosylated proteins, are known to be activated in response to pathogen attacks [21]. Several roles have been attributed to plant POD in host/pathogen interactions [22], including involvement in the HR. These proteins were studied for their potent activities during resistance of cotton to the bacterial blight.

Peroxidases are associated with the oxidative burst in cotton

The oxidative burst in plants infected by incompatible pathogens is a key early event in the expression of resistance [23, 24, 25]. In cotton, the oxidative burst generated AOS during resistance, including anion superoxides ($O_2^{\cdot-}$) and H_2O_2, resulting from dismutation of $O_2^{\cdot-}$ by a MnSOD [59]. Several lines of evidences strongly suggested that a wall-bound peroxidase was involved in the production of $O_2^{\cdot-}$, 3 hpi (Figure 2-A) [26]. Increase in activity of cationic POD isoforms (pI 9-9.4), positive effects of POD inhibitors on $O_2^{\cdot-}$ generation, immunolocalization of POD, and analysis of POD genes transcriptional activity reinforced the idea of a strong role of POD in the cotton burst to *Xcm* [27].

Changes in POD activity during infection was assessed spectrophotometrically according to time course of infection. It increased significantly during the

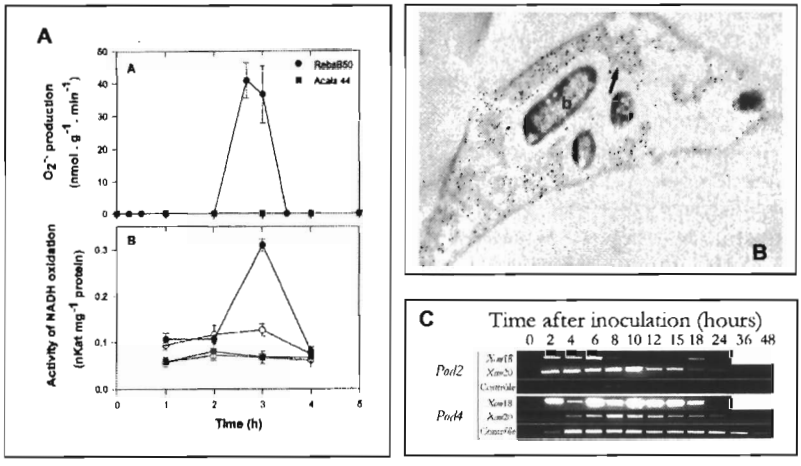


Figure 2. The oxidative burst in *Xcm* race 18-infected cotton cotyledons. A: Production of superoxide anions 3 hpi associated with NAD-oxidation; **B**: Immunolocalization of peroxidases 3 hpi, close to the bacteria, in the intercellular areas of exhibiting-HR cells; **C**: expression of POD genes assessed by RT-PCR: *pod2* was transcripted from 2 to 8 hours during HR, while the expression of *pod4* is higher in HR cells than in susceptible infected cells.

incompatible interaction between 8 and 10 hpi, as compared to the activity in the compatible interaction (Réba B50/*Xcm* race 20) and the control. A subsequent systemic increase in POD activity was also recorded one day after bacterial treatment, both locally in non-infected areas of cotyledons, and in the whole plant [28]. The highly localized accumulation of POD proteins revealed by immunocytochemistry in cotton cells 3 hpi (Figure 2-B) [27] is consistent with diamino-benzidine cytochemistry observations showing that POD activity was confined in the apoplasm and close to the bacteria [26]. POD identified in the apoplastic washing fluid were shown to be (1) cationic, (2) responsible for the production of superoxide anions, subsequently dismutated into wall-bound H_2O_2, (3) pre-formed, since they were detected in healthy cotyledons, and (4) inducible for their $O_2^{\cdot-}$ -generating NADH-oxidase activity in non-infected tissues [26]. POD accumulation in *Xcm*-encapsulating material is a response to HR development and suggests that cotton cells challenged by the pathogen create a localized, highly toxic environment, in line with AOS production, that results in limiting bacterial growth. This confined apoplastic localization of active POD may explain why the change in activity 3 hpi was not observed spectrophotometrically, but only on IEF gels with reference to apoplastic cationic isoforms [26].

Differential expression of peroxidase genes during HR

POD genes were cloned from two cDNA libraries made at early stages (2.5 and 5.5 hpi) of the cotton incompatible reaction [27]. Seven cotton POD genes were cloned (named *pod1*, *pod2*, *pod3*, *pod4*, *pod5*, *pod6* and *pod10*), with similarities to plant class III POD [29]. Putative signal peptide cleavage sites were identified in the coding sequence of all clones and no clear peroxisomal targeting signal sequences were detected, thus suggesting that these POD are located extracellularly.

The expression profile of each cDNA clone was determined by RT-PCR using specific primers, at different pi times for incompatible, compatible reactions and water infiltration. Analysis of gene expression showed variation in transcript accumulation during both compatible (race 20) and incompatible interactions for four of these genes. *pod2* was induced by pathogen infection and weakly stimulated in the control (Figure 2-C); *pod3* was specifically down-regulated during the HR after the oxidative burst; *pod4* and *pod6* were more intensely up-regulated during disease and in the control. All these data suggest that cotton peroxidases may have various functions in the defence response to *Xcm* infections.

Recently, an extensive study on the expression of *A. thaliana* class III POD [30] suggested that POD do not fullfill similar biological roles, even sequences are similar (>70% identical). Hence, it is difficult to predict the role of a particular POD on the basis of its similarity with another known POD. One response of cotton to *Xcm* was found to be a drastic accumulation of flavonoids in walls and cytoplasm of cells undergoing HR [19], suggesting a possible relation between increased POD activity and phenol oxidation. POD were shown to catalyze H_2O_2-dependent oxidation of flavonols [31] suggesting that the flavonoid-POD reaction can function as an H_2O_2 scavenging mechanism. Consistent with observations that H_2O_2 is synthesized apoplastically and flavonoids produced abundantly in cells at the edge of lesions during cotton HR to *Xcm*, the role of POD as an oxidative damage protectant must be examined in light of the fact that flavonoids are electron donors.

Global strategy: Coffee resistance to the orange rust fungus and root-knot nematodes

1. Coffee leaf rust resistance

Resistance of *C. arabica* varieties to leaf rust caused by *H. vastatrix* is conditioned by gene-for-gene interactions [32, 6, 33]. The resistance is expressed by rapid hypersensitive cell death at the infection sites (stomata) as early as 24 hpi [6, 34]. Macroscopically, HR lesions appear as chlorotic flecks 12 days post-inoculation (dpi) [6]. Biochemical and cytological analyses of coffee leaves showed that growth of the fungus in resistant coffee plants

usually ceased in the early stages of infection process, after the formation of the first haustorium. Host cell death was associated with precocious haustoria encasement with callose and ß-1,4-glucans. In addition, a peak of phenylalanine ammonia-lyase (PAL) activity was detected 2 dpi and coincided with the early accumulation of phenolic compounds [34].

Isolation of coffee genes involved in rust resistance: Subtractive EST libraries construction

The suppression subtractive hybridization (SSH) method [35] was used to generate cDNA libraries enriched in sequences expressed in coffee leaves during the early stages of HR [36]. Several studies indicated that a high number of plant genes are transcriptionally regulated upon challenge by a pathogen [37, 38, 39] but that most of them may be common to both compatible and incompatible interactions [40]. To focus on genes strictly involved in the HR, cDNA from plants infected with an avirulent *H. vastatrix* race (incompatible interaction) were subtracted with cDNA from plants infected with a virulent *H. vastatrix* race (compatible interaction).

Twelve, 24 and 48 hours were chosen as the appropriate time-points for isolation of the RNAs used to construct the subtractive cDNA libraries. Indeed, microscopic observations of fungal development indicated that the *H. vastatrix* isolates have germinated and developed appressoria 12 hpi, then penetrated through stomata 24 hpi and finally developed haustorial mother cells 48 hpi. At this time, in the incompatible interaction, the fungus ceased its growth and death of host cells is initiated [34].

In order to obtain a catalogue of expressed genes in rust infected-coffee plants, hundreds of ESTs were generated from the subtracted libraries (GenBank accession numbers: CF588584 to CF589197). ESTs showing similarities to plant protein database entries were classified into functional categories (Figure 3). They were mainly distributed in cell signalling/communication, cell/organism defence, gene/protein expression, and metabolism classes.

The EST from the cell defence category presented homologies with proteins known to be involved in apoptosis regulation in animal cells (Beclin and macrophage migration inhibitory factor), in several oxidative pathways (cytochrome P450 and oxidoreductases), in metal homeostasis and detoxification (metallothioneins) and in response to several stresses (heat-shock proteins). ESTs with homologies to components of *A. thaliana* resistance signalisation pathways such as NDR1 (non race-specific disease resistance) [41] and DND1 (defence no death) [42] proteins were also isolated. Other clones matched proteins involved in defence reactions, such as the pathogenesis-related (PR) proteins (chitinases, ß1,3-glucanases, PR-10), the chalcone synthase and the lipoxygenase enzymes. Finally, ESTs presented similarities to proteins involved in specific resistance to pathogens (tomato Asc-1

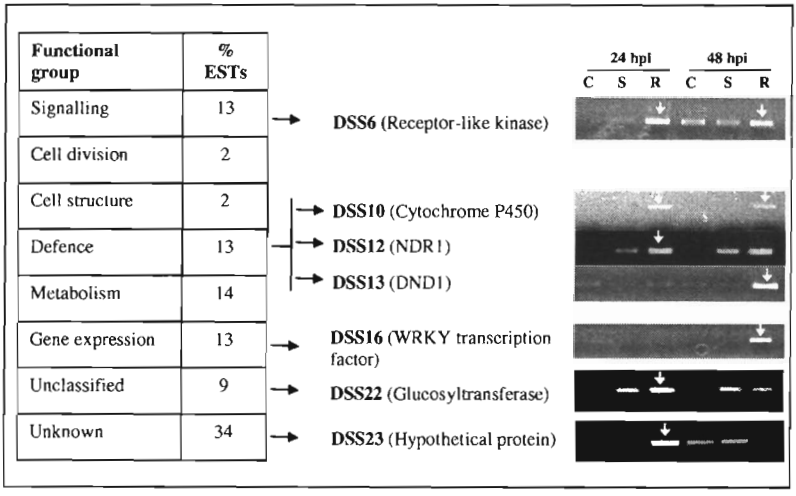


Figure 3. ESTs classification and examples of DSSs expression in *C. arabica* leaves inoculated with *H. vastatrix*. ESTs were classified into functional categories following the Expressed Gene Anatomy Database cellular role classification scheme [43]. DSSs expression analyses were performed by RT-PCR, 24 and 48 hpi. Putative function of DSSs is indicated into brackets. C: non-inoculated plants sprayed with water; S: compatible interaction; R: incompatible interaction. Arrows indicate gene differential expression in incompatible *vs.* compatible interactions. For experimental procedure details see [36].

protein, pepper importin alpha2) and disease resistance (R) proteins (CC-NBS-LRR class, tomato *Cf*-family). These data indicate the conservation in coffee plant of *R*-gene mediated resistance signalling pathways described in model plants.

In addition, a high proportion of the ESTs (34 %) matched genes with unknown function and 17 % had no or low similarity to protein database entries. These cDNA clones may represent an additional source of coffee genes potentially involved in the resistance response.

Gene expression during HR

Selection of cDNA clones specifically expressed in the resistant samples was performed by differential screening of the subtractive libraries. Clones showing a strong hybridization signal with the probes originating from the resistant samples and weak or no hybridization with those from the susceptible samples were selected and called DSS for Differentially Screened Sequences. Differential expression of DSSs during the coffee HR was confirmed by RT-PCR analyses [36]. Tested DSSs clearly showed an enhanced transcript accumulation

in inoculated plants over the time-course experiment as compared with the control plants. In addition, several of them showed induction during the incompatible interaction when comparing with the compatible interaction (Figure 3). Up-regulation of most of the genes occurred around 24 hpi. These results correlate with cytological observations of the coffee rust resistance reaction [34, 44] and suggest that induction of defence responses in cell leaves occurs early after penetration of fungal hyphae into the substomatal chamber.

The majority of the DSSs belonged to the defence, signalling or gene expression categories. DSS12 and DSS13 best matched the *A. thaliana dnd1* and *ndr1* genes. The DND1 protein is a cyclic nucleotide-gated ion channel (AtCNGC2) involved in the HR signalling pathway to *P. syringae* [42]. The NDR1 protein is a key component of the signalling pathway of many CC-NBS-LRR resistance proteins [41]. DSS16 and DSS17 putatively encoded an AP2-type and a WRKY transcription factor. A number of studies have shown the implication of several transcription factors in potentiating the plant responses to pathogen infection [45]. Particularly involved are several WRKY proteins which genes may be rapidly induced by pathogens or treatment with salicylic acid [46, 47, 48].

Although far to be exhaustive, the ESTs reported here may provide a significant set of data for improving our knowledge of coffee resistance to rust. With the availability of high-density cDNA filters technology, the expression profiles of hundreds ESTs will be monitored simultaneously in several coffee/rust interactions to help determine the mechanisms of these biological processes.

2. Coffee root-knot nematode resistance

In the main coffee-growing areas (Latin and Central America), root-knot nematodes (*Meloidogyne* sp.) are the most serious and damaging pest facing coffee production. More than 17 *Meloidogyne* species have been described on coffee but only a limited number is commonly observed (i.e. *M. exigua*, *M. incognita*, *M. arabicida* and *M. paranaensis*) [49]. Some specific nematode-resistant coffee cultivars have recently been engineered by conventional plant breeding based on the introgression of major resistance genes from wild coffee germplasms (*C. canephora*) into the cultivated varieties genomes (*C. arabica*) [50]. Resistance to *M. exigua* is controlled by a single dominant gene called *Mex1* [41] whose phenotypic expression is a HR [51]. When challenged with *M. exigua*, resistant coffee root tips exhibit the characteristic features of localized cell death, similar to those already described in other plant/nematode interactions [52, 53].

Choice of candidates from the coffee/rust EST libraries

Compared to other plant/pathogen interactions (fungal or bacterial), knowledge of mechanisms involve in resistance of woody plants to nematode

remains sketchy. The rare data available in the literature were obtained on *Mi*-resistant tomato since *Mi* is the only resistance gene to root-knot nematodes cloned up to now [54]. They result from the analysis of mutants affected in resistance [55] or from cDNA libraries differential screening [56].

As a consequence, the EST and specific genes (DSS) identified for the coffee/rust interaction represent a pool of potentially interesting candidates to be tested in the nematode interaction. In addition, they offer the opportunity to compare the implication of some genes in the defence response against two different pathogens on a same plant species. Some clones were then chosen based on their homologies with known defence genes and their expression was investigated by semi-quantitative RT-PCR.

Differential expression of selected genes upon nematode infection time-course

The RT-PCR experiments were performed on cDNAs obtained from inoculated and non-inoculated coffee root tips (0.5 cm section) of both susceptible and resistant varieties upon a *M. exigua* infection time-course (2 to 7 dpi). The four time points were conditioned by the results of cytological studies pinpointing the main stages of the HR [51].

The susceptible and resistant cultivars used in this study are both *C. arabica* varieties. Although they display a very low genetic diversity [58], they are not isogenic for the resistance gene *Mex1*. Therefore, differences in the plant genetic backgrounds (in addition to *Mex1*) may explain variations in the basal expression levels of the tested genes in the non-inoculated plants. The expression patterns of the chosen genes are showed in Figure 4. Only the genes displaying a differential expression are presented.

Expression profiles of the selected genes clusterized into 2 groups.

The EST clones, encoding respectively a chitinase, a LTP and a miraculin (LeMir), displayed a somewhat different profile between the susceptible and the resistant hosts. Their RT-PCR patterns in the inoculated susceptible plants showed a constant and linear decrease of the transcript accumulation from 2 to 5 dpi compared to the control, with an assumed return to the basal level by 7 dpi. On the opposite, the infected resistant roots expression pattern transiently peaked at the 3 dpi time point and dropped back to the non-inoculated level by 5 dpi.

Clones 1.1/4B and 16/10 encode a lipid transfer protein (LTP) and a chitinase, respectively. LTP are small, basic cystein-rich proteins proposed to have antimicrobial activities and to be involved in plant defence mechanisms. For instance, they were shown to be implicated in pepper (*Capsicum* sp.) resistance to tobacco mosaic virus [58] and grape (*Vitis* sp.) response to fungal elicitor treatments [59]. Chitinases are part of the PR proteins. They are induced by stress factors (mainly upon infections) and some isoforms show antifungal properties in *in vitro* assays. They play a role in the early stages of

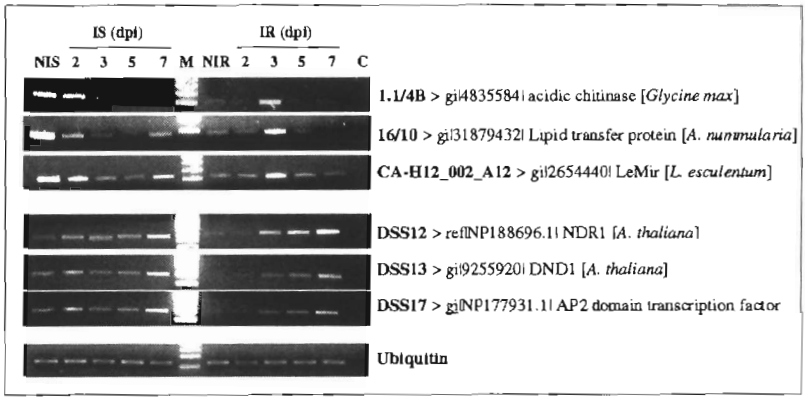


Figure 4. RT-PCR expression patterns of selected genes in *C. arabica* resistant and susceptible roots during an infection time-course with *M. exigua* (from 2 to 7 dpi). NIS: non-inoculated susceptible roots; IS: inoculated susceptible; NIR: non-inoculated resistant; IR: inoculated resistant; C: PCR negative control, M: molecular weight marker. Clone designation and best Blast homology are indicated at the right of each panel. Ubiquitin cDNA was used as internal control.

pathogenesis by releasing elicitor molecules, involved in the transfer of information about the infection [60].

CA-H12_002_A12 encodes the homologue of the tomato protein LeMir (for *L. esculentum* miraculin) shown to be induced specifically on tomato root tips tissues early after infection by *M. incognita* [61]. This protein is secreted within root exudates but its function in resistance to nematode still remains unclear. Nevertheless, its specific transient induction between 2 and 3 dpi may be linked to the penetration/migration stages of the nematodes in the intercellular spaces of the cortical parenchyma and triggering of the early events in HR cell death.

The RT-PCR profiles obtained for the DSS 12, 13 and 17 are less drastically contrasted than the EST clones ones and showed an enhanced expression in the inoculated plants compared to controls, at least at the latest time points of the infection process. The main difference between nematode resistant and susceptible plants lied in the timing and extent of the induction. Transcript accumulation in the susceptible coffee plant, challenged or not with *M. exigua*, remained constant or slightly enhanced from the control to the 5 dpi time point. A faint induction was visible around 7 dpi, whereas under similar conditions, the induction appeared after the 2 dpi time in the resistant plant and continued through the end of the experimental period.

These results are in line with data obtained with the coffee/rust interaction suggesting that DSS12, 13 and 17 encoding NDR1- and DND1-homologs

respectively, and putative AP2 transcription factor, may be key components of the HR regulation in *C. arabica*. To a larger extend, they are consistent with those reported in the literature concerning the response to fungal, bacterial or viral pathogens [62, 63, 64] reinforcing the idea of the occurrence of common points driving the HR process within plant species.

Conclusions

Cotton and coffee defence responses elicited by *Xcm*, the rust fungus and, nematodes respectively were explored by two different strategies (i.e. candidate gene and global approaches) and genes preferentially induced during HR were identified. Their characterization supplies useful data to a better understanding of resistance mechanisms in tropical plants and provides new insights into woody plants responses to biotic stresses.

Some of the identified genes show homologies with well-known components of *R*-mediated resistance in *A. thaliana* and required as regulators of basal defence in several model plants. Although their ubiquitous existence in different plant species suggests the maintenance of common signalling pathways, the role of these effectors in cotton and coffee resistance remains to be validated. So far, current investigations focus on functional analyses performed by *Agrobacterium*-mediated assays.

References

1. Van der Vossen, H.A.M. 2001, Coffee, recent developments, Clarke and Vitzhun (Eds), 184.
2. Campos, V.P., Sivapalan, P., and Gnanapragasam, N.C. 1990, Plant-parasitic nematodes in subtropical and tropical agriculture, M. Luc, RA. Sikora and J. Bridge (Eds), CAB International, 113.
3. Gabriel, D.W., Burges, A., and Lazo, G.R. 1986, Proc. Natl. Acad. Sci. USA, 83, 6415.
4. Noir, S., Anthony, F., Bertrand, B., Combes, M-C., and Lashermes, P. 2002, Plant Pathol., 52, 97.
5. Prakash, N.S., Marques, D.V., Varzea, V.M.P.,. Silva, M.C., Combes, M.C., and Lashermes, P. 2004, Theor. Appl. Genet., 109, 1311.
6. Rodrigues, C.J., Brettencourt A.J., and Rijo, L. 1975, Annu. Rev. Phytopathol., 13, 49.
7. Flor, H.H. 1947, Phytopathology, 45, 680.
8. Heath, M. C. 2000, Plant Mol. Biol., 44, 334.
9. Morel, J.B., and Dangl, J.L. 1997, Cell Death Differ., 4, 671.
10. Pontier, D., Balague, C., and Roby, D. 1998, C. R. Acad. Sci. Paris, 321, 721.
11. Cooper, B., and Clarke, J.D. 2003, Proc. Natl. Acad Sci. USA, 100, 4945.
12. De Torres, M., and Sanchez, P. 2003, Plant J., 33, 665.
13. Narusaka, Y., Narusaka, M. 2003, Plant Cell Physiol., 44, 377.
14. De Feyter, R., Yang, Y., and Gabriel, D.W. 1993, Mol. Plant-Microbe Interact., 6, 225.
15. Innes, N. 1983, Biol. Rev., 58, 157.

16. Nicole, M., Daniel, J.F., Bresson, E., Martinez, C., El Bachir, O., Lopez, F., Assigbetsé, K., Fernandez, D. Montillet, J.L., and Geiger J.P. 1998, Recent Res. Dev. Microbiol., 2, 641.
17. Rustérucci, C., Montillet, J.L., Agnel, J.P., Battesti, C., Alonso, B., Knoll, Bessoule, J.J, Etienne, P., Suty, L., Blein, J.P., and Triantaphylidès, C. 1999, J. Biol. Chem., 274, 36446.
18. Cason, E.T., Richardson, P.E., Essenberg, M.K., Brinkerhoff, L.A., Johnson, W.M., and Venere, R.J. 1978, Phytopahology, 68, 1015.
19. Daï, G.H., Nicole, M., Martinez, C., Bresson, E., Daniel, J.F., Andary, C., and Geiger, J.P. 1996, Physiol. Mol. Plant Pathol., 49; 285.
20. Jalloul, A., Montillet, J.L., Assigbetsé, K., Agnel, J.P., Delannoy, E., Daniel, J.F., Marmey, P., Geiger, J.P., and Nicole, M. 2002, Plant J., 32, 1.
21. Delannoy, E., Marmey, P., Penel, C., and Nicole, M. 2004, Acta Bot. Gal., 151, 353.
22. Chittoor, J.M., Leach, J.E., and White, E.F. 1999, Pathogenesis-Related Proteins in Plants, S. K. Datta and S. K. Muthukrishnan (Eds), CRS Press, New–York, 171.
23. Bolwell, G.P., Bindschedler, L.V., Blee, K.A., Butt, V.S., Davies, D.R., and Gardner, S.L., Gerrish, C., Minibayeva, F. 2002, J. Exp. Bot., 53, 1367.
24. Bolwell, P., Mansfield, J., Bestwick, C., Nicole, M., and Wojtaszek, P. 1999, Trends Plant Sci., 4, 89.
25. Wojtasjeck, P. 1997, Biochem. J., 322, 681.
26. Martinez, C., Montillet, J.L., Bresson, E., Agnel, J.P., Daniel, J.F., Geiger, J.P., and Nicole, M. 1998, Mol. Plant-Microbe Interact., 11, 1038.
27. Delannoy, E., Jalloul, A., Assigbetsé, K., Marmey P., Geiger, J.P., Lherminier J., Daniel, J.F., Martinez C., and Nicole, M. 2003 , Mol. Plant-Microbe Interact., 16, 1030.
28. Martinez, M., Baccou, J.C., Bresson, E., Baissac, Y., Daniel, J., Jalloul, A., Montillet, J.L., Geiger, J.P., Assigbetse, K., and Nicole, M. 2000, Plant Physiol., 122, 757.
29. Tognolli, M., Penel, C., Greppin, H., and Simon, P. 2002, Gene, 288, 129.
30. Welinder, K.G., Justesen, A.F., Kjaergard, I.V.H., Jensen, R.B., Rasmussen, S.K., Jespersen, H.M., and Duroux, L. 2002, Eur. J. Biochem., 269, 6063.
31. Yamasaki, H., Sakihama, Y., and Ikehara, N. 1997, Plant Physiol., 115, 1405.
32. Bettencourt, A.J., and Rodrigues, C.J. Jr. 1988, Principles and practice of coffee breeding for resistance to rust and other disease, R.J. Clarke and R. Macrae (Eds.), Elsevier Applied Science, London, 199.
33. Rodrigues, Jr. C.J., Varzea, V., Godinho, I.L., Palma, S., and Rato, R.C. 1993, Proceedings of the XV Scientific Colloquium on Coffee, ASIC, Montpellier, France, ASIC (Eds.), Paris, France.
34. Silva, M.C., Nicole, M., Guimares, L., and Rodrigues, Jr. C.J. 2002, Physiol. Mol. Plant Pathol., 60, 169.
35. Diatchenko, L., Lau, Y.F.C., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D., and Siebert, D. 1996, Proc. Natl. Acad. Sci. USA, 93, 6025.
36. Fernandez, D., Santos, P., Agostini, C., Bon, M.C., Petitot, A.S., Silva, M.C., Guerra-Guimares, L., Ribeiro, A., Argout, X., and Nicole, M. 2004, Mol. Plant Pathol., 5, 527.
37. Maleck, K., Levine, A., Eulgem, T., Morgan, A., Schmid, J., Lawton, K.A., Dangl, J.L., and Dietrich, R.A. 2000, Nat. Genet., 26, 403.

38. Scheideler, M., Schlaich, N.L., Fellenberg, K., Beissbarth, T., Hauser, N.C., Vingron, M., Slusarenko, A.J., and Hoheisel, J.D. 2002, J. Biol. Chem., 277, 10555.
39. Schenk, P.M., Kazan, K., Wilson, I., Anderson, J.P., Richmond, T., Somerville, S.C., and Manners, J.M. 2000, Proc. Natl. Acad Sci. USA, 97, 11655.
40. Tao, Y., Xie, Z., Chen, W., Glazebrook, J., Chang, H.S., Han, B., Zhu, T., Zou, G., and Katagiri, F. 2003, Plant Cell, 15, 317.
41. Century, K.S., Shapiro, A.D., Repetti, P.P., Dahlbeck, D., Holub, E., and Staskawicz, B.J. 1997, Science, 278, 1963.
42. Clough, S.J., Fengler, K.A., Yu, I.C., Lippok, B., Smith, R.K., and Bent, A.F., Proc. Natl. Acad Sci. USA, 97, 9323.
43. White, O., and Kerlavage, A.R. 1996, Methods Enzymol., 266, 27.
44. Silva, M.C., Nicole, M., Rijo, L., Geiger, J-P., and Rodrigues, Jr. C.J. 1999, Int. J. Plant Sci., 160, 79.
45. Singh, K.B., Foley, R.C., and Onate-Sanchez, L. 2002, Curr. Opin. Plant Biol., 5, 430.
46. Dong, J., Chen, C., and Chen, Z. 2003, Plant Mol. Biol., 51, 21.
47. Turck, F., Zhou, A., and Somssich, I.E. 2004, Plant Cell, 16, 2573.
48. Ulker, B., and Somssich, I.E. 2004, Curr. Opin. Plant Biol., 7, 491.
49. Carneiro, R.M.D.G., Carneiro, R.G., Abrantesn I.M.O., Santos, M.S.N., and Almeida, M.R.A. 1996, J. Nematol., 28, 177.
50. Bertrand, B., Anthony, F., and Lashermes, P. 2001, Plant Pathol., 50, 637.
51. Anthony, F., Topart, P., Martinez, A., Silva, M., and Nicole, M. 2005, Plant Pathol., 54, 476.
52. Bleve-Zacheo, T., Bongiovanni, M., Melillo, M.T., and Castagnone-Sereno, P. 1998, Plant Sci., 133, 79.
53. Williamson, V.M., and Hussey, R.S. 1996, Plant Cell, 8, 1735.
54. Milligan, S.B., Bodeau, J., Yaghoobi, J., Kaloshian, I., Zabel, P., and Williamson, V.M. 1998, Plant Cell, 10, 1307.
55. Martinez de Ilarduya, O., Moore, A.E., and Kaloshian, I. 2001, Plant J., 27, 417.
56. Lambert, K.N., Ferrie, B.J., Nombela, G., Brenner, E.D., and Williamson, V.M. 1999, Physiol. Mol. Plant Pathol., 55, 341.
57. Anthony, F., Combes, M-C., Astorga, C., Bertrand, B., Graziosi, G., and Lashermes, P. 2002, Theor. Appl. Genet., 104, 894.
58. Park, C.J., Shin, R., Park, J.M., Lee, G.J., You, J.S., and Paek, K.H. 2002, Plant Mol. Biol., 1, 243.
59. Gomes, E., Sagot, E., Gaillard, C., Laquitaine, L., Poinssot, B., Sanejouand, Y.H., Delrot, S., and Coutos-Thevenot, P. 2003, Mol. Plant-Microbe Interact., 16, 456.
60. Kasprzewska, A. 2003, Cell Mol. Biol. Lett. 2003, 8, 809.
61. Brenner, D., Lambert, K., Kaloshian, I., and Williamson, V. 1998, Plant Physiol., 118, 237.
62. Glazebrook, J. 2001, Curr. Opin. Plant Biol., 4, 301.
63. Shapiro, A.D., and Zhang, C. 2001, Plant Physiol., 127, 1089.
64. Yu, I., Fengler, K.A., Clough, S.J., and Bent, A.F. 2000, Mol Plant-Microbe Interact., 13, 277.

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

Molecular Biology of Tropical Plants, 2006: 17-39 ISBN: 81-308-0138-8
Editor: Claudine Franche

2 Rice and virus biotechnology: Application for the expression of an anti-*leishmania* vaccine

Siré, C.[1], Piron, F.[1], Bangratz-Reyser, M.[1], Lemesre, J-L.[2] and Brugidou, C.[1]
[1]UMR 5096-Génome et Développement des Plantes and [2]R008-Pathogénie de Trypanosomatidés, Institut de Recherche pour le Développement, 911 avenue Agropolis, BP 64501, 34394 Montpellier, Cedex 5, France

Abstract

Plant systems are showing a considerable potential for the economic production of proteins. Recent experiments of biopharmaceutical production from transient expression in plants are encouraged. We discuss here a general strategy using RYMV-based vectors (Rice yellow mottle virus) to produce an anti-leishmania vaccine in monocot (Oryza sativa L.) and dicot (Nicotiana tabacum or Nicotiana benthamiana) plant species.

Correspondence/Reprint request: Dr. Christophe Brugidou, UMR 5096-Génome et Développement des Plantes Institut de Recherche pour le Développement, 911 avenue Agropolis, BP 64501, 34394 Montpellier, Cedex 5 France. E-mail: christophe.brugidou@mpl.ird.fr

Introduction

Nowadays, many tools are available for production of heterologous proteins, such as bacteria, fungi, mammalian and insect cell systems. Nevertheless, high level of natively folded proteins is limited. Other strategies have been thus exploited to overcome such constraint.

Over the last decade, researches on plants, especially on plant models (i.e. *Arabidopsis thaliana* for dicot species and *Oryza sativa* for monocot species), lead to increase significantly our knowledge of gene regulation and protein synthesis in different plants. Similar to evolution of genomic, new biotechnological tools, with transgenic technologies, were developed to express foreign genes. However, such technologies were exploited not only to improve agronomic performances but also to achieve rapid high level production of valuable proteins (e.g. industrial or pharmaceutical products).

Thus for molecular farming, plants were described as attracting, safe and practical bioreactors.

Nevertheless, technical limitations also existed in such technologies, thus much efforts have been concentrated to overcome drawbacks inherent to transgenic technology. At present day, molecular farming in plants can be achieved by stable or transient expression systems. Moreover, concomitant to genomic studies, research on plant viruses lead to the discovery of many potential of such entities to help the production of heterologous proteins. The potential of plant viruses as tools for genetic engineering, was highlighted, especially as expression vectors for production of proteins in plants.

More recently, the discovery of viral suppressors of gene silencing enhanced knowledge and advances in plant valuable proteins production. Indeed such proteins were exploited to avoid the problem of transgene silencing inherent to transgenic technologies but also to over-expression of foreign genes also in transient systems.

As a model system for cereal genomic and biotechnology, rice could be widely used to develop such strategies in monocotyledon species. In this review, in a first part, we'll discuss about the tools available for rice genomic and biotechnology. Then, in a second part, we'll develop all the strategies described, with their advantages and drawbacks, to widely produce heterologous proteins especially in rice, with the specific aim to produce a therapeutical molecule.

I. Rice and RYMV, tools for functional genomics and biotechnology

I.1. Rice as a model for cereals

Rice belongs to the *Oryza* genus, including 20 species, of which only two are cultivated: *O. glaberrimma* Steud, endemic to Africa and *O. sativa* L.

originates from Asia. *O. sativa* comprises two groups of cultivars analogous of sub-species (*japonica* and *indica*). Asian cultivated rice (*O. sativa*), worldwide cultivated, is an important subsistence crop in tropical regions, for human consumption, providing the staple food for more than a half of the world's population.

Its importance is not only reported at the economic scale but also at the genetic scale. Indeed, rice became a plant model for monocotyledons genomic, especially for cereals because of its genetics features. Actually, rice exhibits a high synteny with the other cereals [4,5] and its genome size is relatively small (i.e. 430 Mbp spread over 12 chromosomes). Thus, many tools have been developed to understand this species better with developing dense molecular genetics maps, YAC[1] and BAC[2] libraries [6,7]. Moreover, improved transformation techniques are now available, with biolistic delivery or explants co-culture with *Agrobacterium tumefaciens*, providing an efficient tool not only for crop improvement but also for functional genomics [8,9]. In this context, essential biological information from the rice genome will be easily assessed and will especially improve our understanding of the basic genomics and genetics of other related significant crops. In this way, a large number of genomic sequences have been generated by sequencing the entire genome that holds fundamental information for its biology, including physiology, genetics, development and evolution [5,10,11]. A large number of studies have been carried out to generate sequence analysis and also to determine complete genomic sequence [12]. Thus, completion of rice sequencing was achieved in 2003 and its genome was automatically annotated, using prediction-based and homology-based searches to identify genes [13]. Moreover several rice genome sequences have been produced, thus providing a unprecedented access to numerous genes [5]. Nevertheless, knowing DNA sequence is the first step for the elucidation of the genomes biology and efficient transformation methods contributed to the deployment of new tools for improving and studying gene function. A large number of tools have been developed to facilitate gene function discovery. Up to date, insertional mutagenesis has been widely used, in this aim [14-16]. However, a complementary method has been undertaken to identify function of genes that have not already been tagged with mutants: RNAi vectors [17]. Such vectors lead to assessment of gene function by suppressing gene expression through specific RNA-mediated RNA degradation mechanism, and can be used either in *Agrobacterium* transformation assays or in transient expression systems. Furthermore, dicotyledonous species, another original technology was used to appreciate rapidly gene function by transitively over-expressing or suppressing gene

[1] Yeast artificial chromosome
[2] Bacterial artificial chromosome

expression, with virus-based vectors. Indeed such technology has been improved for various applications in dicot, it's lacking for efficient use in monocot species.

I.2. Viral vectors: Functional genomics and biotechnological tools

During the last decade, the potential of plant virus-based vectors has been highlighted for functional genomics and study of gene expression. The study of plant viruses has generally permitted the assessment of eucaryotic biology [18,19]. Viral expression systems represent an attractive tool to complement conventional breeding or transgenic methodology. To achieve the expression of heterologous sequences in plants, the use of viral expression systems offers several advantages over stable transgenic expression particularly in regard to the expression levels that are obtained and genes can also be evaluated earlier [20,21].

Molecular genetic studies of plant RNA viruses lead to the generation of infectious RNA (i.e. upon *in vitro* transcription) [22] or directly infectious cDNAs (i.e. downstream from *Cauliflower mosaic virus* 35S promoter). Moreover, the emergence of the green fluorescent protein (gfp) from *Aequorea victoria* as a reporter molecule offers the possibility to assess plant virus infections with a non destructive assay technique [23-26]. Thus, viral infectious clones have been modified to study better different viruses but also viral protein functions by monitoring movement in infected cells, replication and virus spread [27]. Then flexibility of viruses and rapid expression of viral proteins have been exploited for fundamental research or biotechnology application [20,21,28] to produce high level of foreign genes with the development of viral expression vectors. Actually, virus-based expression vectors have number of advantages as gene expression tools including the ability to direct rapid and high level expression of foreign genes in mature and differentiated plant tissues, and have been used for a number of different applications [29,30]. In fundamental virology, fusion with reporter genes (gus or gfp) allows monitoring of viral gene expression and products *in planta* [31,32]. They are also used for production of valuable foreign peptides and proteins in plants [21,29]. Finally, virus-based expression vectors offer advantages to express foreign (or endogenous) genes for functional characterisation of ORFs (open reading frames) [29,33]. The ability of RNA viruses to trigger gene silencing has also been exploited in the construction of VIGS (virus induced gene silencing) vectors to suppress host gene expression in the aim to assign gene function [34,35]. Such vectors carry sequences that share homology with transgenes or endogenous genes and silencing might be initiated by the viral sequence whereas the maintenance step occurred on the nuclear genes targeted independently from the presence of

the virus [36]. VIGS has been validated as an efficient tool for reverse genetic to study genes involved in primary and secondary metabolism, in development, disease resistance [37]. Thus VIGS strategy is an attractive alternative to insertional mutagenesis to investigate gene function particularly in multigene families [38].

Some basis features are required for viruses used to develop plant virus-based vectors. Indeed, they need to i) be autonomous replicating systems, ii) be easily genetically manipulated, iii) have a short cycle time life and finally iv) infectious cDNA clones must be available. Many strategies have been developed for transient expression of foreign genes and two of them were mainly used [30] (figure 1).

The first method consists in the fusion of the protein, or protein domains, with the viral coat protein (CP) for presentation at the surface of the viral particle such as epitope presentation [39-41]. This strategy has been widely exploited for production of vaccine sub-unit [30]. The second method consist in expressing foreign genes and viral genome independently of one another by using duplication of sub-genomic mRNA promoter, gene insertion or gene replacement techniques [20,21,30]. However, whichever the strategy used the insertion of foreign genes can interfere with CP function (e.g. particle assembly, virus movement) and steric constraint limited the size of the inserted sequence in most cases. Such vectors have been improved to preserve viral genome integrity and to promote insert stability with different inoculation methods or gene trans-complementation, for example [42-44].

Transient knock-out strategy through VIGS vectors have also been improved with the assessment of insert features such as insert size [45], insert orientation or the choice of gene fragment. Furthermore viruses have evolved to develop silencing suppression properties [46], and viruses possessing strong silencing suppressors are not suitable tools for the construction of VIGS vectors. Finally in some cases strong VIGS response or symptoms apparition interfering with the assessment of gene function [47] need to be get round.

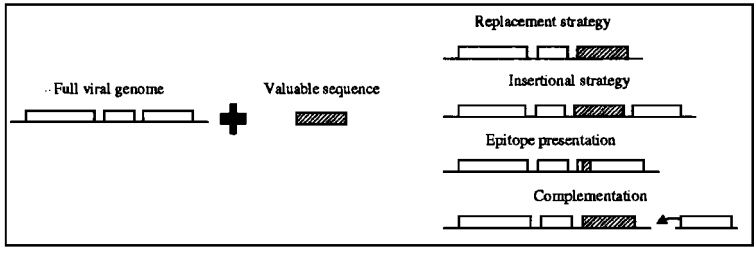


Figure 1. Different strategies to produce virus-based vectors for valuable proteins production.

There are now several plant viruses which have been converted into vectors and implemented for transient over-expression of genes of interest, and also, for VIGS of transgenes, or endogenous genes. Viruses of dicot species have been developed as vectors, but none are functional for cereals. Moreover there are no suitable vector for expression of foreign genes in monocotyledonous plants except a BMV vector [48], thus our lab is interested in developing such technology for rice based on the genome of *Rice yellow mottle virus* (RYMV) (figure 2).

Rice yellow mottle virus is a single-stranded-positive-sense RNA virus that specifically infects rice leaves and causes serious disease in irrigated rice systems in East and West Africa [49-51]. This virus, belonging to the *Sobemovirus* genus, is transmitted by chrysomelid beetles and can be artificially inoculated with sap. RYMV-genome properties or particle structure have been well described [2,3]. RYMV genome is simple and corresponds to a positive single-stranded RNA composed with four partially overlapping ORF [52] (figure 2).

Recently genetic diversity and phylogeny of the virus have been studied [52,53]. This virus thus, represents a good model for developing virus-based expression or VIGS vectors as:

i) infectious cDNA clones are available [54],
ii) RYMV highly replicates in infected cells,
iii) P1 protein involved in PTGS suppression is highly variable among different RYMV isolates [46,55],
iv) tolerant rice genotypes are available to bypass effects of the infection,
v) transgenic plant expressing viral ORF [56,57] are also available for trans-complementation of viral sequences deleted to overcome size constraint.

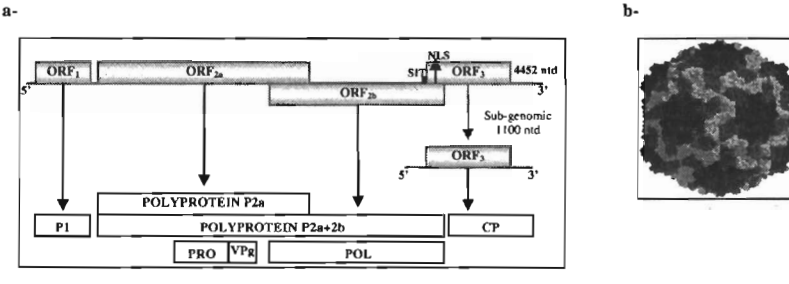


Figure 2. *Rice yellow mottle virus.* **a-** Genomic organisation [54] with SIT : transcription initiation site (GTGGGATAGGGCGAGTCTCCCACAAAGATG [1]), NLS : Nuclear localization signal $(KK(X)_{10}KRKXRR$ [2]), PRO : protease, VPg : *viral protein genome linked*, POL : RNA polymerase, CP : coat protein. **b-** Atomic structure of RYMV particle with a 2,8Å resolution [3].

Nevertheless, we have to keep in mind the major drawback, which consists in the size constraint set by the virus icosaedral particle. Actually, even if the virus genome is highly variable, the size of the sequence and also untranslated regions are conserved among the 15 isolates fully sequenced [55].

As genome size is crucial for icosaedral viruses (e.g. RYMV), to preserve virus infectifivity and integrity, we attempted to develop RYMV-based vectors, in our lab, to highlight maximal size constraint. Thus, interesting tools both for rice functional genomics and for biotechnological applications, with the aim of large-scale production of recombinant proteins, were developed. As, a major size constraint has been demonstrated (Siré *et al.*, unpublished results), other alternative strategies will be undertaken to overcome this drawback. Indeed, the use of replicative viral system, also called amplicon, has been previously reported, in combination with silencing suppressors, to reach high level of protein production [58]. With this aim in mind, RYMV-amplicon tool was developed. Furthermore, large studies have been carried out to better understand mechanism of silencing suppression by RYMV, to improve the amplicon tool for production of recombinant proteins in rice.

I.3. Gene silencing targets viruses and viruses suppress gene silencing: Application for biotechnological tools

Both viruses and genetic invasive elements trigger and target RNA silencing. This mechanism consists in a sequence-specific RNA degradation preventing gene expression.

RNA silencing was firstly reported as PTGS (post-transcriptional gene silencing) in plants and referred to RNAi (RNA interference) in animals, is generally conserved in eukaryote cells [59,60]. Thus RNA silencing consists in an ancient regulatory and adaptive defence mechanism acting at the molecular level against different genetic invasive elements.

In cells where they are detected, double-stranded RNA (dsRNA) molecules trigger PTGS, leading to a dramatic reduction of homologous cytoplasmic mRNA accumulation [61-63]. Both highly transcribed transgenes, transgenes in inverted repeat orientation, and viruses lead to production of dsRNA molecules in higher plants [64,65]. Then the RNAseIII-like enzyme, called DICER, targets dsRNA and cleaves this molecule into small RNA duplexes of 21 to 24 ntd, reported as small interfering RNA (siRNA) [66,67]. These siRNA are then complexed with a large multicoponent RNA-induced-silencing complex (RISC), which is thought to unwind siRNA to help target the appropriate mRNA. The antisense strand of siRNA is used to target homologous cytoplasmic mRNA, which is finally degraded [68,69]. Another step in silencing mechanism has been described as the amplification and propagation steps of the signal in distant tissues, involving RNA-dependent

RNA polymerase (RdRP) activity [70]. Thus newly synthesised dsRNA, with ssRNA (i.e. single-stranded mRNA) as template and siRNA as primer, lead to intense accumulation of siRNA that move in neighbouring cells through plamodesmata [71,72].

Both in higher plants and in animals, PTGS has been reported as an immune system acting at the molecular scale to prevent viruses invasion [73,74].

To counteract such defence mechanism, plant viruses evolved through specialisation of one, or more, of their proteins, to target various steps in silencing pathway (i.e. intercellular or intracellular silencing) [75,76], as a result of diversity and multifunctionnality of viral proteins [77,78]. Silencing suppression with plant-viral proteins, thus facilitate virus replication and movement [46,75,78,79]. Up to date many viral suppressors were identified, encoded by a single ORF or by different ORF for the same virus [80,81] and were generally described as involved in viral pathogenicity and in virus spread [46,79,82]. Features of such proteins were highlighted in biotechnological applications in the aim of enhancing recombinant protein production, which is drastically limited by silencing directed against over-expressed genes [83].

Thus, with the purpose to highly produce protein of interest in rice, with or without RYMV-based vectors, behaviour of P_1 protein from RYMV in silencing suppression, was assessed. Indeed, this protein has been previously described as non autonomous cell silencing suppressor [46,70] and is dispensable for viral replication but is closely related to virus infectivity and also to virus spread [84]. Availability of a large collection of RYMV-isolates allowed an accurate and original study of silencing suppression by entire RYMV particle, and also by its P1 protein (Siré *et al., submitted*). This study characterised silencing suppression features of RYMV under natural infection on rice plants. Biolistic delivery assays on rice leaves and *Agrobacterium*-based leaf infiltration assays on *Nicotiana benthamiana*, determined that different P1 proteins undergo silencing with variable efficiency. Our results also suggested, that silencing suppression occurring under RYMV infection, is a complex mechanism, probably involving more than one viral suppressor.

II. Strategies to produce recombinant proteins: Application for production of anti-*leishmania* vaccine

Many biotechnological applications, like production of vaccines, antibodies, human blood products, hormones and growth factors, require high-level expression of transgenes. In the past few decades, several different systems have been developed, for the production of recombinant proteins at

low cost. However, all these systems are not perfect because in some cases they can be unsafe, or they can lead to production of biologically inactive material. So, production of recombinant proteins in mammalian cells results in products that are identical to those of natural origin but this culture is very expensive and can be carried out on a limited scale. The use of microorganism system allows production on a large scale, but exhibits the major drawback to introduce structural variations in the protein. Plant system is an economical system, where the contamination risks, with human pathogens, are minimised. Furthermore, this system can be developed at an industrial scale. Finally, plant system offers a good compromise with an eukaryotic protein modification machinery allowing subcellular targeting, proper folding and post-translational modifications at low cost. Thus, today all advantages of plant system make it the most attractive technology for the production of recombinant proteins [85,86] such as an anti-*leishmania* vaccine.

Protozoa of the genus *Leishmania* are obligatory intracellular parasites of mammalian macrophages. They are transmitted to vertebrate hosts by sandfly vectors of the genus *Phlebotomus*. They cause a wide spectrum of human diseases in many tropical and subtropical regions of the world that range from a self-healing cutaneous ulcer to a potentially fatal visceral infection (figure 3). Zoonotic visceral leishmaniasis is one of the most important emerging diseases. Wild canids and domestic dogs are the main reservoirs of *L. infantum*

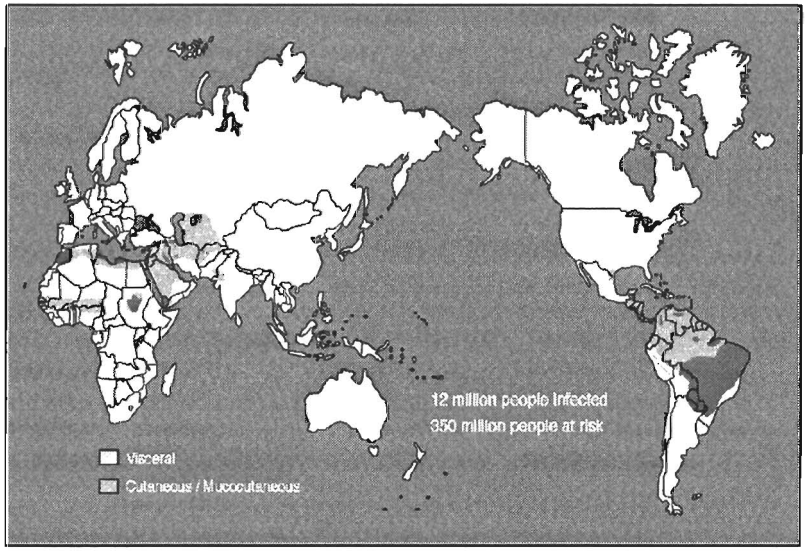


Figure 3. Geographical distribution of leishmaniasis in the world.

in the Mediterranean Basin, extended to several Middle-East and Asian countries, and of *L. chagasi* in South and Central America. Visceral leishmaniasis (VL) due to *L. donovani* is the most severe form of leishmaniasis. Approximately 500,000 new cases of human VL occur annually and the disease is mainly found in Brazil, East Africa and on the Indian sub-continent where devastating outbreaks have occurred and from where most VL cases are reported world wide [87]. In India, millions are at risk, the state of Bihar accounts for nearly 90% of cases, followed by West Bengal and Eastern Uttar Pradesh. Neighbouring countries like Nepal and Bangladesh also report a significant number of VL cases. Affected populations are among the poorest in the world and are not much aware/informed of existing preventive measures. Furthermore miss-use of the first-line drug in these communities is widely spread [88] and the lack of response to the first line drug (pentavalent antimonials) has been increasing sharply these last years in India up to more than 50% of the patients in hyper endemic areas of Bihar [89-91].

There is now an urgent need for new low cost drugs and/or new therapeutic interventions such as a vaccine for the control of this parasitic disease.

Recently, we have developed a vaccine involving *Leishmania* Excreted Secreted Antigen (LESA) [92-94] which was proved efficient both experimentally and in naturally *Leishmania infantum* infected dogs of southern France [95]. Access to a serum-free system for culturing promastigotes of *Leishmania* has improved the feasibility of large-scale production of well-defined parasite material. Using this methodology, it has been possible to easily purify naturally excreted secreted antigens from culture supernatant of *L. infantum* promastigotes successfully cultivated in a completely defined CDM/LP medium [92-94] and to investigate their biochemical properties. LESA only contained few excreted secreted polypeptides and mainly concentrated a major immunogenic protein belonging to the Promastigote Surface Antigen (PSA) family (figure 4).

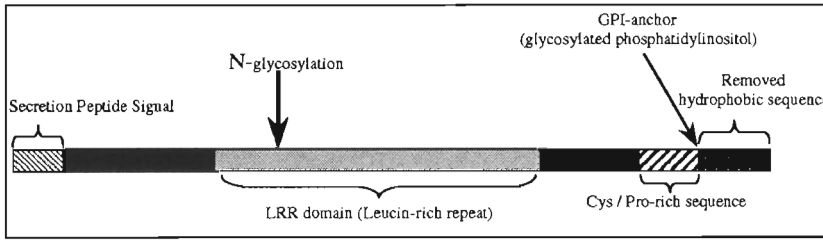


Figure 4. Diagrammatic representation of the structural features of Promastigote Surface Antigen (PSA) of *Leishmania*.

Vaccination with native promastigote surface antigen 2 (PSA-2) of *Leishmania* major has been demonstrated to protect mice from leishmaniasis through a Th1 mediated response. But recombinant PSA-2 purified from *E. coli* was unable to induce protective immunity [96]. These results have important implications for the design of vaccines against leishmaniasis. They strongly suggest that the generation of protective immunity is dependent not only of the induction of a Th1 response, but also indicate that some *Leishmania* antigens may require a near native conformation to be protective [97]. Correct posttranslational modifications and protein folding of antigens may therefore be important not only for the induction of neutralizing antibodies but also for the development of protective CD4+ T cell responses. Finally, the conformation of antigen may play a more major role for the induction of T cell mediated immunity than originally considered.

Thus, expression of the PSA is a good target to test our production system and to demonstrate that is possible to over-express recombinant proteins with the right folding and a good biological activity in cereals.

II.1. Strategies

Up to date, there is not yet a consensus for the best plant species, or tissue, for large-scale recombinant protein production. However, it would be preferable to choose a plant whose genetic manipulation is relatively easy with a large seed production [85]. Then, tobacco seems to represent suitable host for recombinant protein production in sufficient quantities. However, tobacco is not the only plant species used for biopharmaceuticals production. Thus, most antibodies expressed to date have been produced also in potatoes, soybean, alfalfa, rice and wheat. For example, a single-chain Fv antibody (ScFvT84.66) against carcinoembryogenic antigen (CEA) was successfully expressed in the cereal crops rice and wheat [98].

In our study, we decided to focus on two species: rice (*Oryza sativa* L.) and tobacco (*Nicotiana benthamiana* and *Nicotiana tabacum*).

There are currently two methods for protein production from plants: stable transformation and transient transformation. To date, the most common of the methods, stable transformation, has produced all the products available in the marketplace. This system requires a method for transferring the foreign genes into the plant cells, usually using *Agrobacterium tumefaciens* or particle bombardment, in which the genes are taken up and incorporated into the host nuclear genome in a stable manner [99]. This method of transformation presents advantages when performed in a crop species such as grains. Then, the protein product is normally accumulated in seeds that can allow the protein to exist without degradation for at least two years [99].

Nevertheless, regenerating transgenic plants from transformed cells is both labour intensive and time consuming. Moreover, after this hard work,

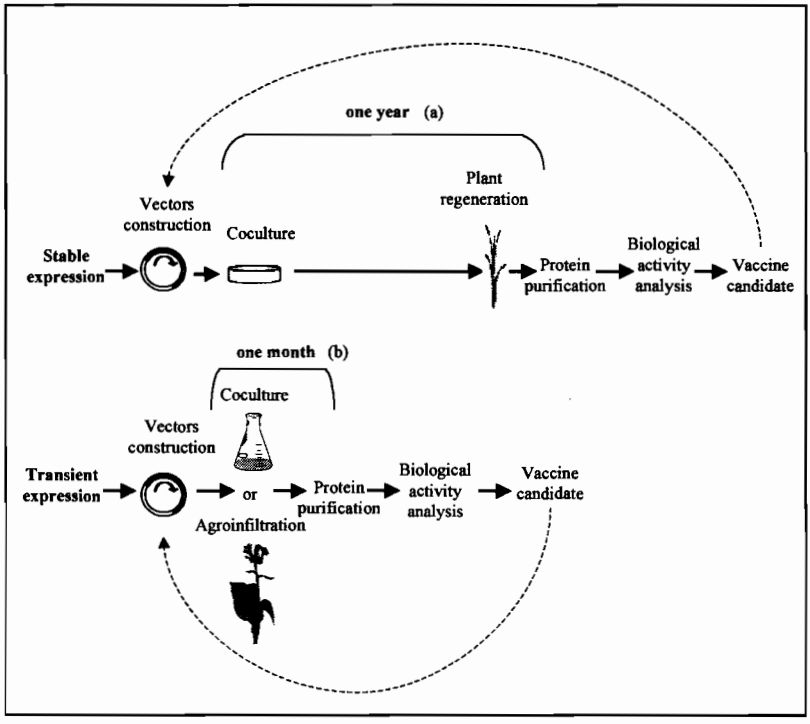


Figure 5. Comparison of time consuming between stable and transient expression systems. (a) Constitutive system is labour intensive and time consuming. To purify recombinant proteins, one year is necessary from the calus coculture to the second-generation plants. (b) Whereas transient expression is a flexible and rapid system for the production of proteins. Only one month is necessary to collect the recombinant proteins with a transient system. This is ideal for verifying functionality, integrity and stability of gene product before large scale developing stable transformed plants.

production of a correct folding protein and a biologically active protein is not guaranteed (Figure 5).

In contrast, transient protein expression using virus-based vectors allows rapid production of recombinant proteins. Thus, this flexible and changeable system is ideal to analyse the structure and the activity of protein produced in plants before developing a constitutive expression system. Transitory system is suitable not only for testing and scoring plant expression constructs and predicting their performance in transgenic plants, but also for purifying the recombinant protein [100]. It was demonstrated that the tobacco agroinfiltration take to the production of functional recombinant proteins [100].

Expression level of recombinant proteins can be variable. Avidin (i.e. immunological regeant) at 3% of extractable protein was produced in transgenic maize seeds [101] whereas only 0.1% of aprotinin [102] and 0.5% of β-glucuronidase [103] was generated with the same plant species. Apoprotinin, the bovine pancreatic trypsin inhibitor which affects known serine proteases such as trypsin, chymotrypsin, plasmin and kallikrein, has been widely used in biochemical research and as a therapeutic agent [102]. It has been hypothesised that the differences between proteins in terms of size, charge and localisation, could play a crucial role in the expression level [102].

Moreover, different levels of protein expression were published according to the use of different tools. Thus, transgenic plants (~0.5 mg/kg) revealed a lower yield of recombinant proteins than agroinfiltrated leaves (~1.5 mg/kg) due to a higher promoter activity and gene dosage during the transient expression [100].

II.2. Production enhancement

Recently, viral vectors have been generated from a large number of different viruses in the purpose to improve speed and yield of expression. In this system, viral vectors were designed to serve as over-expression tools.

RNA viruses can multiply to very high titres in infected plants, which makes them ideal vectors for protein expression. For vector construction, viral RNA genomes are reverse-transcribed *in vitro* and cloned as full-length cDNAs or only amplicon-cDNAs (replicating system of plant virus) in vectors [104]. The idea was that transcription of the amplicon and of the transgene, result in very high levels of the recombinant protein. *Nicotiana benthamiana* leaves, inoculated with *in vitro* transcripts of recombinant tobacco mosaic viral vector, accumulated recombinant protein to level of at least 2% of total soluble protein [105].

Transient expression method depends on the ability of recombinant plant viruses to infect plants and then transiently express a target protein in plant tissue [99]. Moreover, target genes are expressed at high levels consequently to the high level of virus replication [104]. The non-integrated T-DNA (Transferred-DNA) copies remain transitory present in the nucleus, that can be transcribed, leading to transient expression of the T-DNA genes [106]. Efficient of transient expression system has been shown by production and purification of His_6-tagged diabody from a scaled-up agroinfiltration tobacco leaves [100].

As plant viruses have a wide host range, the same vector construct is compatible for different plant species [107]. Plant virus vectors have the potential for becoming a useful tool to express foreign proteins in plants, especially when plant-specific folding and glycosylation of the recombinant proteins are of importance [107].

Then, to increase the level of anti-*leishmania* vaccine expression in tobacco and rice, we used an amplicon system based on the RYMV.

Study on a series of *Tobacco mosaic virus* (TMV)-based hybrid vectors for transient gene expression showed differences in the amounts of recombinant protein produced. These results demonstrated that building an effective vector from a virus is not a trivial exercise. An effective expression vector should contain a combination of *cis*-acting elements that appropriately partitions the limited replicase activity among the various promoters to ensure adequate replication and movement while providing the maximal level of foreign gene expression. It has been demonstrated that the most effective vector based on TMV contained sequences encoding the coat protein subgenomic mRNA promoter, coat protein ORF, and 3' UTR (untranslated region) from *Tobacco mild green mosaic virus* U5. Thus, the recombinant protein (i.e. GFP) accumulated up to 10% of total soluble protein in leaves [108].

However, gene expression in plants is influenced by posttranscriptional controls, known as posttranscriptional gene silencing (PTGS). It has been reported that plant infectious virus or transgene can induce gene silencing in absence of any known homology between viral genome or transgene and host genes [109,110]. Reduced levels of the specific mRNA encoded by the suppressed gene characterise the gene silencing phenotype. But, as a counterdefensive strategy, viruses have evolved proteins that suppress various steps of the RNA silencing mechanism.

In order to suppress PTGS induced by the transgene and to increase the transgene expression, the RYMV P1 and the TBSV (*Tomato bushy stunt virus*) p19 pathogenicity factors were co-infiltrated with the gene of interest. These two proteins have been identified as silencing suppressors of transgene [46]. It has been described two types of silencing suppressors having different actions [70]. Thus, the P1 protein could prevent systemic silencing but not its limited movement at the edge of infiltrated patches: it was the non-autonomous cell action. On the contrary, the TBSV p19 silencing suppressor have autonomous cell action. The effect of p19 was estimated to enhance 50-fold enhancement the abundance of the protein [83]. Our first studies on different RYMV P1 silencing suppressors showed that using the appropriate P1 silencing suppressor, level of expression protein could be strongly increase.

It was reported that high level expression could be achieved by pairing the amplicon approach with the use of a viral suppressor of PostTranscriptional Gene Silencing. Leaves co-expressing Hc-Pro from TEV (*Tobacco etch virus*) and a PVX (*Potato virus X*)/Gus amplicon accumulate GUS to about 3% of total proteins [58].

Free cell suspensions is generally regarded to be the best suitable for large-scale applications in the biotechnology industry. A number of plant species, like *Arabidopsis*, rice, soya bean, alfalfa and tobacco have been used for

generation and propagation of cell-suspension cultures. Moreover, plant-cell suspensions can be cultivated using conventional fermenter equipment. Large-scale fermentations up to a volume of 100 000 litres have been performed successfully [111]. The major advantages associated with *in vitro* plant systems include the ability to manipulate environmental conditions for better control over protein levels and quality, the rapidity of production compared with agriculture, and the use of simpler and cheaper downstream processing schemes for product recovery from the culture medium [112]. Using tobacco BY-2 cell line for fermentation, the cultivation of transgenic suspension cells was scaled-up to a working volume of 40 litres. With a 10% (v/v) inoculum, fermentation times of 150h resulted in a yield of 7.5kg of fresh cell weight, corresponding to 0.4kg dry weight [111].

Moreover, correct processing of protein was demonstrated in this system. For example, correct folding erythropoietin was produced in cultured tobacco BY2 (*Nicotiana tabacum* L. cv. Bright Yellow 2) by introducing human Epo cDNA *via Agrobacterium tumefaciens*-mediated gene transfer [113].

The plant-cell-suspension cultures exhibit the advantage that recombinant proteins can be produced under certified conditions (i.e. certified Good Manufacturing Practice and certified Good Laboratory Practice) [111].

In any systems of heterologuous production, the recombinant molecule must be extracted and purified selectively from total endogenous proteins. Eighty percent of the recombinant protein production cost correspond to this purification step [114].

In whole plant system, an alternative to reduce the expense consists in directing protein synthesis to seed endosperm [115], from where proteins may be easily extracted. Further studies showed that the recombinant protein (gB) behaves like a plant storage protein and is localised almost exclusively in protein storage vesicles, when expressed in tobacco seeds [115].

In the purpose to facilitate the purification of biologically active hirudin (an anticoagulant found to be an inhibitor of thrombin) in tobacco, an Arabidopsis oleosin promoter combined with a plant oleosin "carrier" was used [116]. The fusion protein was then targeted to the oil body membrane. This system was developed to simplify the initial step of purification and to limit the proteolysis [116]. After a correctly targeting to the oil body membrane, the recombinant protein was separated from the majority of other seed proteins by flotation centrifugation [116].

In plant-cell–suspension cultures, recombinant proteins expressed are either found in the culture supernatant or retained within the cells. This localisation depends on two factors: the presence of targeting/leader peptides in the recombinant protein, and permeability for macromolecules allowed by plant cell wall [111]. Targeting signals can be used to direct the protein for secretion or to intracellular organelles [111]. In this way, genetically

modified *Nicotiana tabacum* cells, grown in suspension culture, produced and secreted into the medium, a biologically active human interleukin-2 and interleukin-4 (IL-2 and IL-4). These two proteins were detected at concentrations of 0.10 and 0.18 µg/ml in the medium, respectively for IL-2 and IL-4 [117,118].

Plant suspension culture has be used to produce and secrete into the medium a variety of biologically active mammalian proteins that are clinical and diagnostic relevance [119]. Either human prepro-sequence or the extracellular tobacco protein PR-S were used to secrete the human serum albumin (HSA) in transgenic potato leaf tissues and in tobacco suspensions [118,119].

Moreover, the epitope tagging of expressed proteins is a versatile tool for the detection and purification of proteins [120]. In the purpose to purify the *E. coli* MutS, MutH and MutL (proteins mediating methyl-directed-mismtatch) proteins, genes were cloned into an expression vector, which allows fusion to the His6 affinity tag. These His6-proteins were then purified by variations of batch binding to Ni(2+)-chelation affinity resin. The yield of purified His6-proteins from these procedures was 0,4-0,6 mg from 40 mL of induced culture [121].

Transient expression represents a method for verifying functionality, integrity and stability of gene product before large scale developing stable transformed plants [104].

But, transient expression could be also used to determine the subcellular localisation of proteins. In this way, plasma membrane localisation of the ACBP2 (Cytosyl acyl-CoA-binding proteins) and both nucleus and plasma membrane localisation of AtEBP (*Arabidopsis thaliana* ethylene-responsive element-binding protein), were demonstrated using GFP autofluorescent protein fusions, in transient expression by agroinfiltration of tobacco leaves [122]. Transient expression by agroinfiltration is a powerfull tool for promoter studies. In this way, a preferential expression in vascular tissues of stems and leaves conferred by the promoter of a rice glycine-rich protein gene was highlighted [123].

Morever, transient expression could be an interesting system for functional analysis of different promoters or for identification of genes *via* functional complementation. For example, it can be exploited in sense-antisense systems, normal and mutated genes, or in studies related to disease resistance genes [106]. These studies demonstrated that interaction between host plant and pathogen is not disrupted by infiltration with *Agrobacterium tumefaciens* [106]. This suggests that the procedure can be used for studying plant/pathogen interactions.

II.3. Control of integrity of products

Transient expression can be used to verify protein activity before proceeding to transgenic plants.

The biochemical analyses of the on-step IMAC-purified protein (Immobilised metal ion affinity chromatography) showed tobacco cells expressed and correctly processed the T84.66/GS8 diabody, and preliminary data from mass spectrometry suggested that post-translational modifications did not occur. This study has also revealed only the presence of functional diabody purified from agro-infiltration tobacco leaves.

In therapeutical application, the use of plant system production for glycosylated proteins is still limited because of differences between plant and animal glycosylation machinery. For example, a higher number of Guy's 13 glycoforms in plant than in mammalian expression system has been reported [124]. But despite high structural diversity of the plantibody (antibody product in plant) N-glycans, glycosylation appears to be sufficient for the production of a soluble and biologically active IgG in the plant system and is not a limitation to the use of plantibody Guy's 13 for topical immunotherapy [124]. However, these plants N-glycans are immunogenic. One strategy that could be used to produce recombinant glycoproteins with non-immunogenic glycans is to produce these glycoproteins in plants devoid of one or several enzymes, present in the Golgi apparatus, improved in the N-glycans maturation [114]. Arabidopsis *cgl* mutants, deficient in N-acetylglucosaminyltransferase I (*GnTI*) activity, enzyme initiating the formation of complex N-linked glycans on secretory glycoproteins, were generated. Studies on these mutants demonstrated an effective reduction of *GnTI* activity which can be achieved in mature tissues by means of *GnTI*-mediated gene silencing [125]. These researches opened the way for the production of therapeutic glycoproteins in transgenic plant species carrying minimal compatible *N*-glycans of uniform *N*-acetylglucosamine structure [125]. This methodology has already been reported for mammalian cell lines [126]. Humanisation of the N-glycosylation in transgenic plants contributed to expand the use of plant system for the glycoproteins production [114].

Conclusion

Many biotechnological applications require high-level expression of proteins. Thus, in the past decade, plant-based expression systems have emerged as a serious competitive force with the aim of large-scale production of recombinant valuable proteins. To verify functionality, integrity and stability of gene product proteins, we developed a flexible and versatile system, for a rapid and large production in rice and tobacco. The strategy developed, and first results obtained, using an RYMV-amplicon system combined with RYMV-silencing suppressor, show that this system seem to be a promising tool for the production of recombinant proteins (Figure 6). This system is a good alternative for the production of proteins, like anti-*leishmania* vaccine, having an incorrect folding in bacteria system. Moreover, high-level

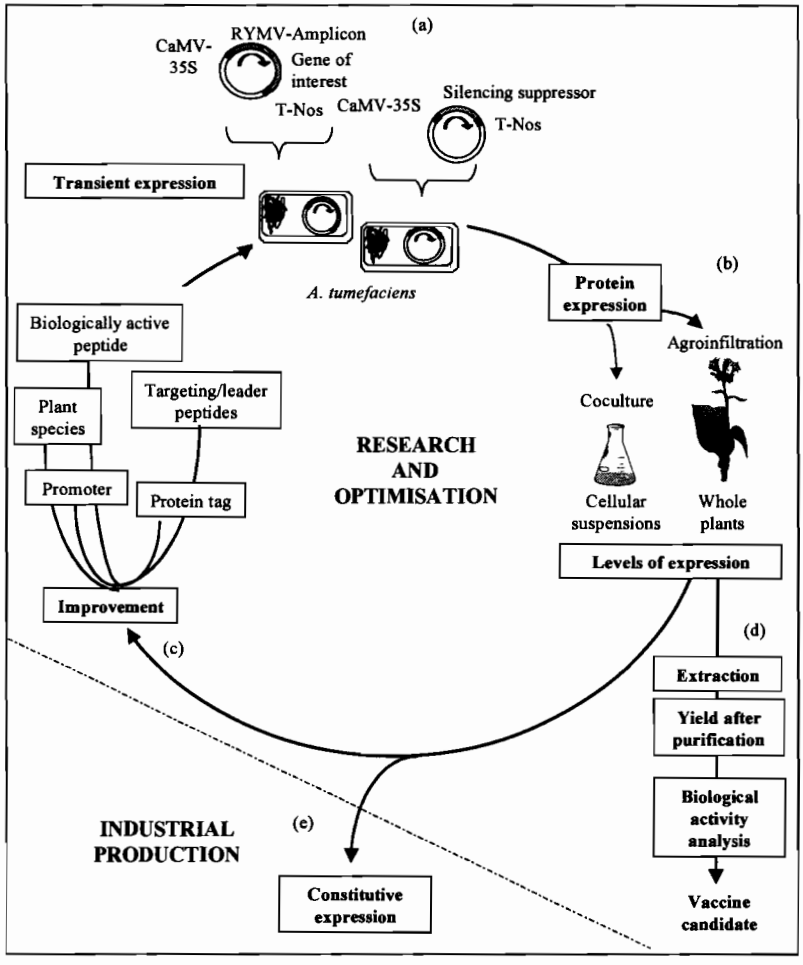


Figure 6. Optimisation of transitory system in order to use the powerful system for constitutive expression of recombinant proteins. (a) Vectors containing gene of interest or silencing suppressors are inserted into *Agrobacterium tumefaciens*. (b) Recombinant proteins are transiently expressed in whole plants with agroinfiltration assays or in cellular suspensions with coculture assays. (c) System of expression could be improved: using other plant species, using other promoters to target the protein, using a tag to facilitate the purification step and using the only biologically active peptide. (d) Products are extracted and analysed to test the structure and the activity of proteins. (e) As soon as the system was optimised to produce a biologically active protein, the production of recombinant proteins could be developed with constitutive system.

expression of recombinant proteins obtained by transitory system is sufficient for structure and activity analysis. The transient expression developed is a powerful tool to choose best expression system to improve yield of production or to facilitate the purification step. Next, production of correct folding and biologically active protein will permit development of a constitutive expression system for the production of the anti-*leishmania* vaccine. This transient system will also be promised to produce another interest protein.

References

1. Hacker, D.L., and Sivakumarann, K. 1997, Virology, 234, 317.
2. Ngon a Yassi, M.N., Ritzenthaler, C., Brugidou, C., Fauquet, C., and Beachy, R.N. 1994, J. Gen. Virol., 75, 249.
3. Qu, C., Liljas, L., Opalka, N., Brugidou, C., Yeager, M., Beachy, R.N., and Fauquet, C.M. 2000, Structure, 80, 1095.
4. Delseny, M., Salses, J., Cooke, R., Sallaud, C., Regad, F., Lagoda, P., Guiderdoni, E., Ventelon, M., Brugidou, C., and Ghesquière, A. 2001, Plant Physiol. Biochem., 39, 323.
5. Delseny, M. 2003, Curr. Opin.Plant Biol., 6, 101.
6. Causse, M.A., Fulton, T.M., Cho, Y.G., Ahn, S.N., Chunwongse, J., Wu, K., Xiao, J., Yu, Z., Ronald, P.C., and Harrington, S.E. 1994, Genetics, 138, 1251.
7. McCouch, S. 2000, Proc. Natl. Acad. Sci., 95, 1983.
8. Tyagi, A.K., and Mohanty, A. 2000, Plant Sci., 158, 1.
9. Christou, P. 1997, Plant Mol. Biol., 35, 197.
10. Goff, S.A., Ricke, D., Lan, T-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., and Varma, H. 2002, Science, 296, 92.
11. Yu, J., Hu, S., Wang, J., Wong, G.K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., and Zhang, X. 2002, Science, 296, 79.
12. Sasaki, T., and Burr, B. 2000, *Curr.* Opin. Plant Biol., 3, 138.
13. Yuan, Q., Ouyang, S., Liu, J., Suh, B., Cheung, F., Sultana, R., Lee, D., Quackenbush, J., and Buell, C.R. 2003, 31, 229.
14. Miyao, A., Tanaka, K., Murata, K., Sawaki, H., Takeda, S., Abe, K., Shinozuka, Y., Onosato, K., and Hirochika, H. 2003, Plant Cell, 15, 1771.
15. Sallaud, C., Meynard, D., van Boxtel, J., Gay, C., Bès, M., Brizard, J.P., Larmande, P., Ortega, D., Raynal, M., and Portefaix, M. 2003, Theor. Appl. Genet., 106, 1396.
16. Sallaud, C., Gay, C., Larmande, P., Bès, M., Piffanelli, P., Piégu, B., Droc, G., Regad, F., Bourgeois, E., and Meynard, D. 2004, Plant J., 39, 450.
17. Miki, D., and Shimamoto, K. 2004, Plant Cell Physiol., 45, 490.
18. Verdaguer, B., De Kochko, A., Fux, C.I., Beachy, R.N., and Fauquet, C. 1998, Plant Mol. Biol., 37, 1055.
19. Verdaguer, B., De Kochko, A. and Beachy, R.N. 1996, Plant Mol. Biol., 31, 1129.
20. Gleba, Y., Marillonnet, S., and Klimyuk, V. 2004, Curr. Opin. Plant Biol., 7, 182.
21. Scholthof, H.B., Scholthof, K-B.G., and Jackson, A.O. 1996, Ann. Rev. Phytopathol., 34, 299.
22. Ahlquist, P., French, R., and Bujarski, J.J. 1987, Adv. Virus Res., 32, 215.

23. Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. 1994, Science, 263, 802.
24. Chiu, W-L., Niwa ,Y., Zeng, W., Hirano, T., Kobayashi, H., and Sheen, J. 1996, Curr. Biol., 6, 325.
25. Haseloff , J., and Amos, B. 1995, Trends Genet., 11, 328.
26. Sheen, J., Hwang, S., Niwa, Y., Kobayashi, H., and Galbraith, D.W. 1995, Plant J., 8, 777.
27. Baulcombe, D.C., Chapman, S., and Cruz, S. 1995, Plant J., 7, 1045.
28. Yusibov, V., Shivprasad, S., Turpen, T.H., Dawson, W., and Koprowski, H. 1999, Curr. Top. Microbiol. Immunol., 240, 81.
29. Pogue, G.P. 2002, Ann. Rev. Phytopathol., 40, 45.
30. Callaway, A., Giesman-Cookmeyer, D., Gillock, E.T., Sit, T.L., and Lommel, S.A. 2001, 39, 419.
31. Hwang, D., Roberts, I.M., and Wilson, T.M.A. 1994, Proc. Nat. Acad. Sci., 91, 9067.
32. Leffel, S.M., Mabon, S.A., and Stewart, C.N. 1997, BioTechniques, 23, 912.
33. Takken, F.L.W., Luderer, R., Gabriels, S.H.E.J., Westerink, N., Lu, R., de Wit, P.J.G.M., and Joosten, M.H.A.J. 2000, Plant J., 24, 275.
34. Lu, R., Martin-Hernandez, A.M., Peart, J.R., Malcuit, I., and Baulcombe, D.C. Methods, 30, 296.
35. Baulcombe, D.C. 1999, Curr. Opin. Plant Biol., 2, 109.
36. Ruiz, M.T., Voinnet, O., and Baulcombe, D.C. 1998, Plant Cell, 10, 937.
37. Robertson, D. 2004, Ann. Rev. Plant Biol., 55, 495.
38. Burton, R.A., Gibeaut, D.M., Bacic, A., Findlay, K., Roberts, K., Hamilton, A., Baulcombe, D.C.,and Fincher, G.B. 2000, Plant Cell, 12, 691.
39. Johnson, J. 1997, Ann. Rev. Phytopathol., 35, 67.
40. Sugiyama, Y., Hamamoto, H., Takemoto, S., Watanabe, Y., and Okada, Y. 1995, FEBS Lett., 359, 247.
41. Wu, L., Jiang, L., Zhou, Z., Fan, J., Zhang, Q., Zhu, H., Han, Q., and Xu, Z. 2003, Vaccine, 21, 4390.
42. Scholthof, H.B. 1999, J. Virol., 73, 7823.
43. Ratcliff, F., Martin-Hernandez, A.M., and Baulcombe, D.C. 2001, Plant J., 25, 237.
44. Wu, L., Fan, J., Jiang, L., Wang, H., Song, R., Zhang, Q., Zhu, H., Li, N., Liu, Z., and Xu, Z. 2003, J. Virol. Meth., 111, 101.
45. Thomas, C.L., Jones L., Baulcombe, D.C., and Maule, A.J. 2001, Plant J., 25, 417.
46. Voinnet, O., Pinto, Y.M., and Baulcombe, D.C. 1999, Proc. Natl. Acad. Sci., 96, 14147.
47. Ratcliff, F.G., MacFarlane, S.A., and Baulcombe, D.C. 1999, Plant Cell, 11, 1207.
48. Holzberg, S., Brosio, P., Gross, C., and Pogue, G.P. 2002, Plant J., 30, 315.
49. Attere, A.F., and Fatokun, C.A. 1983, Plant Dis., 67, 420.
50. Abo, M.E., Sy, A.A., Alegbjo, M.D. 1998, J. Sust. Agri., 11, 85.
51. Bakker, W. 1974, Characterization and ecological aspects of Rice yellow mottle virus in Kenya. [Doctorat], Wageningen.
52. Fargette, D., Pinel, A., Abubakar, Z., Traore, O., Brugidou, C., Fatogoma, S., Hebrard, E., Choisy, M., Sere, Y., and Fauquet, C. 2004, J. Virol., 78, 3252.
53. Fargette, D., Pinel, A., Halimi, H., Brugidou, C., Fauquet, C., and Van Regenmortel, M. 2002, Arch. Virol., 147, 5863.

54. Brugidou, C., Holt, C., Yassi, M.N., Beachy, R.N., and Fauquet, C. 1995, Virology, 206, 108.
55. Fargette, D., Pinel, A., Abubakar, Z., Traore, O., Brugidou, C., Fatogoma, S., Hebrard, E., Choisy, M., Sere, Y. and Fauquet, C. 2004, J. Virol., 78, 3252.
56. Kouassi, N., Brugidou, C., Chen, L., Ngon, A. Yassi, M., Beachy, R.N., and Fauquet, C.M. 1997, Intern. Rice Res. Notes, 14.
57. Bonneau, C. 1998, Etude cytologique et moléculaire du mouvement du virus de la panachure jaune du riz (RYMV), Orsay, Université Paris XI.
58. Mallory, A.C., Parks, G., Endres, M.W., Baulcombe, D.C., Bowman, L.H., Pruss, G.J., and Vance, V.B. 2002, Nature Biotechnol., 20, 622.
59. Susi, P., Hohkuri, M., Wahlroos, T., and Kilby, N.J. 2004, Plant Mol. Biol., 54, 157.
60. Cogoni, C., and Macino, G. 2000, Curr. Opin. Genet. Develop., 10, 638.
61. Aravin, A.A., Klenov, M.S., Vagin, V.V., Rozovskii, Y.M., and Gvozdev, V.A. 2002, Mol. Biol., 36, 180.
62. Carthew, R.W. 2001, Curr. Opin. Cell Biol., 13, 244.
63. Hammond, S.M., Caudy, A.A., and Hannon, G.J. 2001, Nature Rev. Genet., 2, 110.
64. Wang, M.-B., and Waterhouse, P.M. 2000, Plant Mol. Biol., 43, 67.
65. Han, Y., Griffiths, A., Li, H., and Grierson, D. 2004, FEBS Lett., 563, 123.
66. Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. 2002, EMBO J., 21, 4671.
67. Elbashir, S.M., Lendeckel, W., and Tuschl, T. 2001, Genes Develop., 15, 188.
68. Hammond, S.M., Bernstein, E., Beach, and D., Hannon, G.J. 2000, Nature, 404, 293.
69. Matzke, M.A., Matzke, A.J.M., Pruss, G.J., and Vance, V.B. 2001, Curr. Opin. Genet. Develop., 11, 221.
70. Himber, C., Dunoyer, P., Moissiard, G., Ritzenthaler, C., and Voinnet, O. 2003, EMBO J., 22, 4523.
71. Waterhouse, P.M., Wang, M.-B.,and Finnegan, E.J. 2001, Trends Plant Sci., 6, 297.
72. Mlotshwa, S., Voinnet, O., Mette, M.F., Matzke, M., Vaucheret, H., Ding, S.W., Pruss, G., and Vance, V.B. 2002, Plant Cell, 14, S289.
73. Lecellier, C.-H., and Voinnet, O. 2003, Immunol. Rev., 198, 285.
74. Ding, S.-W., Li, H., Lu, R., Li, F., and Li, W.-X. 2004, Virus Res., 102, 109.
75. Kasschau, K.D., and Carrington, J.C. 1998, Cell, 95, 461.
76. Voinnet, O. 2001, Trends Genet., 17, 449.
77. Moissiard, G., and Voinnet, O. 2004, Mol. Plant Pathol., 5, 71.
78. Anandalakshmi, R., Pruss, G.J., Ge, X., Marathe, R., Mallory, A.C., Smith, T.H., and Vance, V.B. 1998, *Proceedings of the National Academy of Sciences* 1998, **95**:13079-13084.
79. Brigneti, G., Voinnet, O., Li, W.-X., Ji, L.-H., Ding, S.-W., and Baulcombe, D.C. 1998, EMBO J., 17, 6739.
80. Roth, B.M., Pruss, G.J., and Vance, V.B. 2004, Virus Res., 102, 97.
81. Lu, R., Folimonov, A., Shintaku, M., Li, W.-X., Falk, B.W., Dawson, W.O., and Ding, S.-W. 2004, Proc. Natl. Acad. Sci., 101, 15742.
82. Voinnet, O., Lederer, C., and Baulcombe, D.C. 2000, Cell, 103, 157.
83. Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D. 2003, Plant J., 33, 949.

84. Bonneau, C., Brugidou, C., Chen, L., Beachy, R.N., and Fauquet, C. 1998, Virology, 244, 79.
85. Daniell, H., Streatfield, S.J., and Wycoff, K. 2001, Trends Plant Sci., 6, 219.
86. Twyman, R.M., Stoger, E., Schillberg, S., Christou, P., Fischer, R. 2003, Trends Biotechnol., 21, 570.
87. Desjeux, P. 1996, Clin. Dermatol., 14, 417.
88. Sundar, S. 2001, Trop. Med. Int. Health, 6, 849.
89. Lira, R., Sundar, S., Makharia, A., Kenney, R., Gam, A., Saraiva, E., and Sacks, D. 1999, J. Infect. Dis., 180, 564.
90. Sundar, S., More, D.K., Singh, M.K., Singh, V.P., Sharma, S., Makharia, A., Kumar, P.C., and Murray, H.W. 2000, Clin. Infect. Dis., 31, 1104.
91. Singh, N., Singh, R.T., and Sundar, S. 2003, J. Infect. Dis., 188, 600.
92. Merlen, T., Sereno, D., Brajon, N., Rostand, F., and Lemesre, J.L. 1999, Am. J. Trop. Med. Hyg., 60, 41.
93. Lemesre, J.L. 1993, IRD-Patent FR N° 9305779.
94. Lemesre, J.L., 1994, IRD-Patent PCT/FR N°94/00577.
95. Lemesre, J.L., Holzmuller, P., Cavaleyra, M., Bras Gonçalves, R., Hottin, G., and Papierok, G. Vaccine, in press.
96. Handman, E. 2001, 14, 229.
97. Sjolander, A., Baldwin, T.M., Curtis, J.M., Bengtsson, K.L., and Handman, E. 1998, Vaccine, 16, 2077.
98. Stöger, E., Vaquero, C., Torres, E., Sack, M., Nicholson, L., Drossard J., Williams, S., Keen, D., Perrin, Y., and Christou, P. 2000, Plant Mol. Biol., 42, 583.
99. Horn, M.E., Woodard, S.L., and Howard, J.A. 2004, Plant Cell Rep., 22, 711.
100. Vaquero, C., Sack, M., Schuster, F., Finnern, R., Drossard, J., Schumann, D., Reimann, A., and Fischer, R. 2002, FASEB J., 16, 408.
101. Hood, E.E., Witcher, D.R., Maddock, S., Meyer, T., Baszczynski, C., Bailey, M., Flynn, P., Register, J., Marshall, L., and Bond, D. 1997, Mol. Breeding, 3, 291.
102. Zhong, G.-Y., Peterson, D., Delaney, D.E., Bailey, M., Witcher, D.R., Register III, J.C., Bond, D., Li, C.-P., Marshall, L., and Kulisek, E. 1999, Mol. Breeding, 5, 345.
103. Witcher, D.R., Hood, E.E., Peterson, D., Bailey, M., Bond, D., Kusnadi, A., Evangelista, R., Nikolov, Z., Wooge, C., and Mehigh, R. 1998, Mol. Breeding, 4, 301.
104. Fischer, R., Vaquero-Martin, C., Markus, S., Drossard, J., Emans, N., and Commandeur, U. 1999, Biotech. Appl. Biochem., 30, 113.
105. Krebitz, M., Wagner, B., Ferreira, F., Peterbauer, C., Campillo, N., Witty, M., Kolarich, D., Steinkellner, H., Scheiner, O., and Breiteneder, H. 2003, J. Biology, 329, 721.
106. Kapila, J., De Rycke, R., Van Montagu, M., and Angenon, G. 1997, Plant Sci., 122, 101.
107. Wagner, B., Fuchs, H., Adhami, F., Ma, Y., Scheiner, O., and Breiteneder, H. 2004, Methods, 32, 227.
108. Shivprasad, S., Pogue, G.P., Lewandowski, D.J., Hidalgo, J., Donson, J., Grill, L.K., and Dawson,W.O. 1999, Virology, 255, 312.
109. Ratcliff, F., Harrison, B.D., and Baulcombe, D.C. 1997, Science, 276, 1558.
110. Al-Kaff, N.S., Covey, S.N., Kreike, M.M., Page, A.M., Pinder, R., and Dale, P.J. 1998, Science, 279, 2113.

111. Fischer, R., Emans, N., Schuster, F., Hellwig, S., and Drossard, J. 1999, Biotechnol. Appl. Biochem., 30, 109.
112. Doran, P.M. 2000, Curr. Opin. Biotechnol., 11, 199.
113. Matsumoto, S., Ikura, K., Ueda, M., and Sasaki, R. 1995, Plant Mol. Biol., 27, 1163.
114. Faye, L., Landry, N., Lerouge, P., Gomord, V., and Vézina, L.-P. 2001, Médecine/Sciences, 17, 867.
115. Wright, K.E., Prior, F., Sardana, R., Altosaar, I., Dudani, A.K., Ganz, P.R., and Tackaberry, E.S. 2001, Transgenic Research, 10, 177.
116. Parmenter, D.L., Boothe, J.G., van Rooijen, G.J., Yeung, E.C., and Moloney, M.M. 1995, Plant Mol. Biol., 29, 1167.
117. Fischer, R., Drossard, J., Emans, N., Commandeur, U., and Hellwig, S. 1999, Biotechnol. Appl. Biochem., 30, 117.
118. Magnuson, N.S., Linzmaier, P.M., Reeves, R., An, G., HayGlass, K., and Lee, J.M. 1998, Prot. Expres. Purif., 13, 45.
119. Sijmons, P.C., Dekker, B.M., Schrammeijer, B., Verwoerd, T.C., van den Elzen, P.J., and Hoekema, A. 1990, Biotechnol., 8, 217.
120. Darain, F.B.C., and Shim, Y.B. 2004, Biosensors Bioelectronics, 20, 857.
121. Feng, G.W.M. 1995, Biotechniques, 19, 956.
122. Li, H.-Y., and Chye, M.-L. 2004, Plant Mol. Biol., 54, 233.
123. Liu, Z.-Z., Wang, J.-L., Huang, X., Xu ,W.-H., Liu, Z.-M., and Fang, R.X. 2003, Planta, 216, 824.
124. Cabanes-Macheteau, M., Fitchette-Laine, A., Loutelier-Bourhis, C., Lange, C., Vine, N., Ma, J., Lerouge, P., and Faye, L. 1999, Glycobiology, 9, 365.
125. Wenderoth, I., and von Schaewen, A. 2000, Plant Physiol., 123, 1097.
126. Stanley, P. 1989, Mol. Cell. Biol., 9, 377.

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

Molecular Biology of Tropical Plants, 2006: 41-56 ISBN: 81-308-0138-8
Editor: Claudine Franche

3 Advances in genetic transformation of *Eucalyptus* species

Quoirin, M.[1] and Quisen, R.[2]
[1]Department of Botany, Federal University of Parana, Curitiba, PR, Brazil
[2]Embrapa Amazônia Ocidental, Manaus, AM, Brazil

Abstract

The genus Eucalyptus comprises approximately 700 species and 1% of them are used for industrial purposes. They are fast growing species with short rotation, widely cultivated for wood-pulp, timber, veneer, firewood, shelter and essential oil production. Their economic importance is therefore enormous. Genetic improvement of eucalypts has paralleled the increasing importance of plantations and resulted in the development of several breeding and biotechnology programmes world-wide. To complement this activity,

Correspondence/Reprint request: Dr. Quoirin, M., Department of Botany, Federal University of Parana, Curitiba PR, Brazil. E-mail: mquoirin@ufpr.br

it would be interesting to improve or introduce specific desirable traits through genetic engineering. In this chapter, we review the major results achieved to date in organogenesis by adventitious bud formation and somatic embryogenesis. Advances in genetic transformation are also reported, especially those obtained and published during the last decade, and some information is given on private forest companies biotechnological programmes.

Introduction

Eucalyptus species are extensively used world-wide in commercial afforestation programmes. In 2000, the total area planted with these species was estimated at 17.8 millions hectares [1]. Their main advantages are their fast growth rates, short rotations and their wood properties, which allow production of high quality pulp for paper and cellulose. These species are also used for fuel wood and essential oil production. Each species is grown in a particular region depending on climate, soil fertility and water availability.

Genetic engineering could assist the genetic improvement of *Eucalyptus* species, as it enables specific traits to be added to elite genotypes without altering other characteristics. For improvement of eucalypt trees, conventional methods are very low, due to long breeding cycles, including long juvenile phase, high levels of heterozygosis, and are limited by the tree size. Most of the breeding programmes are focused on productivity and wood quality.

Excellent reviews on *Eucalyptus* transformation were published in 1994 [2] and in 1997 [3]. This chapter reports on the main progress made in genetic transformation of *Eucalyptus* species over the last decade, including the *in vitro* methods used for plant regeneration.

1. Plant regeneration

The establishment of a good regeneration system is fundamental for further genetic transformation of *Eucalyptus* species. During the last years, there has been considerable progress in this area. In many cases, the development of adventitious buds has been reached, but some papers also described the process of somatic embryogenesis.

The organogenesis process generally starts from juvenile parts of the plant, such as cotyledons, hypocotyls and leaf fragments excised from young plantlets developed from seeds *in vitro* (Table 1). In some cases, it was also initiated from zygotic embryos [4]. This process starts with callus formation. Bud regeneration is often achieved on the same medium used for callogenesis. Mineral media have to be adapted in function of the species. MS [5] medium is generally used, but also others like B_5 [6] for *E. camaldulensis* [7, 8, 9], Lainé and David (1994) specific media [10] or modifications of MS and WPM media [13]. In most cases, callus development occurred in the presence of an auxin

(NAA) and a cytokinin (BA). Some authors used a thiourea-derivated cytokinin (4-CPPU, TDZ) instead of a purine-derivative [8, 11, 12]. In some cases, both types are used in a sequence of media [13]. Picloram was also applied for callogenesis and organogenesis of *E. gunnii* explants [14].

During the first days or the first month of the culture, it is sometimes important to keep the explants in the dark or under low intensity light in order to avoid explant oxidation. This was observed for callus formation and bud regeneration for *E. camaldulensis* [15], for *E. grandis* [10], for *E. grandis x urophylla* [13, 16] and for *E. tereticornis* [17]. Some authors also recommend

Table 1. *In vitro* plant regeneration of *Eucalyptus* sp. (growth regulators in µM).

Species	Explant	Callogenesis medium	Regeneration medium	Rooting medium	References
E. camaldulensis	Cotyledon	MS salts; NN vit.; NAA 5.3; BA 3.1	Same as for callogenesis	NAA 0.5 – 5.3	[18]
E. camaldulensis	Leaflet from micropropagated plants	WPM; casein 1000 mg/L; NAA 16; BA 0.44; sucr. 5%	WPM; BA 1.32	WPM	[19]
E. camaldulensis	Hypocotyl	B_5; CW 100 ml/L; Glut. 200 mg/L; HC 100 mg/L; BA 4.4; NAA 16; sucr. 3%	Same as for callogenesis	MS macro; IBA 5	[7]
E. camaldulensis	Cotyledon Hypocotyl	B_5; 4-CPPU 1; NAA 0.1; sucr. %	B_5; NAA 0.1; BA 0.9; sucr. 1%	1/4 B_5; NAA 0.05; sucr. 1%	[8]
E. camaldulensis	Leaf explant	B_5, combinations of CK and Aux	B_5; TDZ 0.5-2.5	Without GR	[9]
E. globulus	Cotyledon and embryonary axes	MS; 2,4-D 7.5; BA 0.44; sucr. 3%	WPM; NAA 1; BA 2.6; arginine 574; sucr. 3%	MS salts; White vit; IBA 7.5; sucr. 2%	[5]
E. globulus	Cotyledon Hypocotyl	MS; BA 1; TDZ 1			[20]
E. globulus	Fragments of young plantlets	MS; 2,4-D 7; BA 0.44; sucr. 3%	WPM; NAA 1; BA 2.6; arginine 574; sucr. 3%	MS salts; White vit; IBA 7.5; sucr; 2%	[21]
E. globulus	Cotyledon Hypocotyl	MS; TDZ 0.05; 2,4-D 0.2 or NAA 5.0	Same as for callogenesis, then BA 5	Not indicated	[22]
E. globulus	Cotyledon Hypocotyl	MS; BA 2.2; NAA 5.3	MS; BA 4.4; NAA 2.65	IBA	[27]
E. grandis	Hypocotyl	NAA 21; Kin 4.7	Nodules and shoots: without GR	IBA 0.5 - 1.5	[23]
E. grandis	Leaf explants	G22 medium, BA 2;	GBA medium,	KG médium;	[10]

Table 1. Continued

	from micropropagated plants	NAA 2 or R5; Zea 36.3; NAA 8	BA 6.6; NAA 0.5	IBA 1.2; sucr. 2%	
E. grandis hybrids	Leaf and node segments	MS; BA 1.1; 2,4-D 7.5	WPM; TDZ 3; NAA 0.01	Not indicated	[24]
E. urophylla	Hypocotyl fragments	MS; BA 0.4-44; NAA 0-2.6 or TDZ 0.045-10; NAA 0-16	MS; NAA 1; BA 0.9	MS; NAA 5.3; BA 0.044	[12]
E. grandis x urophylla	Hypocotyl Cotyledon	Not indicated	SP; 2,4-D and BA 5 of each	Not indicated	[25]
E. grandis x urophylla	Hypocotyl Cotyledon Leaf	MS; TDZ 2	BA 2.7-1 and NAA 0.48 or Zea and NAA	Elong. without GR, rooting: IBA 12.5	[26]
E. grandis x urophylla	Leaf fragments	R5; TDZ 2.4; NAA 0.5	R5 with MS iron and micronutr.; BA 4.4	MS salts, White vit.; PVP 800 mg/L inositol 555; IBA 7.5	[11]
E. grandis x urophylla	Leaf fragments	BIP medium; 2,4-D 0.5; BA 1	BIT medium; NAA 0.1; TDZ 3 SDM medium; BA 2; NAA 0;5	Ex-vitro	[13]
E. gunnii	Leaf and node fragments	MS; picloram 0.04; BA 1 (1 week) and then 2.2	Same as for callogenesis	Not indicated	[14]
E. nitens	Hypocotyl Cotyledon	MS; BA 2.2; NAA 5.3	MS; BA .4; NAA 2.65	IBA	[27]

Culture media: B_5 = Gamborg et al., 1968 [6]; G22, GBA, KG = Lainé & David, 1999 [10]; MS = Murashige & Skoog, 1962 [5]; NN = Nitsch & Nitsch, 1969 [28]; R5 = Lainé & David, 1994 [10]; White, 1963 [29]; WPM = Lloyd & McCown, 1980 [30]. Others: Argin. = arginine; BA = benzyladenine; CW = coconut water; 2,4-D = 2,4-dichlorophenoxyacetic acid; Glut. = glutamine; GR = growth regulator; HC = casein hydrolysate; IBA = indolbutyric acid; Kin = kinetin; NAA = naphthalenacetic acid; TDZ = thidiazuron; Zea = zeatin.

the use of an anti-oxydant solution during the excision of the explants [13]. Shoot elongation often takes place on regeneration medium, but in some cases a specific medium without growth regulators has been used [13]. Few articles report the presence of somaclonal variants or plants with abnormal haploidy. It was the case for *E. urophylla* shoots obtained by organogenesis from hypocotyls: 28% of 25 regenerated plants were haploid and 12% triploid [12]. This aspect must be considered when studying transformed plants.

Somatic embryogenesis is another regenerative procedure that may be used to obtain plants from transformed callus. It is of great interest as it allows the direct formation of a complete plantlet with a good rooting system, what is not always possible with adventitious rooting. The formation of somatic embryos or embryo-like structures has been described for *E. citriodora* [31, 32,

33], *E. dunnii* [34, 35] and *E. grandis* [35]. Some results were also obtained for *E. globulus* [22, 36] and for *E. nitens* [27, 36], but plant regeneration was not achieved. In *E. urophylla* hypocotyl explants, NAA in combination with TDZ induced both organogenesis and the production of heart shaped embryo-like structures that developed into shoots on the same medium [12]. Recently, the induction of the embryogenic process from juvenile explants of *E. globulus* was described [37]. Embryogenic callus was obtained from cotyledons in the presence of NAA and somatic embryos developed on media containing 3 to 5 $mg.L^{-1}$ NAA alone or with 1 $mg.L^{-1}$ 2,4-D. The same auxin was used for somatic embryo induction in the *E. citriodora* [30, 32] and *E. dunnii* [33] species. Conversion of *E. globulus* somatic embryos into plantlets was observed in 21% of them. When the explants were cultured under light, phenolic compounds accumulation was observed, which impeded the establishment of embryogenic cultures. However, light is important for somatic embryos formation [34, 37].

2. Genetic transformation

2.1. Direct transformation

The first results on direct transformation were published in 1990 for *E. saligna* [38] and in 1991 for *E. gunnii* [39]. In both cases, protoplasts were used for transformation. Regeneration of *E. saligna* plantlets from protoplasts was reported [38]. In the case of *E. gunnii* protoplasts, two methods of transformation were tested: polyethylene glycol (PEG) treatment and electropulsation [39]. Protoplasts were isolated from calli and from cell suspensions. The optimum pulsing conditions were three 1-second-spaced pulses of 1200 $V.cm^{-1}$ applied during 1 ms for callus protoplasts and 3 pulses of 800 $V.cm^{-1}$ during 1 ms for cell suspension protoplasts. A heat shock pre-treatment at 45°C for 5 min stimulated PEG mediated DNA uptake, possibly through an effect on membrane permeability. Chloramphenicol acetyl transferase (CAT) and *gus* genes were expressed. PEG treatment greatly stimulated CAT activity and transient expression of this gene was observed. The physiological conditions of the cells were important, as well as the type of promoter associated with reporter genes [39]. Unfortunately, plantlet regeneration was not achieved [39].

The biolistic method was used for transformation of zygotic embryos of *E. globulus* [4]. Plants were recovered through organogenesis from callus. The integration of *gus* and *nptII* genes into the plant cells was confirmed by Southern blot analysis. The same method was used with calli derived from cotyledons and hypocotyls of the hybrid *Eucalyptus grandis x E. urophylla* [40]. In this case, the *gus* and *nptII* genes were introduced into the calli through the acceleration of tungsten particles, and their presence demonstrated by PCR analysis after 105 days. However, the recovery of transgenic plants from the transformed calli was not possible.

2.2. Indirect transformation

E. camaldulensis was the first species used for experiments of transformation mediated by *Agrobacterium tumefaciens* [20, 41, 44]. It is also the most studied species: among 24 published results, twelve were obtained with *E. camaldulensis* [7, 20, 41, 49]. The choice of this species is due to both its economic importance and its good regeneration capacity. Results were also published for *E. globulus* [4, 22, 25, 50], *E. grandis* [25, 51, 52, 53] and for *E. gunnii* [50]. A procedure of *E. camaldulensis* shoot transformation was patented by Teasdale [54]. Studies were also carried out with juvenile material of *E. grandis x urophylla,* and the transient expression of the marker gene *gus* was evaluated in function of several co-culture parameters [26].

As for crops, several factors affect transformation efficiency of tree tissues. The most important aspects of eucalypt transformation protocols are reviewed below. Some of them are correlated with the bacterium, others with the plant tissues.

Agrobacterium strains

Agrobacterium tumefaciens strains and their plasmids are classified in function of the opine genes they carry. These opines, which are synthesised in the infected plant cells, are mainly agropine, nopaline and octopine. For *E. globulus* transformation the following strains were tested: LBA 4404 pAL4404 (octopine), C58C1 pMP90 (nopaline), EHA101 pEHA101 (agropine) and AGL1 pTiBo542 (agropine) [21]. EHA101 yielded four times more transient expression than LBA4404 and twice as much as C58C1. According to these authors, tree species generally respond better to the nopaline than to the octopine strains. For stable expression of *gus* gene, it was shown that strains C58C1, EHA101 and AGL1 were the most efficient, while LBA4404 resulted in low expression [55]. However, the integration into the plant genome also depended on plant genotype. In another study conducted with *E. camaldulensis* explants [19], five strains were compared: A6, LBA4404, GV3111, AGL1 and GV3850. The best transformation results (rooted transformed plants) were obtained with the two last strains.

Promoters

The constitutive promoter CaMV35S was used in all the published studies, sometimes with double enhancer [46]. However, it is necessary to select promoters to drive specific spatial and temporal expression or inactivation of genes used for wood modification.

Co-culture duration

Co-culture can be applied to the explants transferred to a liquid or a solid medium. For tree species like eucalypts, this duration must be of several days,

unlike herbaceous species. Some authors recommended three days [7, 19, 56], others four [25, 46], five [13, 24] or even six days [21].

Phenolic compounds and osmoprotectants

The bacterium virulence (*vir* genes expression) is induced naturally by phenolic compounds released by the wounded plant tissues. These compounds are used in transformation protocols in order to stimulate bacteria virulence. Most of them are acetosyringone (AS), hydroxy-acetosyringone, sinapyl and coniferyl alcohols. Some authors recommended the addition of 50 µM AS to solid media used for explant activation and co-culture of *E. grandis x urophylla* hybrid [13]. In transformation experiments conducted with *E. globulus*, it was observed that the optimum concentration of AS depended on the clone and the medium used for *A. tumefaciens* growth [57]. This author recommended 20 or 100 µM added to MYA [58] medium. AS (100 mM) may also be added to the bacterial suspension medium, as was shown for *E. grandis* and *E. grandis x urophylla* tissues infection [16]. Proline is an osmotically active compound which acts as an osmoprotectant. It is sometimes added to the bacterial solution (at 0.5 to 1 mM) used for tissue inoculation [13]. In the case of *E. globulus* transformation [55], proline addition to culture medium had no effect.

Composition of bacterial solution

It is generally recommended that the bacteria be resuspended in a MS type solution. The important point at this stage is to use a slightly acidic solution. The pH medium seems to have an effect on transformation efficiency. However, our results indicated that it is also possible to obtain a transient transformation without resuspending the bacteria.

Plant genotypes

Some genotypes do not respond to bacterial infection. For example 13 out of the 23 clones of *E. camaldulensis* tested for regeneration efficiency were selected [19]. Then 5 of them were used for transformation experiments. It was observed that the clones that were the most efficient in regeneration were also the most efficiently transformed. For *E. grandis x urophylla* hybrid too, gene transfer and integration were greatly dependent on genotype [12].

Type of tissues

The use of young tissues is highly recommended: cotyledonary leaves, hypocotyls, young leaves, young plantlets (2 to 15 days after germination) [56]. In the case of elite material where only adult trees are available, micropropagation from epicormic buds is a preliminary step towards obtaining rejuvenated material.

Sonication treatment and vacuum infiltration

According to several authors [13, 16], sonication treatment (for 15 s) of the explants during bacterial inoculation greatly enhanced the efficiency of transformation. Generally the sonicator is of water-bath type, delivering a frequency of 40 kHz. This treatment provokes micro-injuries at the tissue surface and induces the release of phenolic compounds. Vacuum infiltration eliminates the air present in the intercellular spaces and helps bacteria penetration into the tissues.

Pre-culture

Some authors investigated whether further increase in the efficiency of gene transfer and subsequent formation of transgenic calli could be gained from the introduction of a pre-culture step before explant co-culture. The pre-culture period varied from 0 to 5 days, followed by a co-culture period [7, 13, 21, 24]. After a 4 to 6 days pre-culture on 2,4-D and BA containing medium, the level of *gus* transient expression in *E. globulus* explants was significantly greater than in those without pre-culture or that of 8 days [21]. For *E. grandis* hybrids, four days appeared to be the most indicated pre-culture period [24] while Tournier et al. [13] recommended 2 days. According to Sangwan et al. [59], the presence of an auxin and a cytokinin in a pre-culture medium induced cellular dedifferentiation, increasing transformation frequency. This improvement of DNA uptake could be due to a stimulation of cell division by the hormone in the pre-culture medium, since mitotic cells would be more susceptible to *Agrobacterium* or would have a higher level of transcription [21].

3. Applications of genetic engineering to *Eucalyptus* improvement

The main target traits considered for *Eucalyptus* improvement by genetic engineering are: growth enhancement, fibre property modification, cellulose polymerisation increase; lignin biosynthesis modification, or lignin content reduction, herbicide and insect resistance, and abiotic stress tolerance. Sterility may also be desired in order to prevent dissemination of the transgenes in regions where cross-pollination with non-GM or with native trees is possible, such as Australia. An interesting review on biotechnology of forest trees was published recently by Cooke *et al.* [60]. Interest in manipulating wood properties has considerably increased during the last two decades. Nowadays trees are planted and managed under intensive regimes, like agricultural crops, and these practices affect wood structure and fibre quality [60]. The progress made in genomics and proteomics approaches allows better knowledge of physical and chemical characteristics of wood fibres. Constituents of the fibre cell wall are a target for improvement of wood characteristics. Discovering

cellulose, hemicelluloses and lignin biosynthesis pathways and cloning the key enzyme genes are also important for designing strategies aiming at modifying these compounds. Modification of cellulose may be reached through manipulation of cellulose synthase or precursors of cellulose biosynthesis. Other cell wall proteins may be a target for wood modification. For example, recombinant cellulose-binding domains (CBD or expansin) have been shown to modulate the elongation of different plant cells in vitro [61]. In *Acetobacter xylinum*, CBD increased the activity of the cellulose synthase. When introduced into poplar, *cbd* gene raised biomass production, thanks to fibre cell elongation and to an increase in the average degree of cellulose polymerisation [61]. Other genes of interest are those which encode endo-1,4-ß-glucanase proteins. These enzymes are implicated in cell wall enlargment. The structure of the primary cell wall is determined by interaction between cellulose microfibrils and the xyloglucan network. Cell expansion occurs through the regulated reorientation of these cell wall components. Endoglucanases are encoded by multi-gene families. They hydrolyse polysaccharides, which contain a 1,4-glucan backbone, to non-substituted glucose residues. During cell elongation, native endoglucanases hydrolyse cellulose-xyloglucan links, allowing cellulose chain to move freely. By overexpressing these genes in plant, cells are induced to elongate more rapidly, resulting in faster growth and development [62]. CBD and *Arabidopsis thaliana* endoglucanase genes were fused to different promoters and introduced in *E. camaldulensis*, *E. grandis* and *E. grandis* hybrid commercial clones [62]. Plants were recovered and 25 transgenic lines are presently tested in the field [62].

Lignin is a complex phenolic polymer present in the cell wall of higher plants which has essential functions as it confers mechanical support, impermeability and disease resistance to the plants. Detailed reviews of lignin biosynthesis and manipulation were written recently [63, 64, 65], where the reader will find details on these aspects. During pulp and paper production, it is necessary to separate lignins from cellulose, a procedure that is costly, energy consuming and polluting [13]. Several genes coding for enzymes acting in the different steps of lignin pathway may be down-regulated in order to simplify the lignin elimination process: first, the enzymes of ring hydroxylation (phenyl ammonia lyase, cinnamate 4-hydroxylase, p-coumarate 3-hydroxylase and ferrulate 5-hydroxylase); secondly, those of ring methylation (caffeic acid methyltransferase and caffeoyl-CoA-3-O-methyltransferase) and third the side chain modification enzymes (4-coumarate-CoA ligase, cinnamoyl-CoA reductase and cinnamyl alcohol dehydrogenase) [59]. Some of these genes seem appropriate for genetic manipulation of lignin as they did not demonstrate deleterious effects of the transgenes in tobacco or poplar [59].

In *E. camaldulensis*, modification of lignin metabolism was achieved through the introduction of a cinnamate 4-hydroxylase gene from poplar [44]

or cinnamyl alcohol dehydrogenase antisense constructs (*cad* antisense) [46]. In the first case, transgenic lines carrying sense or antisense gene were produced. The gene integration into the plant genome was confirmed by PCR and Southern blot analyses and more than 100 cuttings were generated from transformed plants [44]. Among the shoots of *E. camaldulensis* transformed with an *E. gunnii cad* antisense gene and selected by PCR analysis, 32% exhibited a significant reduction of CAD activity but the tobacco construct was less efficient. However, there were no significant changes in lignin profiles (quantity, composition and pulp yield) in transgenic lines analysed after 10 months, indicating that CAD activity was not sufficiently suppressed [46]. The same construct from *E. gunnii* was successfully introduced into the hybrid *E. grandis x urophylla* [13]. In this case, 58% of 120 transgenic shoots were significantly inhibited for CAD activity, and nine exhibited the highest down-regulation. After Northern analyses, two lines were considered as the most interesting with 26 and 22% of residual activity.

Other aspects of lignin biosynthesis pathway may be manipulated through modification of transcription factors and regulators genes (MYB, LIM genes). Transgenic tobacco plants with *Ntlim1* antisense gene showed low levels of transcripts from some key phenylpropanoid pathway genes such as phenylalanine ammonia lyase, hydroxycinnamate-CoA ligase and cinnamyl alcohol dehydrogenase [48]. Furthermore, a reduction in lignin content of over 20% was observed in these plants. The same construct was also introduced into *E. camaldulensis* tissues and the resulting plants showed a 20% reduction of lignin content in cell wall residues of stem xylem tissues [48]. Another approach consists in manipulating two genes of lignin pathway in the same plant. Some research has already been carried on model plants and the recovered plants were promising for industrial applications [60]. To our knowledge, this strategy has not been applied to eucalypt species to date.

Genetic manipulations of flowering genes are aimed to shorten flowering and reproductive cycle. Some genes homologues of *Arabidopsis thaliana AP1* and *LEAFY* genes, were identified in *Eucalyptus* [66, 67]. Their overexpression in *Eucalyptus* species might help reduce juvenile phase and shorten generation time, allowing to evaluate these lines earlier than non transgenic trees. Another application of the manipulation of genes that control flower development is the production of sterile trees, with high impact for the safety of transgenic tree release.

Insect resistance may be launched by the insertion of *Bt* (*Bacillus thuringiensis*) toxin genes as for agricultural crops. An example of this was the introduction of *Bt cry3A* gene in *E. camaldulensis* tissues and the recovery of transformed plants that were insect resistant [8].

Trees may also be improved for tolerance to abiotic stresses such as frost and cold, water deficit and salinity. An example is the use of *dreb* tolerance genes for transformation of *E. camaldulensis, E. grandis* and hybrids [68].

During the 1990 decade, several private companies started developing biotechnology programmes aimed at modifying important traits in commercial eucalypt species (Table 2). Some of them have already obtained stably transformed trees that are being tested in the field. Unfortunately, few information is available about their research, due to intellectual property and patent rights.

An example of research developed by these companies is the generation of highly salt-tolerant eucalyptus plants by Nippon Paper Industries Co [69]. According to the company web site, this success was correlated with the application of its proprietary genetic engineering technology, the Multi-Auto-Transformation (MAT) Vector System. The gene introduced in this experiment

Table 2. Main forest companies applying genetic engineering research to *Eucalyptus* species.

Company and reference	***Eucalyptus* species**	**Genes**	**Modified characteristics**
Aracruz (Brazil) [13]	*grandis x urophylla* hybrids	Cinnamyl alcool dehydrogenase (*cad*) antisense	Lignin components
Advanced Technology Cambridge (UK) [43, 46]	*camaldulensis, globulus, grandis, grandis* hybrids, *saligna, urophylla*	NI	Lignin components
Arborgen (USA) [70]	NI	NI	Wood quality, growth enhancement, herbicide resistance
CBD (USA, Israel) [49] and Suzano (Brazil) [61]	*camaldulensis, grandis, grandis* hybrids	Cellulose binding domains (*cbd*) and *cel1*	Fibre properties, biomass enhancement
Nippon Paper Industries Co. Ltd (Japan)	*camaldulensis*		
[47]		Choline oxydase (*codA*)	Abiotic stress resistance (salt)
[48]		Antisense *Ntlim1* (transcription factor)	Reduction of lignin content
Shell Forestry (Netherlands) [46]	*camaldulensis, dunnii, grandis, grandis* hybrids, *saligna, urophylla*	NI	Herbicide tolerance
Oji Paper (Japan) (71)		NI	Adaptation of stock material to acidic soil
[45, 68]	*camaldulensis*	Dehydration-responsive element binding (*Dreb1A*)	Abiotic stress resistance
Mitsubishi Paper Mills (Japan) [72]	NI	NI	Reduction of lignin content

NI: not indicated

is the *codA* gene, named from choline oxidase of the soil bacterium *Arthrobacter globiformis*. Choline oxidase catalyzes the oxidation of choline and promotes the conversion of choline to glycine betaine. Betaine works to protect cells from salt stress. It is known that salt-tolerant plants as *Chenopodium album* L. and barley accumulate substantial amounts of betaine in their cells to balance their cellular solute concentrations. The MAT Vector System allows the introduction of a number of genes and multiple transformations without the use of antibiotic or herbicide resistance genes as selectable marker genes. This technique also has the capability to remove the marker genes from the transgenic plants. Consequently, it is deemed a safer and more environmentally friendly method of gene alteration. The company's internal tests have shown that the genetically altered eucalyptus grew in environments that contained salt at a sodium chloride concentration as high as 200 mM, which is equivalent to one-third of the NaCl concentration of seawater. According to the company, this genetically engineered eucalyptus presents a possible new solution for the global problem of soil deterioration. The modified tree may help to meet the environmental challenges raised by dramatically expanding deserts and salt-containing barrens in many parts of the world. The altered eucalyptus has not only shown a high tolerance to salt but also exhibited a tendency to resist other environmental stresses, such as excessive heat and very low temperatures [69].

Conclusions

Eucalypt species are particularly recalcitrant to organogenesis and genetic transformation. However, considerable progress has been accomplished in the establishment of gene transfer systems as well as in genomics of some commercial species and knowledge of important processes like wood formation and lignin synthesis. However, as in the case of other tree species, the commercial use of transgenic eucalypts in forest plantations faces important technical, ecological and political challenges [73]. Some features inherent in the biology of trees present real challenges to researchers. For instance, the recalcitrance of tissues from adult trees to in vitro culture frequently impedes the regeneration of plants from elite trees. The mechanism of gene silencing needs to be elucidated and somaclonal variation studied [73]. Transgenic lines need to be evaluated in long-term field trials to verify that their yield and adaptability characteristics remain intact. The genes that confer insect, disease or herbicide resistance have to be managed carefully and their ecological impact studied. When we consider the present tendencies of research in forest tree improvement, we can imagine that, in the future, some eucalypt clones will be designed for solid wood production, others for paper or cellulose synthesis, others for phytoremediation or air pollution control, resulting in "specialised" plantations.

Public and private companies have to deal with public opinion opposed to GM trees when they decide about research programmes and investments. At the same time, efforts must be done by researchers to educate the public and make them understand the real advantages and risks associated with the use of genetically modified trees for forest plantations. This is the only way to address irrational fears about transgenic trees.

References

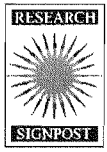
1. FAO. Global Forest Resources Assessment 2000 – Main report. FAO Forestry paper. ISSN 0258-6150, 2000. 479 p. Available HTTP: www.fao.org/forestry/fo/fra/main/index.jsp.
2. Teulières, C., Marque, C., and Boudet, A.M. 1994, Biotechnology in agriculture and forestry, Vol. 29, Plant protoplasts and genetic engineering V, Y.P.S. Bajaj (Ed.), Biotechnology in agriculture and forestry, Vol. 29, Plant protoplasts and genetic engineering V, Springer-Verlag, Berlin, Heidelberg, New York, 289.
3. MacRae, S., and van Staden, J. 1999, Biotechnology in agriculture and forestry, Vol. 44, Transgenic trees VII, Y.P.S. Bajaj, Y.P.S. (Ed.), Springer-Verlag, Heidelberg, New York, 88.
4. Serrano, L., Rochange, F., Semblat, J.P., Marque, C., Teulières, C., and Boudet, A.-M. 1996, J. Exp. Bot., 47, 285.
5. Murashige, T., and Skoog, F. 1962, Physiol. Plant., 15, 473.
6. Gamborg, O. L., Miller, R. A., and Ojima, K. 1968, Exp. Cell. Res., 50, 151.
7. Ho, C.-K., Chang, S.-H., Tsay, J.-Y., Tsai, C.-J., Chiang, V.L., and Chen, Z.-Z. 1998, Plant Cell Rep., 17, 675.
8. Harcourt, R.L., Kyozuka, J., Floyd, R.B., Bateman, K.S., Tanaka, H., Decroocq, V., Llewellyn, D.J., Zhu, X., Peacock, W.J., and Dennis, E.S. 2000, Mol. Breeding, 6, 307.
9. Chang, S.-H., Ho, C.-K., Tsai, J.-Y., and Chen, Z.-Z.. 2000,. Taiwan J. For. Sci., 15, 81.
10. Lainé, E. and David, A. 1994, Plant Cell Rep., 13, 473.
11. Alves, E.C.S.C., Xavier, A., and Otoni, W. C. 2004, Pesq. Agropec. Bras., Brasilia, 39, 421.
12. Tibok, A., Blackhall, N. W., Power, J.B., and Davey, M.R. 1995, Plant Sci., 110, 139.
13. Tournier, V., Grat, S., Marque, C., El Kayal, W., Penchel, R., Andrade G. de, Boudet, A.-M., and Teulières, C. 2003, Trans. Res., 12, 403.
14. Hervé, P., Jauneau, A., Pâques, M., Marien, J.-N., Boudet, A.M., and Teulières, C. 2001, Plant Sci., 161, 645.
15. Dibax, R. 2004, MSc Thesis, Universidade Federal do Paraná, Curso de Pós-graduação em Agronomia – Produção Vegetal, Curitiba, PR, Brazil.
16. Gonzales, E.R., Andrade, A. De, Bertolo, A. L., Lacerda, G.C., Carneiro, R. T., Defávari, V.A.P., Labate, M.T.V., and Labate, C. A. 2002, Funct. Plant Biol., 29, 97.
17. Subbaiah M.M., and Minocha, S. C. 1990, Plant Cell Rep., 9, 370.
18. Diallo N., and Duhoux, E. 1984, J. Plant Physiol.,115, 177.
19. Mullins, K. V., Llewellyn, D. J., Hartney, V.J., Strauss, S., and Dennis, E.S. 1997, Plant Cell Rep., 16, 787.

20. Azmi, A., Dewitte, W., Devret, C., Van Onckelen, H., Landré, P., Boudet, A. M., and Chriqui, D. 1997, Plant Sci., 127, 81.
21. Moralejo, M., Rochange, F., Boudet, A.-M., and Teulières, C. 1998, Aust. J. Plant Physiol., 25, 207.
22. Nugent, G., Chandler, S., Whiteman, P., and Stevenson, T. 2001, In Vitro Cell. Dev. Biol. - Plant , 37, 388.
23. Warrag, E., Lesney, M.S., and Rockwood, D.J. 1991, Plant Cell Rep., 9, 586.
24. Machado, L. de O. R., Penchel, R. M., Teulières, C., Bertolucci, F.L., and Boudet, A.M. 1997, IUFRO Conference on Silviculture and Improvement of Eucalypts, Salvador, Brazil, 192.
25. Esmeraldo, M.V., Carvalheira, S.B.R.C., Brasil, A. A., Figueiredo, S. A., Barrueto Cid, L. P., and Brasileiro, A. C. M. 1997, IUFRO Conference on Silviculture and Improvement of Eucalypts, Salvador, Brazil, 19.
26. Barrueto Cid, L.P., Machado, A.C.M.G., Carvalheira, S.B.R.C., and Brasileiro, A.C. 1999, Plant Cell Tissue Org. Cult., 56, 17.
27. Bandyopadhyay, S., Cane, K., Rasmussen, G., and Hamill, J. 1999, Plant Sci., 140, 189.
28. Nitsch, J.P., and Nitsch, C. 1969, Science, 163, 85.
29. White, P.R. 1963, The Cultivation of Animal and Plant Cells, Ronald Press, New York.
30. Lloyd, G., and McCown, B. 1981, Comb. Proc. Intern. Plant Propagators Soc., 30, 421.
31. Muralidharan E. M., and Mascarenhas A.F. 1987, Plant Cell Rep., 6, 256.
32. Muralidharan, E.M., and Mascarenhas, A.F. 1995, Somatic embryogenesis in woody plants, vol. 2, S. Jain, P. Gupta and R. Newton (Eds), Kluwer, Dordrecht, 23.
33. Muralidharan, E.M., Gupta, P.K., and Mascarenhas, A.F. 1989, Plant Cell Rep., 8, 41.
34. Termignoni, R., Wang, P.J., and Hu, C.Y. 1996, Plant Cell Tissue Org. Cult., 45, 129.
35. Watt, M.P., Blakeway, F.C., Termignoni, R., and Jain, S.M. 1999, Somatic embryogenesis in woody plants, vol 5, S.M. Jain, P.K. Gupta and R.J. Newton (Eds), Kluwer, UK, 63.
36. Bandyopadhyay, S., and Hamill, J. 2000, Ann. Bot., 86, 237.
37. Pinto, G., Santos, C., Neves, L., and Araújo, C. 2002, Plant Cell Rep., 21, 208.
38. Kawazu, T., Doi, K., Ohta, T., Shinohara, Y., Ito, K., and Shibata, M. 1990, Abstracts VIIth International Congress on Plant Tissue and Cell Culture, Amsterdam, Netherlands, 64.
39. Teulières, C., Grima-Pettenati, J., Curie, C., Teissie, J., and Boudet, A.M. 1991, Plant Cell Tissue Org. Cult., 25, 125.
40. Sartoretto, L.M., Barrueto Cid, L.P., and Brasileiro, A.C.M. 2002, Funct. Plant Biol., 29, 917.
41. Kawazu, T., Doi, K., Tatemichi, Y., Ito, K., and Shibata, M. 1996, Tree improvement for sustainable tropical forestry, QFRI-IUFRO Conference, Caloundra, Queesland, Australia, 492.
42. Llewelyn, D.J. 1999, Molecular Biology of Woody Plants, Vol. 2, S.M. Jain and S.C. Minocha, Kluwer Academic Publ., Dordrecht, Netherlands, 439.
43. Maunders, M. 1997, Second International Wood Biotechnology Symposium, Canberra, Australia.

44. Chen, Z.-Z., Chang, S.-H., Ho, C.-K., Chen, Y.C., Tsai, J.-B., and Chiang, V. L. 2003, Taiwan J. For. Sci., 16, 249.
45. Hibino, T., Koshiyama, J., Kondo, K., Doi, K., Ishige, N., Asada, T., Sukeno, S-I., and Yamaguchi-Shinozaki, K. 2002, Intern. Plant & Animal Genome Conf. San Diego, CA, USA.
46. Valério, L., Carter, D., and Rodrigues, J.C. 2003, Mol. Breed., 12, 157.
47. Yamada-Watanabe, K., Kawaoka, A., Matsunaga, E., Nanto, K., Sigita, K., Endo, S., Ebinuma, H., and Murata, N. 2003, Tree Biotechnology, IUFRO, Umeå, Sweden, S7.9.
48. Kawaoka, A., Nanto, K., Sugita, K., Endo, S., Yamada-Watanabe, K., Matsunaga, E. and Ebinuma, H. 2003, Tree Biotechnology, IUFRO, Umeå, Sweden, S10.6.
49. Shani, Z., Dekel, M., Cohen, B., Barimboim, N., Kolosovski, N., Safranuvitch, A., Cohen, O., and Shoseyov, O. 2003, Tree Biotechnology, IUFRO, Umeå, Sweden, S10.26.
50. Chriqui, D., Adam, S., Caissard, J.-C., Noin, M., and Azmi, A. 1991, Intensive Forestry: The Role of Eucalypts, Proc. IUFRO Symposium, Durban, South Africa, 70.
51. Purse, J. 1997, Tree Biotechnology Liaison Group 17th annual Meeting, University of Nottingham, Moralejo, 21.
52. Gonzales, E.R., Toro, A.A., Labate, M.T.V., Dias, C.T.S., and Labate, C.A. 1999, Anais do Congresso de Genética e Biologia Molecular, Porto Alegre, Brazil, 677.
53. Gonzales, E.R., Andrade A., Lacerda, G.C., Defávari, V.A.P., Bertolo, A.L., Carneiro, R. T., Labate, M.T.V., and Labate, C.A. 2001, Tree Biotechnology in the New Millenium, IUFRO, Stevenson, Washington, USA.
54. Teasdale, R.D. Patent W098113503.
55. Rochange, F. 1996, Doctorate Thesis, Université Paul Sabatier, Toulouse, France.
56. Gonzales, E.R. 2002, Doctorate thesis. Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, São Paulo, Brazil.
57. Valério, L.A.L. 2000, Doctorate thesis, Universidade Técnica de Lisboa, Instituto Superior de Agronomia, Lisbon, Portugal.
58. Tepfer M., and Casse-Delbart , F. 1987, Microbiol. Sc., 4, 24.
59. Sangwan, R.S., Bourgeois, Y., Brown, S., Vasseur, G., and Sangwan-Norreel, B. 1992, Planta, 188, 439.
60. Cooke, J.E.K., Morse, A.M., and Davis, J.M. 2004, Handbook of Plant Biotechnology, P. Christou and H. Klee (Eds.), John Wiley and Sons, Ltd. Vol. 2, 46, 881.
61. CBD Technologies, Inc. and Companhia Suzano de Papel e Cellulose conclude a strategic agreement. Available HTTP: http://www.cbdtech.com/agreement.html, access on 29/10/2004.
62. Shani, Z., Dekel, M., Cohen, B., Barimboim, N., Cohen, O., Halaly, T., Kolosovski, N., and Shoseyov, O. 2004, Eucalyptus in a changing world, N. Borralho (Ed), Proceedings of the international IUFRO Conference, Aveiro, Portugal, 668.
63. Baucher, M., Monties, B., Van Montagu, M., and Boerjan, W. 1998, Crit. Rev. Plant Sci., 17, 125.
64. Whetten, R.W., MacKay, J.J., and Sederoff, R.R. 1998, Annu. Rev. Plant Physiol. Plant Mol. Biol., 49, 585.
65. Grima-Pettenati, J., and Goffner, D. 1999, Plant Sci., 145, 51.

66. Kyozuka, J., Harcourt, R.L., Peacock, W.J., and Dennis, E.S. 1997, Plant Mol. Biol., 35, 573.
67. Southerton, S.G., Strauss, S.H., Olive, M.R., Harcourt, R.L., Decroocq, V., Zhu, X., Llewellyn, D.J., Peacock, W.J., and Dennis, E.S. 1998, Plant Mol. Biol., 37, 897.
68. Kondo, K., Koshiyama, J., Doi, K., Ishige, N., Hibino, T., Kasuga, M., Shinozaki, K., and Shinozaki, K. 2002, Plant Cell Physiol., 43, S113 Suppl. S.
69. Nippon Paper Group (08/03/2002). Available HTTP: http://www.np-g.com/e/news/news02030801.html, access on 30/10/2004.
70. Hinchee, M. 2003, Tree Biotechnology 2003, IUFRO. Umeå, Sweden, S9.1.
71. Genetically modified eucalyptus grows in acidic soil. Available HTTP: http://www.asahi.com/english/business/K2003071600271.html.
72. Mitsubishi Paper Mills Limited, Pulp & Paper. Available HTTP: http://web.infoweb.ne.jp/mpm/eng/kaihatu/pulp1-e.html.
73. Strauss, S.H., Meilan, R., and DiFazio, S.P. 2004, 6th International Symposium on the Biosafety of GMs. Available HTTP: http://www.ag.usask.ca/isbr/Symposium/Proceedings/Section10.htm.

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



Molecular Biology of Tropical Plants, 2006: 57-74 ISBN: 81-308-0138-8
Editor: Claudine Franche

4 Molecular biology and genomics of the nitrogen-fixing tree *Casuarina glauca*

Hocher, V.[1], Peret, B.[1], Wall, L.G.[2], Obertello, M.[1], Sy, M-O.[1], Santi, C.[1] Svistoonoff, S.[1], Laplaze, L.[1], Auguy, F.[1], Franche, C.[1] and Bogusz, D.[1]
[1]UMR 1098, Institut de Recherche pour le Développement (IRD), BP 64501 911 Avenue Agropolis, 34394 Montpellier cedex 5, France; [2]Departamento de Ciencia y Tecnologia, Universidad Nacional de Quilmas, Roque Saénz Peña 180, bernal (B1876BXD), Buenos Aires, Argentina

Abstract

Actinorhizal species are non-leguminous perennial plants belonging to 8 angiosperm families. They are able to form root nodules as a result of infection by a nitrogen-fixing actinomycete called Frankia. Actinorhizal nodules consist of multiple lobes, each of which represents a modified lateral root with Frankia infected cells in the expanded cortex. This chapter reviews the latest knowledge in molecular biology about this original symbiotic process in Casuarina glauca, a tropical actinorhizal tree belonging to the

Correspondence/Reprint request: Dr. Hocher, V., UMR 1098, Institut de Recherche pour le Développement (IRD), BP 64501, 911 Avenue Agropolis, 34394 Montpellier cedex 5, France. E-mail: valerie.hocher@mpl.ird.fr

Casuarinaceae family. Data on several C. glauca genes expressed during the development and functioning of nodules are discussed. We also review current progress in the contribution of genetic transformation of C. glauca and Allocasuarina verticillata (a closely related species of C. glauca) in exploring plant gene function during the different steps of the development of the symbiotic nodule. We describe the genomic approaches developed in our group with the aim of isolating new actinorhizal symbiotic genes.

Introduction

Two root nodule symbioses are known between nitrogen fixing soil bacteria and higher plants: Legumes associated with *Rhizobia* (including the non-legume *Parasponia*) and so-called actinorhizal plants that interact with *Frankia*. Inside root nodules, the microsymbionts provide the plant with fixed nitrogen and, in exchange, the bacteria are supplied with carbon by the host plant. Although striking differences are observed between Legume/*Rhizobia* and actinorhizal plants /*Frankia* symbiotic systems [1, 2], it has been suggested that both groups of plants belong to the same clade, and thus share a single origin of the predisposition for root nodule symbiosis [3]. Although the symbiosis between *Rhizobia* and Legumes involves more than 1700 plant species of the *Fabaceae* (Legumes) family, few model legume species have been the subject of extensive and detailed studies that resulted in the discovery of the molecules and transducing signal pathways involved in plant-host recognition [4]. Conversely, the symbiotic association between *Frankia* and actinorhizal plants is still poorly understood at the molecular level [5, 6] because of technical difficulties involved in studying the bacteria. Nevertheless different actinorhizal plant species belonging to different actinorhizal plant clades [7] have been described anatomically, histologically, and physiologically, and provided relevant information [8] that can be analyzed when building a molecular model of how actinorhizal symbiosis evolved its own mechanisms to achieve a functional association.

Actinorhizal plants represent about 200 [8] species distributed among 24 genera and 8 angiosperm families. These plants are distributed worldwide, from cool, high latitudes with strong seasonal influences to warm tropical regions with no pronounced difference between seasons [9]. Actinorhizal plants are perennial dicotyledon angiosperms, and are, with the exception of the genera *Dastica*, woody trees or shrubs. Examples of well known genera include *Alnus* (alder), *Eleagnus* (automn olive), *Hippophae* (sea buckthorn) and *Casuarina* (beef wood). Actinorhizal plants are capable of high rates of nitrogen fixation comparable to those found in Legumes [10]. In Egypt, a nitrogen-fixing potential of 288 kg N ha^{-1} has been reported for *Casuarina* [11]. These plants are able to grow in poor and disturbed soil, they are important pioneer species in plant communities worldwide and play an

essential role in land stabilization and soil reclamation [9]. In addition, some actinorhizal species can grow well under a range of environmental stresses such as high salinity, heavy metal and extreme pH [12]. Recognition of current and potential benefits of actinorhizal plant to forestry and agriculture recently focused research on molecular biological aspects [1, 5, 13].

During the last few years our group has concentrated on understanding the plant molecular mechanisms involved in the symbiosis between *Casuarina glauca* and *Frankia*. Molecular tools including a cDNA nodule library, genetic transformation of *Casuarinaceae* and more recently a root/nodule EST databank have been developed [1, 14]. Such tools allowed us to clone and characterize several *Casuarina* plant genes regulated during the symbiotic process.

Here, we present a brief review of the symbiotic partners - *Casuarina*, the host and *Frankia*, the bacteria - and a short description of the morphological and cytological symbiotic events leading to the development of actinorhizal nodules. We also review the current state of knowledge on the molecular biology and genomics of the symbiotic interaction between the tropical actinorhizal tree *Casuarina* and the actinomycete *Frankia*.

Two partners: *Casuarina* and *Frankia*

The *Casuarinaceae* family

The *Casuarinaceae* family includes 90 species of trees and shrubs divided into 4 genera: *Casuarina*, *Allocasuarina*, *Ceuthostoma* and *Gymnostoma* [15]. *Casuarinaceae* are primarily native to the Southern hemisphere, and are found from Australia to South East Asia in tropical, subtropical and temperate coastal regions as well as in arid regions. All members of *Casuarinaceae* are characterized by highly reduced leaves and photosynthetic deciduous branchlets that limit loss of water by evapotranspiration and allow their survival in hot dry areas [15]. They are pioneer species able to colonize severely disturbed sites and contribute to the rehabilitation of these sites by stabilizing the soil and building up its nitrogen content. Some species like *Casuarina glauca* can even grow in a wide variety of soil types, including sandy and saline soils, which has led to the introduction of species belonging to the *Casuarina* and *Allocasuarina* genera in most tropical and sub-tropical areas worldwide and particularly in coastal areas to anchor dunes and to protect crops from wind [16]. Their rapid growth combined with their tolerance of poor fertility and low soil moisture makes them very useful for agroforestry and land reclamation, as well as being valuable sources of fuel wood and charcoal, and generating income for smallholders in tropical countries like India, China and Vietnam [11, 16].

The ability of *Casuarinaceae* to adapt to a range of environmental conditions is due to the exceptional plasticity of their root system, which

enables them to adapt to different environmental stresses. The symbiotic association with the actinomycete *Frankia* allows them to grow in soil deficient in nitrogen. In addition, *Casuarina* roots have a symbiotic relationship with endo- and ectomycorhizal fungi that facilitate the uptake of minerals, notably phosphorus. Furthermore, in phosphorus or iron deficient soil, *Casuarina* roots produce short, densely clustered lateral roots called proteoid (or cluster) roots, which help to absorb phosphorus and other vital minerals insoluble nutrients required for growth and nitrogen fixation [17].

The actinomycete *Frankia*

The microsymbiont *Frankia* is a filamentous, branching, gram-positive actinomycete and is characterized by a slow growth rate, and high G+C DNA content [18, 19]. The first successful isolation of *Frankia* in culture was reported in 1978 [20]. In pure culture, *Frankia* presents three major structures: vegetative hyphae (multiplication form), vesicles that are the site of nitrogen fixation, and sporangia (dissemination form). Due to the lack of genetic tools [21] most aspects of *Frankia* biology, particularly symbiosis, are still unknown [1]. Several trials of genetic transformation, mutagenesis, and functional complementation failed not provide conclusive results [16]. Only the *Frankia* genes involved in nitrogen metabolism have been isolated, and so far, efforts to detect any *Frankia* genes homologous to the *nod* genes of rizobia have failed [22]. Nevertheless preliminary analyses of *Frankia* genome sequences revealed some disperse putative *nod*-like genes although they do not appear to be organized in clusters as in rhizobia and at least *nodA* is not present [P. Normand, personal communication].

Morphological and cytological description of the development of actinorhizal nodules

The morphological steps in the development of actinorhizal nodules have been described in details in several reviews [2, 23, 24].

Infection process

Depending on the host plant, two modes of infection of actinorhizal plants by *Frankia* have been described: intercellular root invasion and intracellular root hair infection [6, 24, 25]. Intracellular infection *via* root hairs (e.g. of *Casuarina, Alnus, Myrica*) starts with root hair curling induced by an unknown *Frankia* signal (Figure 1). Signal exchange between *Frankia* and the host plant has been investigated by several laboratories but the active plant and *Frankia* molecules have not yet been identified [26, 27, 28]. After invagination of growing filaments of *Frankia* into the curled root hairs, infection proceeds intracellularly in the root cortex. *Frankia* hyphae become encapsulated by a

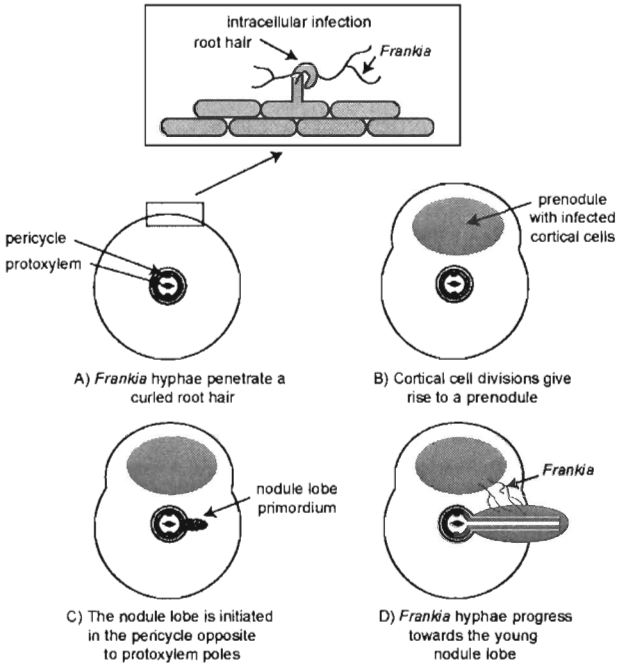


Figure 1. Infection and early organogenesis of a nodule lobe in actinorhizal plants.

cell wall deposit that is believed to consist of xylans, cellulose, and pectins of host origin [29, 30]. At the same time, limited cell divisions occur in the cortex near the invading root hair leading to the formation of a small external protuberance called the prenodule [24]. Infection threads consisting of lines of encapsulated *Frankia* hyphae progress intracellularly toward this mitotically active zone and finally invade most cells of the prenodule [31]. *Frankia* filaments inside the infected plant cells are always surrounded by plant plasma membrane.

As the prenodule develops, cell divisions are induced in pericycle cells opposite a protoxylem pole and give rise to the nodule primordium. While cortical cell divisions lead to the formation of a nodule primordium in Legumes, actinorhizal prenodules do not evolve in nodules. The function of the *C. glauca* prenodule is not yet fully understood but a study of the expression of symbiosis-related genes (*cg12*, *cghb*, see above for details) coupled to cellular modification (cell wall lignification) indicated that the prenodule displays the same characteristics as the nodules and can be considered as a very simple symbiotic organ [32]. The prenodule could thus be

a parallel symbiotic organ on its own or the remaining form of a common nodule ancestor for Legumes and actinorhizal plants [32, 33].

Concerning the intercellular root invasion pathway (e.g. *Discaria, Ceanothus, Elaeagnus*, *Hypophae*), *Frankia* hyphae penetrate between two adjacent rhizoderm cells and progress apoplastically through cortical cells within an electron-dense matrix secreted into the intercellular spaces [34, 35, 36, 37]. Unlike the intracellular mode of infection, no prenodule is formed in the root cortex. Once the nodule primordium has developed from the pericycle, intracellular penetration by *Frankia* and the formation of infection threads is initiated acropetally in developing cortical cells of the nodule lobe primordium, following a pattern similar to that described in plant species invaded through root-hairs.

Nodule development

For both intracellular and intercellular modes of infection, nodule development starts with the induction of mitotic activity in pericycle cells opposite a protoxylem, giving rise to an actinorhizal lobe primordium [10]. An apical meristem is responsible for primordium growth towards the root surface in regions not infected by *Frankia*. The primordium does not incorporate the prenodule but gets infected by hyphae coming from the prenodule [25, 38]. Further development of the primordium gives rise to an indeterminate actinorhizal nodule lobe (Figure 2). New lobes arise continuously to form a coralloid nodule. Mature actinorhizal nodules consists of multiple lobes. In each lobe there

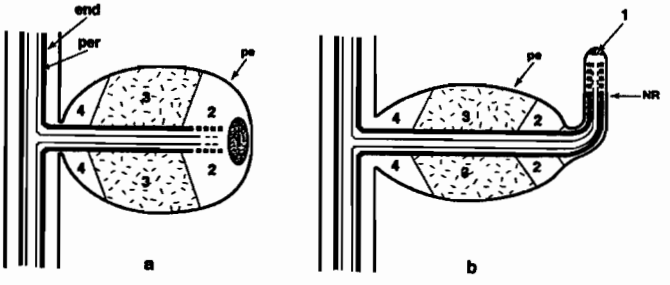


Figure 2. Structure of actinorhizal nodule lobe. Nodule consists of discrete or densely packed lobes. Each nodule lobe is a modified lateral root without root cap, including central vascular tissue, cortical parenchyma infected with *Frankia* and a superficial periderm (pe). A zonation of the cortex with four different zones can be defined : (1) meristem, (2) infection zone, (3) fixation zone, (4) senescence zone. (end) endoderm; (per) pericycle. a : *Alnus* type lobe. b : *Myrica, Casuarina* type lobe. These lobes exhibit a nodule root (NR) at the apex of nodule lobe. Nodule roots are devoid of *Frankia* hyphae.

is a central vascular bundle surrounded by an endoderm, an expanded cortex and a periderm. *Frankia* is restricted to the cortical cells. Some species like *Casuarina* or *Myrica* develop a so-called root nodule at the apex of each lobe [25] (Figure 2b). This root nodule lacks root hairs, has a reduced root cap and displays negative geotropism. It might be involved in the diffusion of gas, especially oxygen, in and out of the nodule lobe [39].

Nodule functioning

Four zones have been morphologically defined in studies of nodules originating from intracellular infection [25, 40] or intercellular root invasion [41] and gene expression [42] (Figure 2). (i) The **apical meristem** is free of *Frankia*. (ii) Adjacent to the meristem is an **infection zone** where some of the young cortical cells resulting from the meristem activity are infected by *Frankia*. The bacterium starts to proliferate but remains encapsulated in a plant-derived matrix, and the plant cells enlarge; (iii) the subsequent **fixation zone** contains both infected and uninfected cortical cells. Infected cells are hypertrophied and are filled with *Frankia* filaments that differentiate vesicles where nitrogen fixation takes place. The appearance and shape of these vesicles are controlled by the plant. In some species like *Casuarina*, infected cells have a lignified cell wall and there is no vesicle differentiation. Uninfected cells are smaller and in some species contain amyloplast and phenolic compounds, and might be involved in nitrogen and carbon metabolism. Finally a basal **senescence zone** (iv) is observed in old nodules; plant cells and bacteria degenerate and nitrogen fixation is switched off. More recently, a second level of compartmentation was described in *Casuarina glauca* nodules based on the accumulation of flavans, which occurs in uninfected cells in the endodermis and in the cortex. These cells form layers that delimit *Frankia* infected compartments in the nodule lobe and may play a role in restricting bacterial infection to certain zones of the nodule [43].

Molecular events that occur during *C. glauca-Frankia* symbiosis

During differentiation of the symbiotic actinorhizal root nodule, a set of genes -called actinorhizal nodulin genes- is activated in the developing nodules [44, 45] (Figure 3). Similarly to Legumes, two major types of actinorhizal nodulin genes have been defined by their pattern of expression and function. Early nodulin genes are expressed before the beginning of nitrogen fixation; they are thought to be involved in plant infection or in nodule organogenesis whereas late nodulin genes comprise sequences involved in different metabolic activities necessary for the functioning of the nodule.

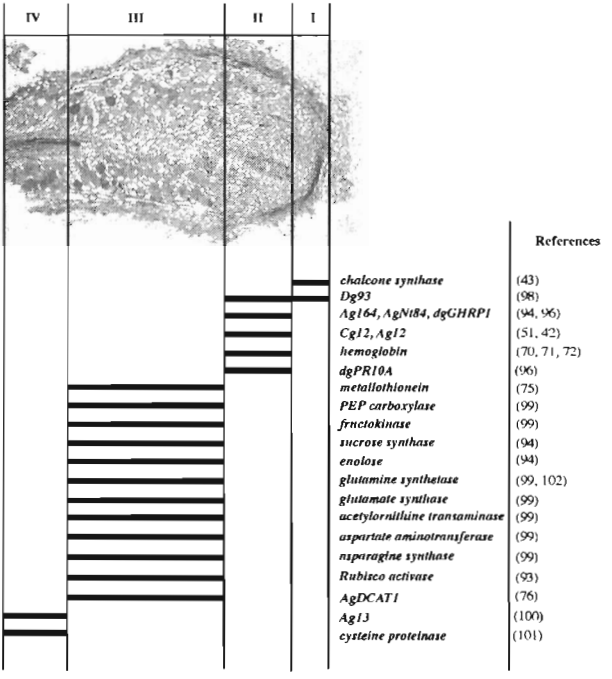


Figure 3. Gene expression map in different zones of actinorhizal root nodules. Nodule zones are indicated. I meristem zone, II infection zone, III nitrogen fixation zone, IV senescence zone. Black bars indicate the presence of mRNA transcripts. *Dg: Datisca glomerata, Ag: Alnus glutinosa, Cg: Casuarina glauca.* (modified from Obertello *et al.* (13))

Casuarina glauca is a good model for studying symbiotic gene expression as it is the only actinorhizal species with *Allocasuarina* that can be genetically transformed. Using *Agrobacterium* as a biological vector for gene transfer, transgenic plants have been recovered for both *C. glauca* and *Allocasuarina verticillata* [46, 47, 48, 49]. These transgenic *Casuarinaceae* trees provide valuable tools, first, to investigate the molecular mechanisms involved in actinorhizal symbiosis and, second, to establish comparisons with Legumes [13].

cg12 an early expressed symbiotic gene in *C. glauca*

cg12 is an actinorhizal symbiotic gene isolated from *C. glauca* [50] that is homologous to *ag12* previously described for *A. glutinosa* [42]. As described for *ag12*, *cg12* encodes for a subtilisin-like serine protease (subtilases) and is specifically expressed during plant cell infection in young prenodule and

nodule infected cells just before plant cells differentiate to fix nitrogen [50]. The regulation of *cg12* expression and its possible role during actinorhizal nodule infection were investigated with a transgenic approach by introducing *cg12* promoter-reporter fusions in *Allocasuarina verticillata* and in *Casuarina glauca*. Expression of the reporter gene was observed during the first steps of the infection process, i.e. when *Frankia* was invading deformed root hairs and in root and nodule cortical cells containing growing infection threads. *cg12* expression seems to be correlated with plant cell invasion by the endosymbiont from the very start of the symbiotic process [51]. A study of the promoter expression of *ara12*, the *Arabidopsis* homologue of *cg12*, revealed expression in roots and shoots and in developing trichomes and siliques suggesting a role in cell elongation and/or differentiation [52].

Subtilases are a super-family of proteases and are thought to play a role in several different aspects of plant development including epidermal surface formation and stomatal density and distribution in *Arabidopsis* [53, 54], response to pathogens [55], lateral root development [56], and microsporogenesis [57]. Using anti-CG12 polyclonal antibodies our group recently investigated CG12 cytolocalization. The CG12 protein was only detected in nodules in *Frankia*-infected cells. Furthermore, microscopical observations showed that CG12 was associated with the plant cell wall and the polysaccharidic matrix surrounding *Frankia* filaments. Although the implication of the localization of CG12 in this compartment is not yet understood, it has been suggested that CG12 may play a role in the maturation of a polypeptide involved in signalling cascades activated upon *Frankia* infection [58].

enod 40: Involvement in actinorhizal symbiosis

enod40 is an early nodulin gene first isolated from soybean [59]. In Legumes, *enod40* genes are highly conserved and are key genes for nodule organogenesis and a limiting factor in nodule development [60]. They also play a role in mycorrhizal symbiosis [61]. *enod40* genes encode transcripts of about 0.7 kb that are characterized by the absence of a long open reading frame (ORF); they all contain two conserved regions, named regions I and II [62]. A small ORF encoding a peptide of 12 or 13 amino acids has been identified in region I and the translation of an ORF spanning region II has been demonstrated to be necessary for the biological activity of *ENOD40* [63]. In Legumes, *enod40* expression is induced at a very early stage by nodulation factors, and is localized in the vascular system of roots, shoots, mature nodules as well as in nodule primordia [60, 62]. Recent work has revealed that *enod40* encodes two peptides that bind to sucrose synthase which suggests a role in increasing phloem unloading and/or sink strength determination to induce nodulation [60, 64].

A homolog of *enod40* was isolated from *C. glauca* (*cgenod40*) [65]. Sequence comparison with other *ENOD40* from Legumes and non Legumes revealed that in addition to significant similarities, the ORF peptide in region I was lacking in both *C. glauca* and *A. glutinosa,* another actinorhizal tree. RNA gel blot analysis revealed a lower level of *cgenod40* expression in actinorhizal nodules than that observed in Legume nodules. Expression of *cgenod40-gus* fusion was then studied in transgenic *A. verticillata* and *C. glauca* and expression in the vascular tissue of the roots, shoots and nodules was observed. However expression was found neither in the early stages of infection by *Frankia* including prenodules and nodule primordia, nor in response to nod factors [65]. These results are different from the scheme described in Legumes and suggest *enod40* plays a different role in actinorhizal plants. In Legumes, phloem unloading is mostly apoplastic in the root nodulation zone, but mostly symplastic (due to a lignified root system) in actinorhizal species, thus explaining why *enod40* is not involved in nodule induction [65]. However the role of *enod40* in actinorhizal symbiosis is not yet understood.

Late actinorhizal nodulins

Hemoglobin

Biological nitrogen fixation is an ancient biochemical process that evolved before photosynthesis and it is absolutely O_2 sensitive. At the same time the reduction of N_2 to NH_3 consumes a lot of energy in the cell, so O_2 is useful to generate ATP. Different strategies have evolved in nature to handle this paradox, and examples can be found in the diversity of nitrogen fixing microorganisms. Actinorhizal symbiosis also show different nodule anatomy development as different solutions for this physiological problem [39]. Except in *Casuarina* and *Allocasuarina*, when associated with actinorhizal plants, *Frankia* forms vesicles that limit O_2 diffusion to protect nitrogenase. In *C. glauca* nodules, an oxygen diffusion barrier is created by lignification of the cell wall of the infected and adjacent uninfected cortical cells [66]. A large amount of the O_2-transport protein hemoglobin (hb) has also been found in *Casuarina* nodules [67]; the purified protein was shown to be similar to the legume leghemoglobin, thus suggesting a similar function [68]. The large amount of hb and the lignified cell walls of infected cells are consistent with the absence of *Frankia* vesicles in *Casuarina* nodules. Symbiotic hb genes [69] and a corresponding cDNA were isolated from *Casuarina glauca*. Localisation of hb mRNA in nodules by *in situ* hybridization showed that the corresponding *Hb* symbiotic genes are induced in young infected cells prior to the detection of *Frankia nifH* mRNA suggesting that hb contributes to reducing O_2 tension before *nif* gene expression [70]. In *C. glauca* nodules it has been demonstrated by immunogold localisation that hb is found in the cytoplasm and nuclei of infected host cells and is not associated with *Frankia* membranes. Thus, in

Casuarina it seems that, just like in the nodules of Legumes, O_2 regulation is mediated by a host-derived O_2 diffusion barrier and O_2 transport protein. It is interesting to note that hb was found in nodules of *Myrica gale* [71] and *A. glutinosa* [72] where *Frankia* vesicles are present. This suggests that even in the presence of vesicles, symbiotic hb ensures the flow of O_2 within infected cells.

Metallothioneins

Metallothioneins (MTs) are defined as low molecular weight cysteine-rich proteins that can bind heavy metals and may play a role in their intracellular sequestration and transportation. Although their exact function remains unclear, plant metallothioneins are thought to be involved in response to stresses like wounding, pathogen infection, and leaf senescence [73]. It has recently been argued that they also function as antioxidants and play a role in plasma membrane repair [74].

A clone for type I metallothionein (*cgMT1*) was isolated from a *C. glauca* nodule cDNA library [75]. *In situ* hybridization revealed localisation of the transcripts in mature *Frankia* infected cells and in the pericycle. The *gus* gene under the control of the *cgMT1* promoter was introduced into *Casuarina* and *Allocasuarina*. In transgenic plants the *cgMT1* promoter was shown to be primarily active in large *Frankia* infected cells of the nodule nitrogen-fixing zone, in roots, and in the oldest parts of the shoots. Induction experiments performed on transgenic *Arabidopsis* plants carrying the P*cgMT1-gus* construct revealed that the promoter P*cgMT1* responded to wounding, oxidative stress and pathogen infection. Our current hypotheses is that the metallothionein gene *cgMT1* could be involved in metal ion transport required for nitrogenase function in nodules, in metal homeostasis in roots, and/or in antioxidant defence against reactive oxygen species induced during the symbiotic process [75].

Other proteins involved in nodule physiology

Several actinorhizal nodulin genes encoding enzymes involved in nitrogen and carbon metabolism have been characterized in different actinorhizal species [For recent reviews see 13, 32, 33]. More recently, the isolation of a nodule-specific dicarboxylase transporter in *A. glutinosa* nodules was reported, which may be involved in carbon metabolism [76]. In *C. glauca*, our group isolated a cDNA encoding for a chalcone synthase (chs), the corresponding mRNA was localized in the flavan-containing cells of the apex of the nodule lobe. Since chalcone synthase is a key enzyme in the flavonoid biosynthesis pathway, our data suggest that flavonoid synthesis depends on the developmental stage of the cells within the nodule lobe [43]. A cDNA corresponding to an Acyl Carrier Protein was also isolated from the *C. glauca*

nodule cDNA library which could be involved in fatty acid biosynthesis occurring during plant cell infection by *Frankia* [77].

The actinorhizal nodule, a modified lateral root

Legume nodules have a stem-like anatomy with peripheral vascular bundles and infected cells in the central tissue and they originate in the root cortex. In contrast, actinorhizal nodules have the same origin and structure as lateral roots [5]. Thus, we wonder to what extent lateral root and actinorhizal nodule can be compared.

Comparison between actinorhizal nodule and lateral root development

Actinorhizal nodule and lateral root development have features in common. Both organs originate from divisions in pericycle cells situated in front of a xylem pole. Moreover, both nodule and lateral root vasculatures are central, in contrast with legume nodule vasculature which is peripheral. In addition, some actinorhizal nodules (*e.g. Casuarina glauca* nodules) show a so-called "nodule root" at their apex, highlighting the indeterminate growth characteristic of these nodules. Because of their common developmental origin, their similar structure and the presence of the nodule root, actinorhizal nodules are considered to be modified lateral roots.

Interestingly, the distribution of lateral roots remains unchanged in nodulated *Alnus glutinosa* plants [40] and also in nodulated *Discaria trinevis* [78]. This suggests that the infection by *Frankia* does not "hijack" a young forming lateral root to produce a nodule but induces *de novo* nodule formation. It is very tempting to speculate that part of the lateral root genetic program has been recycled during evolution to create the nodule genetic program. To what extent the lateral root genetic program is used to complete nodule formation remains to be determined. To this end, we are currently studying the expression of genes specifically involved in lateral root development during the course of nodule organogenesis.

One of these genes, *HRGPnt3,* encodes a plant-cell wall protein expressed during early stages of *Nicotiana tabacum* lateral root development [79, 80]. The promoter of this gene was fused to the β-glucuronidase coding sequence and used as a molecular marker for lateral root development in the actinorhizal tree *A. verticillata*. Unfortunately, no GUS activity was detected either in lateral roots or in nodules suggesting that the regulation of this gene is not conserved between the two species. The isolation from actinorhizal plants of homologs of genes known to be involved in lateral root development in model species should help us to further compare nodule and lateral root development.

Involvement of auxin

Considering the major role of auxin in lateral root formation [81, 82] one would expect this hormone to play a key role in actinorhizal nodule formation. Indeed, it has long been known that nodulated roots contain high levels of auxin compared to non-nodulated roots [83, 84]. Moreover, some *Frankia* strains secrete natural auxins in culture such as phenylacetic acid (PAA) and indolacetic acid (IAA) [85] which are thought to be involved in nodule induction. It has also been suggested that this auxin production by the endosymbiont contributes to the differentiation of the hypertrophied *Frankia* infected cells.

Genes from the AUX1-family are involved in auxin influx transport (entry of auxin into the cell) which is known to be important for lateral root formation [86]. *Arabidopsis* mutants for two of the four members of the gene family, namely *aux1* and *lax3*, have half the number of lateral roots than the wild type. We recently cloned *C. glauca* homologs of *Arabidopsis aux1* and *lax3* and we are in the process of comparing the expression patterns of these genes during lateral root and nodule development. These auxin influx transporters might also play an important role during the early stages of the symbiosis by enabling the perception of bacterial auxins. The use of molecular markers of *in situ* auxin accumulation such as *DR5-gus* [87] and *iaa2-gus* [88] should help to compare the localisation of influx transporters and auxin flux in these organs thus enabling us to better understand the role of this hormone in nodule formation.

Looking for new early expressed genes: Analysis of *Casuarina glauca* EST banks

The early molecular mechanisms involved in the *Casuarina–Frankia* symbiosis are still poorly understood. Besides the differential hybridization approach, we recently developed a more global non-targeted approach by means of expressed sequence tag (EST) analysis [14].

A total of 3 000 ESTs were obtained from cDNA libraries corresponding to mRNA extracted from (1) young nodules induced by *Frankia* and (2) non infected roots. The raw EST sequences obtained were stored in an in-house database and an automatic treatment pipeline was designed to analyze and annotate them. 70% of the sequences (root and nodule) were considered of high quality and were submitted to a clustering program in order to eliminate redundant ESTs. Each EST or cluster was annotated using the BLAST algorithm by sequence comparison against known proteins of non-redundant database (SWISSPROT, Trembl and PIR) [89]. The e value was fixed at 10^{-5}. Around 60% of root and 40% of nodule sequences (ESTs and clusters) were annotated and the identified sequences were subsequently assigned to 14

functional categories on the basis of the classification developed for the *Medicago truncatula* EST databank [90, 91]. For both nodulated and non nodulated root ESTs, the largest predominant categories were "protein synthesis" and "primary metabolism". It is worth noting that these categories were also described as being predominant for *M. truncatula* EST [91]. The three largest predominant categories in nodule were: "cell division", "vesicular and cell trafficking", and "defence and cell rescue". This may reflect the development of the nodular structures and the induction of defence genes upon the infection of plant cells by the actinomycete [90, 91, 92]. Not surprisingly, in the nodule EST database, several EST/cluster sequences corresponded to proteins previously described as actinorhizal nodulins. For example, the following ESTs corresponded to identified actinorhizal nodulins genes: hemoglobin [70], metallothioneins [75], subtilisin [51], rubisco activase [93], saccharose synthase [94], glycine and histidine rich proteins [95]. A set of nodule specific sequences was selected and a study of their expression profiles during early symbiotic events is underway. Furthermore, we developed a subtractive hybridization approach using 24 h infected roots *versus* non induced roots to generate nodule sequences of genes that are expressed very early.

Conclusion

The input of fixed nitrogen by actinorhizal plants on a global scale is enormous; they contribute 15% of symbiotic nitrogen fixation. *Casuarinaceae* species are widely distributed and contribute to maintaining/rehabilitating marginal lands, as well as to providing incomes for smallholders in different tropical and sub tropical countries. Understanding the development and functioning of actinorhizal nodules is thus an important challenge. In the past decade, molecular tools have been developed and considerable advances have been made in the identification and characterization of genes involved in actinorhizal symbiosis. The genetic transformation procedures developed for *Casuarinaceae* made it possible to perform functional analysis of the isolated symbiotic genes. However, our understanding of the early events occurring when the *Casuarina-Frankia* symbiosis takes place is still poor. For example, nodulation signals produced by *Frankia* and plant factors required for the initiation of nodule morphogenesis have not been described to date. Emerging genomic resources such as EST libraries, *Frankia* genome sequences [P. Normand and L. Tisa, personal communication] have profound implications for the study of actinorhizal symbioses and may reveal novel mechanisms of plant-microbe recognition.

Which specific properties permitted actinorhizal plants to form root nodules induced by the nitrogen-fixing actinomycete *Frankia*? Recent phylogenetic studies suggest a single origin for the predisposition to form Legumes/*Rhizobium*

and actinorhizal plants/*Frankia* symbioses [96]. Using the transgenic *Casuarinaceae*/reporter gene approach our group has shown that common mechanisms of transcriptional gene regulation activated during bacterial infection and nodule functioning may be part of the common heritage [97].

Acknowledgements

The majority of the work cited in this chapter was funded by IRD. B. Peret is granted by the MENRT (France) and M. Obertello is granted by DSF, IRD (France).

References

1. Franche, C., Laplaze, L., Duhoux, E., and Bogusz, D. 1998, Crit. Rev. Plant Sci., 17, 1-28.
2. Wall, L.G. 2000, J. Plant Growth Regul., 19, 167-182.
3. Soltis, D.E., Soltis, P.S., Morgan, D.R., Swensen, S.M., Mullin, B.C., Dowd, J.M., and Martin, P.G. 1995, Proc. Natl. Acad. Sci. U.S.A., 92, 2647-2651.
4. Cohn, J., Day, B.R., and Stacey, G. 1998, Trends Plant Sci., 3, 105-110.
5. Pawlowski, K., and Bisseling, T. 1996, Plant Cell, 6, 1899-1913.
6. Wall, L.G., and Berry, A.L. 2005, Actinorhizal Symbioses, K. Pawlowski, W Newton (eds.), Springer Verlag. (in press).
7. Swensen, S.M., and Mullin, B.C. 1997, Physiol. Plant., 99, 565-573.
8. Huss-Danell, K. 1997, New Phytol., 136, 375-405.
9. Spent, J.I., and Parsons R. 2000, Field Crops, 65, 183-196.
10. Torrey, J.G. 1976, Amer. J. Bot., 63, 335-345.
11. Diem, H.G., and Dommergues, Y.D. 1990, The Biology of *Frankia* and Actinorhizal Plants, C.R. Schwintzer, J.D. Tjepkema (eds), Academic Press, New York, 317-342.
12. Dawson, J.O. 1990, The Biology of *Frankia* and Actinorhizal Plants, C.R. Schwintzer, J.D. Tjepkema (eds), Academic Press, New York, 299-316.
13. Obertello, M., Sy, M.O., Laplaze, L., Santi, C., Svistoonoff, S., Auguy, F., Bogusz, D., and Franche, C. 2003, African J. Biotech., 2, 528-538.
14. Hocher, V., Argout, X., Auguy, F., Svistoonoff, S., Laplaze, L. and Bogusz, D. 2004, 6th European Nitrogen Fixation Conference, Toulouse, France, July 24-27 2004, abstract.
15. National Research Council 1984. *Casuarinas*: nitrogen-fixing trees for adverse sites, National Academic Press, Washington DC, USA.
16. Duhoux, E., and Franche, C. 2003, Biofutur, 235, 45-49.
17. Arahou, M., and Diem, H.G. 1997, Plant Soil, 196, 71-79.
18. Benson, D.R., and Silvester, W.B. 1993, Microbiol Rev., 57, 297-319.
19. Simonet, P., Normand, P., Hirch, M., and Akkermans, A.D.L. 1990, Molecular Biology of Symbiotic Nitrogen Fixation P.M. Gresshoff (ed), CRC Press, Bocaraton, 77-109.
20. Callaham, D., Del Tredia, P., and Torrey, J.G. 1978, Science, 199, 899-902.
21. Mullin, B.C., and Dobritsa, S.V. 1996, Plant Soil, 186, 9-20.
22. Cérémonie, H., Cournoyer, B., Maillet, F., Normand, P., and Fernandez, M.P. 1998, Mol. Gen. Genet., 260, 115-119.

23. Newcomb, W.R., and Wood S. 1987, Int. Rev. Cytol., 109, 1-88.
24. Berry, A.L., and Sunnel, L.A. 1990, The Biology of *Frankia* and Actinorhizal Plants C.R. Schwintzer, J.D. Tjepkema (eds), Academic Press, New York, 61-81.
25. Duhoux, E., Diouf, D., Gherbi, H., Franche, C., Ahée, J., and Bogusz, D. 1996, Act. Bot. Gall., 143, 593-608.
26. Prin, Y., and Rougier, M. 1987, Plant Physiol., 6, 99-106.
27. van Ghelue, M., Lovaas, E., Ringo, E., and Solheim, B. 1997, Physiol. Plant., 99, 579- 587.
28. Cérémonie, H., Debellé, F., and Fernandez, M.P. 1999, Can. J. Bot., 77, 1293-1301.
29. Lalonde, M., and Knowles, R. 1975, Can. J. Bot., 5, 1951-1971.
30. Berg, R.H. 1990, Protoplasma, 159, 35-43.
31. Berg, R. H. 1999, Can. J. Bot., 77, 1327-1333.
32. Laplaze, L., Svistoonoff, S., Santi, C., Auguy, F., Franche, C., and Bogusz, D. Nitrogen fixation research: origins and progress Vol. VI: Actinorhizal symbioses. W.E. Newton (ed). Kluwer, in press
33. Gualtieri , G. and Bisseling, T. 2000, Plant Mol. Biol., 42, 181-194.
34. Miller, I.M., and Baker, D.D. 1985, Protoplasma, 128, 107-119.
35. Racette, S., and Torrey, J.G. 1989, Can. J. Bot., 67, 2873-2879.
36. Valverde, C., and Wall, L.G. 1999, New Phytol., 141, 345-354.
37. Liu, Q., and Berry, A.M. 1991, Protoplasma, 163, 82-92.
38. Wall L.G., and Berry, A.L. Nitrogen fixation research: origins and progress Vol. VI: Actinorhizal symbioses. W.E. Newton (ed). Kluwer, in press
39. Silvester, W.B., Harris, S.L., and Tjepkema, J.D. 1990, The Biology of *Frankia* and Actinorhizal Plants. C.R. Schwintzer, J.D. Tjepkema (eds), Academic Press, New York, 157-176.
40. Angulo Carmona, A.F. 1974,. Acta Bot. Neerl., 23, 257-303.
41. Valverde, C. and Wall, L.G. 1999. New Phytol., 141, 345-354.
42. Ribeiro, A., Akkermans, A.D.L, van Kammen, A., Bisseling, T., and Pawlowski, K. 1995, Plant Cell, 7, 785-794.
43. Laplaze, L., Gherbi, H., Frutz, T., Pawlowski, K., Franche, C., Macheix, J-J., Auguy, F., Bogusz, D., and Duhoux, E. 1999, Plant Physiol., 121, 113-122.
44. Mullin, B.C., and Dobritsa, S.V. 1996, Plant Soil, 186, 9-20.
45. Pawlowski, K. 1997, Physiol. Plant., 99, 617-631.
46. Diouf, D., Gherbi, H., Prin, Y., Franche, C., Duhoux, E., and Bogusz, D. 1995, Mol. Plant Microbe Interac., 8, 532-537.
47. Franche, C., Diouf, D., Le, Q.V., N'Diaye, A., Gherbi, H., Bogusz, D., Gobé, C., and Duhoux, E. 1997, Plant J., 11, 897-904.
48. Smouni, A., Laplaze, L., Bogusz, D., Guermache, F., Auguy, F., Duhoux, E., and Franche, C. 2002, Funct. Plant Biol., 29, 649-656.
49. Santi, C., Svistoonoff, S., Constans, L., Auguy, F., Duhoux, E., Bogusz, D. and Franche, C. 2003, Plant Soil, 254, 229–237.
50. Laplaze, L., Ribeiro, A., Franche, C., Duhoux, E., Auguy, F., Bogusz, D., and Pawlowski, K. 2000, Mol. Plant-Microbes Int., 13, 113-117.
51. Svistoonoff, S., Laplaze, L., Auguy, F., Runions, J., Duponnois, R., Haseloff, J., Franche, C., and Bogusz, D. 2003, Mol. Plant-Microbes Int., 16, 600-607.
52. Svistoonoff, S., Laplaze, L., Auguy, F., Santi, C., Fontanillas E., Duhoux, E., Franche C., and Bogusz D. 2003, Plant Soil, 254, 239-244.

53. Tanaka, H., Onouchi, H., Kondo, M., Hara-Nishimura, I., Nishimura, M., Machida, C. and Machida, Y. 2001, Development, 128, 4681-4689.
54. Berger, D., and Altmann, T. 2000, Genes Dev., 1, 1119-1131.
55. Jordá, L., Coego, A., Conejero, V., and Vera, P. 1999, J. Biol. Chem., 22, 2360-2365.
56. Neuteboom, L.W., Veth-Tello, L.M., Clijdesdale, O.R., Hooykaas, P.J.J., and van der Zall, B.J. 1999, DNA Res., 26, 13-19.
57. Taylor, A.A., Horsch, A., Zepczyk, A., Hasenkampf, C.A., and Riggs, C.D. 1997, Plant J., 12, 1261-1271.
58. Svistoonoff *et al.*, Planta , submitted.
59. Yang, W., Katinakis, P., Hendriks, P., Smolders, A., de Vries, F., Spee, J., van Kammen, A., Bisseling, T., and Franssen, H. 1993, Plant J., 3: 573-585.
60. Charon, C., Sousa, C., Crespi, M., and Kondorosi, A. 1999,. Plant Cell, 11, 1953-1965.
61. Staehelin, A., Charon, C., Boller, T., Crespi, M., and Kondorosi, A. 2001, Proc. Natl. Acad. Sci. USA, 98, 15366- 15371.
62. Crespi, M., Jurkevitch, E., Poiret, M., d'Aubenton-Carafa, Y., Petrovics, G., Kondorosi, E., and Kondorosi, A. 1994, EMBO J., 13, 5099-5112.
63. Sousa, A., Johansson, C., Charon, C., Manyani, H., Saulter C., Kondorosi, A., and Crespi, M. 2001, Mol. Cell. Biol., 21, 354-366.
64. Röhrig, H., Schmidt, J., Miklashevichs, E., Schell, J., and John, M. 2002, Proc. Natl. Acad. Sci. USA, 99, 1915-1920.
65. Santi, C., von Groll, U., Chiurazzi, M., Auguy, F., Bogusz, D., Franche, C., Pawlowski, K. 2003, Mol. Plant-Microbes Int., 16, 808-816.
66. Berg, R.H., and Mc Dowell, L. 1988, Can. J. Bot., 66, 354-366.
67. Fleming, A.I., Wittenberg, J.B., Wittenber, B.A., Dudman, W.F., and Appleby, C.A. 1987, Biochem. Biophys. Acta, 911, 209-220.
68. Gibson, Q.H., Wittenberg, J.B., Wittenberg, B.A., Bogusz, D., and Appleby, C.A. 1989, J. Biol. Chem., 264, 100-107.
69. Jacobsen-Lyon, K., Jensen, E.O., Jorgensen, J-E., Marcker, K.A., Peacock, W.J., and Dennis, E.S. 1995, Plant Cell, 7, 213-222.
70. Gherbi, H., Duhoux, E., Franche, C., Pawlowski, K., Nassar A, Berry A., and Bogusz D. 1997, Physiol. Plant., 99, 608-616.
71. Pathirana, S.M., and Tjepkema, J.D. 1995, Plant Physiol., 107, 827-831.
72. Suharjo, U.K.J., and Tjepkema, J.D. 1995, Physiol. Plant., 95, 247-252.
73. Cobbett, C, and Goldsbrough P. 2002, Ann. Rev. Plant Biol., 53, 159–182.
74. Hall, J.L. 2002, J. Exp. Bot., 53, 1–11.
75. Laplaze, L., Gherbi, H., Duhoux, E., Pawlowski, K., Auguy, F., Guermache, F., Franche, C., and Bogusz, D. 2002, Plant Mol. Biol., 49, 81-92.
76. Jeong, J., Suh, S., Guan, C., Tsay, X.F., Moran, N., Oh, C., An, C.S., Demchenko, K.N., Pawlowski, K., Lee, Y. 2004, Plant Physiol., 134, 969-978.
77. Laplaze, L., Gherbi, H., Franche, C., Duhoux, E., and Bogusz, D. 1998, Plant Physiol., 116, 1605.
78. Valverde, C., and Wall, L.G. 1999, Can. J. Bot., 77, 1302-1310.
79. Keller B., Lamb C.J. 1989, Mol. Plant-Microbes Int., 15, 209-215.
80. Vera P., Lamb C., Doerner P.W. 1994, Plant J., 6, 717-727.
81. Casimiro, I., Marchant, A., Bhalerao, R.P., Beeckman, T., Dhooge, S., Swarup, R., Graham, N., Inze, D., Sandberg, G., Casero, P.J., and Bennett M.J. 2001, Plant Cell., 13, 843-852.

82. Bhalerao, R.P., Eklo, J., Ljung, K., Marchant, A., Bennett, M., and Sandberg, G. 2002, Plant J., 29, 325-332.
83. Dullaart, J. 1970, J. Exp. Bot., 21, 975-984.
84. Henson, I.E., and Wheeler C.T. 1977, J. Exp. Bot., 28, 1076-1086.
85. Wheeler, C.T., Henson, I.E., and MacLaughlin M.E. 1979, Bot. Gaz., 140, 52-57.
86. Reed, R.C., Brady, S.R., and Muday, G.K. 1998, Plant Physiol., 118, 1369–1378.
87. Ulmasof, T., Murfett, J., Hagen, G., and Guilfoyle, T.J. 1997, Plant Cell, 9, 1963-1971.
88. Luschnig, C., Gaxiola, R., Grisafi, P., and Fink, G. 1998, Genes Dev., 12 , 2175-2187.
89. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990, J. Mol. Biol. 215, 403-410.
90. Covitz, P. A., Smith, L. S., and Long, S. R. 1998, Plant Physiol., 117, 1325-1332.
91. Journet, E-P., van Tuinen, D., Gouzy, J., Crespeau, H., Carreau, V., Farmer, M-J., Niebel, A., Schiex, T., Jaillon, O., Chatagnier, O., Godiard, L., Micheli, F., Kahn, D., Gianinazzi-Pearson, V., and Gamas, P. 2002, Nucl. Acid Res., 30, 5579-5592.
92. Fedorova, M., van de Mortel, J., Matsumoto, P.A., Cho, J., Town, C.D., Van den Bosch, K.A., Gantt, J.S., and Vance, C.P. 2002, Plant Physiol., 130, 519-537.
93. Okubara, P. A., Pawlowski, K., Murphy, T.M., and Berry, A.M. 1999, Plant Physiol., 120, 411-420.
94. Van Ghelue, M., Ribeiro, A., Solheim, B., Akkermans, A.D.L., Bisseling, T., and Pawlowski, K. 1996, Mol. Gen. Genet., 250, 437-446.
95. Pawlowski, K., Twigg, P., Dobritsa, S., Guan, C-H., and Mulin, B.C. 1997, Mol. Plant-Microbes Int., 10, 656 -664.
96. Pawlowski K., Swensen, S., Guan, C., Hadri, A-E., Berry, A.M., and Bisseling, T. 2003, Mol. Plant-Microbes Int., 16, 796-807.
97. Svistoonoff, S., Laplaze, L., Liang, J., Ribeiro, A., Gouveia, M.C., Auguy, F., Fevereiro, P., Franche, C., and Bogusz D. 2004, Plant Physiol., 136, 3191-3197.
98. Okubara, P. A., Fujishige, N. A., Hirsch, A. M., and Berry, A. M. 2000, Plant Physiol., 122, 1073-1079.
99. Guan C., Wolters, D.J., van Dijk, C., Akkermans, A.D.L., van Kammen, A., Bisseling, T., and Pawlowski, K. 1996, Act. Bot. Gall., 143, 613-620.
100. Guan C., Akkermans, A.D.L., van Kammen, A., Bisseling, T., and Pawlowski, K. 1997, Physiol. Plant., 99, 601-607.
101. Goetting-Mineski, M.P., and Mullin, B. 1994, Proc. Natl. Acad. Sci. USA, 91, 9891-9895.
102. Berry, A.M., Murphy, T.M., Okubara, P.A., Jacobsen, K.R., Swensen, S.M., and Pawlowski, K. 2004, Plant Physiol., 135, 1849-1862.

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

Molecular Biology of Tropical Plants, 2006: 75-103 ISBN: 81-308-0138-8
Editor: Claudine Franche

5 Genome mapping in tropical tree species

Siming Gan, Fagen Li and Mei Li
Research Institute of Tropical Forestry, Chinese Academy of Forestry
Longdong, Guangzhou 510520, Guangdong Province
People's Republic of China

Abstract

Genome mapping is a very active research area in tree genetics. It has focused on construction of three types of map, namely, the genetic map, the physical map, and the complete DNA sequence map. However, for tropical trees most of the progress has been made in construction of genetic maps. Genetic maps with large numbers of genetic markers have been constructed for more than 10 tropical tree species including tropical eucalypts, tropical pines, acacia, coffees, and rubber trees. Meanwhile, in this mapping work several issues have emerged including distorted segregation of markers, availability of suitable DNA markers, low density of mapped markers, and slow

Correspondence/Reprint request: Dr. Siming Gan, Research Institute of Tropical Forestry, Chinese Academy of Forestry, Longdong, Guangzhou 510520, Guangdong Province, People's Republic of China
E-mail: smggan@pub.guangzhou.gd.cn

progress in comparative mapping. The genetic maps have been used to characterize agronomically important QTLs in a variety of tropical trees. More recently physical maps have been initiated for a tropical eucalypt species and coffee, while whole genome sequencing is in progress for Eucalyptus camaldulensis. These advances in genome mapping should lead to a much greater understanding of the structural and functional basis of the genomes and eventually enable the more efficient production of superior varieties in breeding programs of tropical trees.

Introduction

The genome is defined as the sum of the genes and intergenic sequences of the haploid cell of an organism [1]. Genomics is the study of how genes and genetic information are organized within the genome, and how this organization determines their function [2]. Genomics has been described as either structural or functional. Structural genomics involves sequencing and mapping of genomes and a study of the structure of all gene sequences encoded in a fully sequenced genome [2]. Genome mapping, as part of structural genomics, focuses on producing three map types, namely, the genetic map, the physical map, and the complete nucleotide sequence map, as described in the Human Genome Project [3].

A genetic map is usually constructed with a large number of genetic markers by determining how frequently two markers, such as physical traits or detectable DNA sequences, are inherited together in certain pedigree(s). Markers that lie close together on a chromosome have a much higher chance of being inherited together than do markers that lie farther apart. Relative distance between gene markers on a chromosome is measured in centimorgans (cM). Two markers are one centimorgan apart if they are separated one percent of the time during transmission from parents to sibs. The physical distance to which a centimorgan corresponds varies a great deal, but the genome-wide average distance for a centimorgan is believed to be roughly 1 million base pairs in human [4].

A physical map can be constructed with microscope-based cytogenetic markers (Class I, low resolution), chromosome-derived pieces (clones) (Class II), or sequence-tagged sites (STSs) (Class III). Distance between markers is represented physically in Mb, kb, or bp. In recent times, physical maps commonly refer to Class III markers. In this class, an STS could be any mapped element [YAC (Yeast Artificial Chromosome), BAC (Bacterial Artificial Chromosome), PAC (Plastimid Artificial Chromosome), or cosmid based individual clone, contig, and/or sequenced region], which is basically a short DNA sequence that has been shown to be unique. A sequence map is the complete DNA base sequence of the genome of an organism and is invaluable for understanding the structure and function of genes in the genome.

In early years, genome mapping was restricted to only genetic map construction, which was hampered by the lack of powerful marker techniques, though it evoked great enthusiasms of the geneticists. Phenotypic characters were pursued as markers for genetic mapping at the initial stage. For instance, the early genetic linkage map of fruit fly (*Drosophila melangogaster*) was consisted of 56 phenotypic markers across the four linkage groups [5]. Later in the late 1960s, allozyme technique began to be used widely in biology and in genetic mapping as well, but its usefulness was limited mainly by the small number of allozyme loci available [6, 7]. The new era of genome mapping came with the advent of DNA-based molecular marker technology, especially the two milestone techniques, namely, RFLP (Restrictive Fragment Length Polymorphisms) proposed in 1974 [8] and PCR (Polymerase Chain Reaction) in 1986 [9], which could theoretically provide unlimited markers for genetic mapping and other purposes. Since then, a great number of molecular marker systems have been developed on basis of the above techniques and widely used in genome mapping studies in tree species. So far, high and ultra-high density genome maps have been constructed using a variety of molecular marker types for several plant species, for example, rice (*Oryza sativa*) [10, 11] and wheat (*Triticum aestivum*) [12], and complete genome sequence maps are available for three plant species, *Arabidopsis thaliana* [13], rice [14,15], and *Populus trichocarpa* (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html).

Trees are generally long-lived perennials, out-crossing, and with high levels of heterozygosity, and differ greatly in genetic background from such model plants as *Arabidopsis* and rice. As a result specific mapping strategies tend to have been used for genome map construction in tree species [e.g. 16, 17]. Nowadays, more than 40 tree species have been involved in genome mapping all over the world, many focusing on genetic map construction following a pseudo-testcross strategy. Although these maps vary to some degree in map density, marker type, and mapping pedigree type, their potential applications to breeding are widely recognized by the tree genetics community. Thus the completion of the whole genome sequencing (a sequence map) of *Populus* in 2004 will help to provide a solid base in tree genomics to advance biological knowledge and aid in application of molecular technologies to breeding programs (http://www.jgi.doe.gov/News/news_9_21_04.html).

For tropical trees the genomic technologies are similar to those used in other animal or plant species. Here we deal with genome mapping of tropical tree species in respect to geographical rather than technical considerations. For a number of major genera such as *Eucalyptus*, *Acacia* and *Pinus* the majority of species are of temperate origin. However some species do have tropical ranges or partially so (eg *E. camaldulensis, E. tereticornis,* and *E. grandis*) and these as well as other species planted in tropical environments have been considered as tropical tree species for this review. The aims of this chapter are

to present an overview and put forward perspectives on genome mapping in tropical tree species.

Molecular marker types

Molecular markers, or DNA markers, serve as a biological tool for genetic mapping in tropical trees. A DNA marker locus refers to the unique position within the genome of a specific DNA fragment and is not necessarily part of a gene. The genetic polymorphisms at these DNA loci are the consequence of mutations and chromosomal rearrangements that occur during the evolutionary process. DNA marker technology has been developed on basis of several basic techniques, e.g. restriction enzyme digestion, PCR, Southern blot, electrophoresis, and sequencing. Though a large number of marker types have been developed, they can be divided into three basic categories based on their methods of characterization : (I) electrophoresis- and Southern-blot-based markers, (II) PCR- and electrophoresis-based markers, and (III) DNA-sequencing-based markers.

Category I: Electrophoresis- and Southern-blotting-based markers

This category includes mainly RFLP [8] and DNA fingerprinting [18]. In practice the genomic DNA is digested with restriction enzymes, agarose gel electrophoresed, and then transferred to N^+ membranes by Southern blotting, then following by hybridization with DNA probes [see 19]. Probes used in RFLP are preferably low-copy DNA fragments, including genomic DNA, cDNA, and ESTs (Expressed Sequence Tags), whereas those for DNA fingerprinting are repetitive sequences, including tandem repetitive sequences (satellite, minisatellite, and microsatellite) and dispersed repetitive sequences (transposable element and anti-transposable element).

Mendelian RFLPs in the nuclear genome are codominant and can be multiallelic. Moreover RFLPs transfer very successfully to other species in the same genus and this transferability even extends across closely related genera. Nevertheless, they are technically challenging to do in the laboratory, because of requirement of large amounts of DNA, utilization of radioactive isotypes, and necessity of labor-intensive and time-consuming hybridization and autoradiography procedure. There is currently a trend to convert hybridization-based markers to PCR-based procedures, e.g. PCR-RFLP (or Cleaved Amplified Polymorphic sequence, CAPS) especially when the RFLPs occupy key location in a genetic linkage map.

Category II: PCR- and electrophoresis-based markers

The main types of markers in this category are RAPD (Random amplified polymorphic DNA) [20], SSR (Simple sequence repeats, or microsatellites)

[21], ISSR [22], and AFLP (amplified restriction fragment polymorphisms) [23]. PCR-based markers, in contrast to hybridization-based techniques, require less input in time, manpower, and DNA quantities and are therefore preferred for linkage mapping. RAPD, ISSR, and AFLP are dominant markers whereas SSR are co-dominant and multiallelic and hence genetically more informative for mapping. The downside of SSRs has been the considerable resources required to develop large numbers of them that would be required for comprehensive map construction. RAPD and AFLP loci can be assayed within species but not readily across species so their application in comparative mapping is limited.

Conservation of SSR marker loci appears to reduce as the evolutionary distance between species increases but can vary considerably depending on the genera. For instance in eucalypts about half of the SSR loci are conserved across subgenera and around 22% across related genera [24] whereas theses figures are considerably less for acacias [25]. Microsatellite motifs exist in plants, but at significantly lower densities than in mammalian genomes. For example, dinucleotide repeats are found in *Pinus taeda* at frequencies of one every 520 kb (AC stretches), trinucleotide repeats (namely AAT stretches) can be as frequent as dinucleotides, and tetranucleotides are at least as rare as one every 1,500 kb [26]. Recently, more and more EST-derived SSRs have been reported, e.g. in conifers [27, 28], *Coffea* spp. [29] and eucalypts [30].

Category III: DNA-sequencing-based markers

ITS (Internal transcribed spacers) [31] and SNP (Single nucleotide polymorphisms) [32] fall into this category. Sequencing-based markers are of substantial cost to develop, especially for characterization of SNPs, and thus they are not yet in common usage outside of the major commercial tree species. However they will become the marker of choice for mapping and association studies of candidate genes in trees [33] and a significant advantage will be that large scale automated genotyping is possible.

It should be noted that the classification of three marker categories is simply a rough definition for operational convenience purposes rather than a strict description, and some marker types may span two or three of the categories mentioned above. For example, a SCAR (Sequence characterized amplified region) marker is usually developed from sequencing of a specific fragment [34], usually an RAPD or AFLP marker, so categories II and III are involved. STS (Sequence-tagged sites) spans the three categories when development of a pair of PCR primers is based on sequencing of an RFLP probe [35], but categories II and III when based on a YAC (Yeast artificial chromosome) contig [36].

In addition, a number of gel-running techniques have been proposed for improving the efficiency in detecting SNP-based fragment polymorphisms,

e.g. SSCP (Single-strand conformation polymorphism) [37], DGGE (Denatured gradient gel electrophoresis) [38], and CAPS [39]. As PCR primer pairs are designed always from sequenced fragments, these techniques are generally classified as combination of marker categories II and III.

Another classification of molecular markers could be in terms of dominant or codominant. Dominant markers, such as RAPD, AFLP, and ISSR, require no prior sequence information, and can, therefore, quickly generate a high number of anonymous markers for genetic mapping. Codominant multiallelic markers, such as RFLP, SSR, STS, and CAPS, however, are sequence or at least probe dependent and more informative than dominant markers. Generally codominant markers are much more transferable across species and can be used as orthologous landmarks in conjunction with dominant-marker-based maps to determine synteny among homoeologous linkage groups [40]. SNPs are mostly biallelic and hence less informative than say SSRs. and it is also not known how often they will be polymorphic and assayable across species. However any SNP markers in the same sequence region will suffice for mapping purposes.

Some basic concepts in tree genetic mapping

Mapping pedigree

In trees pedigree types used for genetic mapping are generally outbred full-sib families or crosses. In most species long generation times have meant that only two generation pedigrees are readily available for mapping. Similarly full-sib crosses between species have been widely used so that simultaneous construction of maps for the two parental species could be made with dominant markers. In addition, haploid megagametophytes [e.g. 41], double haploids [e.g. 42], outbred three generation pedigrees [e.g. 43], inbred F_2 [e.g. 44], and selfed F_2 [e.g. 44] have been employed in genetic mapping for various tree species.

Marker segregation types in an outbred full-sib pedigree

Segregation type of a locus describes the alleles present in the parents of a cross or family, e.g. *ab*×*cd*, and thereafter the possible distribution of genotypes in the sibs. In some certain cases, segregation type applies to only the maternal parent for a half-sib family and to grandparents as well as parents for a three-generation pedigree.

Outbred species, such as trees, have high levels of heterozygosity in their genetic background, resulting in a much more complex genotypic configuration in full-sib pedigrees than that of crops. For instance, the percentage of heterozygous loci is from 16.2% to 39.5% in eucalypts [45], but only 3.6% in *Brassica napus* [46]. On one hand, the higher the heterozygosity in the parents, the more polymorphisms in the offspring, resulting in a greater

number of markers for genetic mapping. On the other hand, heterozygosity complicates the marker segregation types in an outbred full-sib family and raises unique considerations in mapping analysis.

Maliepaard *et al* (1997) summarized that in an outbred full-sib family there were seven essentially distinct segregation types for a locus (marker) providing recombination information [47]: (1) two alleles, one parent heterozygous (*ab*×*aa*), or (2) the other parent heterozygous (*aa*×*ab*), (3) two alleles, both parents heterozygous (*ab*×*ab*), (4) four alleles (*ab*×*cd*), (5) two alleles, of which one is a null-allele, both parents heterozygous (*a0*×*a0*), (6) three alleles, of which one is a null-allele (in one copy), two parents heterozygous, the null-allele in the one parent (*ab*×*a0*), or (7) in the other (*a0*×*ab*). The segregation expectations of these segregation types are shown in Table 1.

Table 1. Informative segregations at a locus in an outbred full-sib family.

Parent genotypic configuration	Segregation expected in progeny
(1) *ab*×*aa*	1 (*aa*):1 (*ab*)
(2) *aa*×*ab*	1 (*aa*):1 (*ab*)
(3) *ab*×*ab*	1 (*aa*):2 (*ab*):1 (*bb*)
(4) *ab*×*cd*	1 (*ac*):1 (*ad*):1 (*bc*):1 (*bd*)
(5) *a0*×*a0*	3 (*a*-):1 (*00*)
(6) *ab*×*a0*	2 (*a*-):1 (*ab*):1 (*b0*)
(7) *a0*×*ab*	2 (*a*-):1 (*ab*):1 (*b0*)

Note: The numeral "*0*" represents a null allele, and the symbol "-" any possible allele, e.g. *0*, *a*, or *b*, if applicable.

Table 2. Configuration numbers of all pairwise marker combinations of segregation types.

Locus one	Locus two						
	ab×*aa*	*aa*×*ab*	*ab*×*ab*	*ab*×*cd*	*a0*×*a0*	*ab*×*a0*	*a0*×*ab*
ab×*aa*	1	*	2	3	4	5	6
aa×*ab*		(1)	(2)	(3)	(4)	(6)	(5)
ab×*ab*			7	8	9	10	(10)
ab×*cd*				11	12	13	(13)
A0×*a0*					14	15	(15)
ab×*a0*						16	17
A0×*ab*							(16)

Note: The numeral "*0*" stands for a null allele. The asterisk "*" represents unavailability in recombination information. The number in parentheses indicates the configuration equivalent to its reciprocal cross. When no number is given the configuration is equivalent to that of the loci exchanged (Source: adopted from Maliepaard *et al* (1997) [47]).

For any two loci or markers, consequently, there are 17 possible types of recombination in an outbred full-sib family (Table 2) [47]. A recombination event refers to the fact that an allele at a certain locus in a sib is from one homologue of a parent and the allele at the next locus from the other. The process of linkage analysis is essentially to detect recombination frequency between loci in both parental meioses.

Mapping strategies

A variety of mapping strategies have been commonly adopted to construct genetic maps of trees, e.g. the pseudo-testcross strategy, the outbred three generation model, and the double-haploid model. All these strategies follow those practiced in mammalian or crop species and facilitate linkage analysis by using the software readily available, such as MAPMAKER [48] and MAPMANAGER [49]. However, as we mentioned above, there may be seven types of marker segregation for an outbred cross or family, so single model could not take full advantage of all the marker configurations, especially when dominant and co-dominant markers are combined together.

For out-breeders, such as trees, linkage estimation must distinguish between coupling and repulsion phase, for both dominant and codominant markers, and must accommodate as many as four alleles (in a diploid) segregating at a locus [50]. Few software packages available currently can handle phase-unknown marker data. Though JOINMAP can map all segregation types of marker [51], it uses only pairwise recombination fractions between markers to estimate marker order and thus cannot be expected to be as accurate as multi-locus likelihood-based packages. Mapping programs optimal for outcrossed pedigrees should be able to handle phase-ambiguous data and use multiple locus information when identifying the order of loci within a linkage group. In this respect, OUTMAP [52] could be a sound choice for linkage analysis of marker data for outbreeders.

OUTMAP is specifically designed for analyzing segregation data from codominant loci in outcrossed pedigrees and deals effectively with phase ambiguous data. It can successfully handle all segregation types, determine phase, provide a choice of three optimization methods and calculate the likelihood of alternative marker orders. The risk of introducing errors when recoding data to suit the input format of different programs is avoided. In addition, there is no need to divide segregation data into separate data sets for male and female meiosis. Butcher and colleagues [53] demonstrated that the marker orders produced using OUTMAP were consistently of higher likelihood than those produced by JOINMAP and distances between markers often varied from those calculated by JOINMAP, resulting in an increase in the estimated genome length.

Genetic maps constructed in tropical tree species

To date genome mapping in tropical trees has primarily consisted of the construction of initial low density genetic maps in some of the commercial species. Genetic maps are available for species from five genera including *Eucalyptus, Pinus, Acacia, Coffea*, and *Hevea*. Table 3 gives details of the genetic maps reported for tropical tree species and their main features are summarized below.

Genetic maps of tropical eucalypts

Several papers have demonstrated the utilization of different types of DNA markers for constructing genetic linkage maps in tropical eucalypt species, including *Eucalyptus urophylla*, *E. grandis*, *E. tereticornis*, and *E. camaldulensis*.

Grattapaglia and Sederoff (1994) proposed a "two-way pseudo-testcross" mapping strategy and constructed two moderate density genetic maps for *E. grandis* and *E. urophylla* with RAPD markers [16]. They used 151 RAPD primers to amplify a total of 558 markers on the mapping population, including 272 markers from *E. grandis* and 286 from *E. urophylla*. Finally they mapped 240 markers into 14 linkage groups (1552 cM) in maternal *E. grandis* and 251 markers in 11 linkage groups (1101 cM) in paternal *E. urophylla* (n = 11 in *Eucalyptus*). From the 558 RAPD markers, 516 (92.5%) were fully repeatable in both replicates, indicating a reliable repetibility of RAPD markers. They found clustering of markers occurring throughout both linkage maps, which could be an artifact resulting from the limited resolution of the mapping population. They revealed that 53% of 48 mapped RAPD markers were amplified from low copy genome regions. It was pointed out that pseudo-testcross/RAPD mapping strategy should be efficient at both inter- and intraspecific levels.

Verhaegen and Plomion (1996) constructed two single-tree linkage maps for *E. urophylla* and *E. grandis*, respectively, using RAPD markers [54]. Totally 480 RAPD markers were scored in an F_1 interspecific mapping population, including 244 1:1 from female *E. urophylla*, 211 1:1 from male *E. grandis*, and 25 3:1 shared by both parents. They used 1:1 segregation (testcross) markers to establish separate maternal and paternal maps, while 3:1 segregation (intercross) markers were integrated to identify homology between linkage groups between the two maps. The maternal *E. urophylla* map they constructed was consisted of 269 markers covering 1331 cM in 11 linkage groups, and the paternal *E. grandis* map 236 markers covering 1415 cM in 11 linkage groups. The genome coverage was 95% and 99% for the maternal and paternal maps, respectively. They compared their maps with those reported by Grattapaglia and Sederoff [16] and identified seven parallel linkages, which demonstrated the usefulness of RAPD fragments segregating in a Mendelian

Table 3. Genetic linkage maps available for tropical tree species.

Species	Marker type (mapped marker No.)	Map length (cM)	Linkage group No.	Coverage (%)	Reference
Eucalyptus grandis	RAPD(240)	1552	14	95	[16]
	RAPD(236)	1415	11	95-99	[54]
	SSR(63)	*NA*	*NA*	*NA*	[57]
Eucalyptus urophylla	RAPD(251)	1101	11	95	[16]
	RAPD(269)	1331	11	95-99	[54]
	RAPD(160)	1504.6	23	94.9	[59]
	SSR(53)	*NA*	*NA*	*NA*	[57]
Eucalyptus tereticornis	AFLP(268)	919	14	80-100	[55]
	RAPD(126)	1035.7	23	68.7	[59]
Eucalyptus camaldulensis	RAPD(73), RFLP(81), SSR(14)	1236	11	*NA*	[56]
Pinus elliottii	RAPD(73)	782	13	64-75	[60]
	RAPD(91)	952.9	13	61.7	[61]
	AFLP(71), SSR(7)	1170	23	82	[62]
Pinus caribaea	AFLP(100), SSR(9)	1658	27	88	[62]
Acacia mangium	RFLP (219), SSR(33)	966	13	85-90	[17]
Coffea canephora	AFLP(96), RAPD(11), SSR(18), RFLP(35)	1041	11	74.4	[64]
	RFLP(47), RAPD(100)	1402	15	*NA*	[42]
Coffea ozanguebariae & C. liberica	AFLP(167), RFLP(13)	1144	14	*NA*	[54]
Coffea arabica	AFLP(177)	1802.8	31	*NA*	[65]
Hevea brasiliensis × *H.benthamiana*	RFLP(301), AFLP(388), SSR(18), isozyme(10)	2144	18	*NA*	[66]

Note: Numbers flanking the slash '/' refer to the maternal/paternal parents, respectively, of the mapping population. *NA*: not available.

mode in two unrelated trees of the same species for establishing with high confidence a consensus species map.

Marques and colleagues (1998) used AFLP markers to generate genetic maps of *E. tereticornis* and *E. globulus* (a temperate species) [55]. They scored 606 polymorphic fragments on the mapping population of *E. tereticornis* × *E. globulus*, including 487 segregating in a 1:1 ratio (testcross loci). The proportion of segregation-ratio distorted fragments was fairly high in this study, accounting for 15%. They ordered 268 markers in 14 linkage groups in the maternal *E. tereticornis* map (967 cM), and 200 markers in 16 linkage groups in the paternal *E. globulus* map (919 cM). The genome coverage ranged 80%-100%. They detected straightforward homologies between 9 linkage groups in *E. tereticornis* and 10 linkage groups in *E. globulus* using 15 markers segregating 3:1. They also highlighted that the inclusion of fragments showing segregation-ratio distortion would result in some increase in map length, e.g. 20% increase for *E. tereticornis* and 14% for *E. globulus*.

Agrama and colleagues (2002) used RAPD, RFLP, and SSR markers to construct a linkage map for *E. camaldulensis* [56]. Linkage analysis resulted in 11 linkage groups covering 1236 cM of the genome and comprising 168 markers, including 73 RAPDs, 81 RFLPs, and 14 SSRs. They identified 90 orthologous markers that were common in both mapping parents, which served as consensus markers to determine homologies of linkage groups in the two maps and integrate the two maps into one.

Brondani and colleagues (2002) [57] integrated 70 genomic SSR markers on the RAPD framework maps of *E. grandis* and *E. urophylla* previously reported by Grattapaglia and Sederoff (1996) [16]. They placed 63 SSR markers on 11 linkage groups for *E. grandis*, and 53 on 10 linkage groups for *E. urophylla*. In *E. urophylla*, the addition of microsatellite markers to the RAPD framework map did not change the number of linkage groups. In *E. grandis*, however, a merger of groups 7 and 14 at LOD score ≥ 3.5 was observed, resulting in the reduction of the number of linkage groups from 12 to the expected final number of 11. At a likelihood support of 1000:1, the locus order was colinear in the two species for 39 (97.5%) out of the 40 loci that could be compared between the two parental maps. They also observed that numerical differences, though not significant, in the estimates of recombination frequency for some locus pairs between the two maps. They demonstrated that the collinear arrangement of microsatellite markers along the linkage maps of the two *Eucalyptus* species could set the stage for the final construction of a genus-wide reference map that would be useful for the great majority of commercially important eucalypt species.

Myburg and colleagues (2003) conducted comparative genetic mapping between *E. grandis* and *E. globulus* using an F_1 hybrid of the two species [58]. They estimated that approximately 20% of loci in the genome of the F_1 hybrid

were hemizygous due to a difference in genome size between *E. grandis* (640 Mbp) and *E. globulus* (530 Mbp). They investigated the extent of colinearity between the two genomes and the distribution of hemizygous loci in the F_1 hybrid using high-throughput, semi-automated AFLP marker analysis. They used more than 800 AFLP markers to genotype two pseudobackcross families (backcrosses of an F_1 individual to non-parental individuals of the parental species) and found popular colinearity for all shared AFLP marker loci in the three single-tree parental maps and little evidence for gross chromosomal rearrangements. They pointed out that hemizygous AFLP loci could disperse throughout the *E. grandis* chromosomes of the F_1 hybrid.

Gan and colleagues (2003) constructed moderate-density molecular maps for the genomes of *E. urophylla* and *E. tereticornis* using RAPD markers and an interspecific cross between the two species [59]. One hundred and eighty-three primers were employed to generate 245 and 264 parent-specific markers in *E. urophylla* and *E. tereticornis*, respectively, as well as 49 parent-shared markers. The normally segregating markers, including 208 (84.9%) specific to maternal *E. urophylla*, 175 (66.3%) to paternal *E. tereticornis*, and 48 shared by both parents, were used for framework map construction for each parental species. For maternal *E. urophylla*, the linkage map consisted of 23 linkage groups, 160 framework markers, and 60 accessory markers, defining a total map distance of 1504 cM and an average interval of 11.0±8.07 cM. For paternal *E. tereticornis*, the linkage map contained 23 linkage groups, 126 framework markers, and 92 accessory markers, defining a total map distance of 1035 cM and an average interval of 10.1±7.23 cM. Genome length was estimated at 1585 and 1507 cM for *E. urophylla* and *E. tereticornis*, respectively, indicating map coverage of 94.9 and 68.7% of the corresponding genome.

Genetic maps of tropical pines

So far, two commercially important tropical pine species have been involved in genetic map construction, that is, *Pinus elliottii* and *Pinus caribaea*, and for both species interspecific crosses and dominant PCR markers were used.

Nelson and colleagues (1993) constructed the first genetic map of *Pinus elliottii* [60]. They grouped 73 RAPD markers into 13 linkage groups and nine pairs spanning a genetic map distance of approximately 782 cM. The map represented a genetic distance of 2160 cM, approximately two-thirds of the slash pine genome. The map coverage was estimated to be 64-75% of the genome.

Kubisiak and colleagues (1995) used RAPD markers to construct linkage maps of the parents of a longleaf pine (*Pinus palustris*) × slash pine (*Pinus elliottii*) F_1 population [61]. They mapped 91 markers into 13 groups and six pairs for slash pine, with a total map distance 952.9cM. The total map coverage was estimated to be 1462.9 cM, approximately 61.7% of the slash

pine genome. They used the 3:1 loci to identify homologous linkage groups between longleaf pine and slash pine and found that four of the longleaf-pine linkage groups appeared to be potentially homologous to five different slash-pine linkage groups.

Shepherd and colleagues (2003) constructed genetic maps for individual *Pinus elliottii* var. *elliottii* and *Pinus caribaea* var. *hondurensis* trees using AFLP and SSR markers [62]. They mapped 78 markers into 23 linkage groups for *P. elliottii*, spanning 1170 cM in total map distance and covering 82% of the genome. For *P. caribaea*, they mapped 109 markers into 27 linkage groups, with a total map distance 1658 cM and 88% in genome coverage. Additionally, they established homologous linkage groups between 11 of the 24 *P. elliottii* groups and 10 of the 25 *P. caribaea* groups using 19 "bridge" markers.

Genetic maps of tropical acacias

Butcher and Moran (2000) constructed an integrated genetic linkage map of *Acacia mangium*, using two outbred pedigrees [17]. Basing on construction of individual maps for each pedigree, they assigned 219 RFLP and 33 microsatellite markers in 13 linkage groups, with total map length 966 cM and genome coverage 62%. They found that differences in recombination rates between linked loci in male and female meioses as well as between parents were confined to specific regions and were not uniform across the genomes. They proposed that the integrated map would provide a sound basis for QTL detection, leading to marker-assisted selection in *A. mangium*, and syntenic mapping between *Acacia* species would be possible using microsatellites and RFLPs.

Genetic maps of coffees

Paillard and colleagues (1996) constructed the first genetic linkage map in coffee (*Cofffea canephora*) using doubled haploids [42]. They placed a total of 47 RFLP and 100 RAPD loci on 15 linkage groups, totaling 1402 cM in map distance. They detected rather low DNA polymorphism rate in the mapping population, e.g. 18% for RFLP probes and 29% for RAPD primers.

Ky and colleagues presented in 2000 an interspecific partial genetic linkage map of *Coffea* sp. based on 62 backcross hybrids [63]. The mapping population was a backcross of (*C. pseudozanguebariae* × *C. liberica* var. *dewevrei*) × *C. liberica* var. *dewevrei*. They assembled 167 AFLP and 13 RFLP loci into 14 linkage groups, covering a total map distance 1144 cM. They observed a high ratio of segregation distortion of markers, for example, 30%.

Lashermes and colleagues (2001) used a doubled diploid (DH) population and a testcross (TC) pedigree of *Coffea canephora* from the same clone to carry out genome mapping work [64]. Based on the DH population, they identified 11 linkage groups with 160 markers, putatively corresponding to the

11 gametic chromosomes of *C. canephora*. Segregation distortion of markers was especially high in the DH population. They found indistinguishable differences in the recombination frequencies in both populations, indicating the lack of significant sex differences in recombination in *C. canephora*. Among the markers mapped, as they stated, the single-copy RFLP probes and microsatellites could serve as standard landmarks in coffee-genome analyses.

Pearl and colleagues (2004) used AFLPs to construct a genetic linkage map of arabica coffee (*Coffea arabica*) using a pseudo-F_2 population derived from a cross between the cultivars Mokka hybrid and Catimor [65]. Their analysis resulted in 16 major linkage groups containing 4–21 markers and 15 small linkage groups consisting of 2–3 linked markers each, with a total map length of 1,802 cM and an average distance of 10.2 cM between adjacent markers.

In addition, Poncet and colleagues (2004) defined an initial set of 54 highly conserved, single copy genes (COS) as markers for comparative mapping between the tomato and coffee genomes (http://www.asic-cafe.org/pdf/abstract/B204_2004.pdf). Crouzillat and colleagues (2004) constructed genetic maps of *Coffea canephora* var. *robusta* using 453 RFLP and SSR markers as well as a cross between elite clones BP409 and Q121 (http://www.asic-cafe.org/pdf/abstract/B202_2004.pdf). They identified 11 linkage groups, covering a total map distance of 1258 cM.

Genetic maps of rubber tree

Lespinasse and colleagues (2000) presented the first and only one published genetic map for *Hevea* spp. (2n=36) [66]. The mapping pedigree comprised of 106 F_1 individuals of *H. brasiliensis* (clone PB260) × (*H. brasiliensis* × *H. benthamiana*) (clone RO38). They assembled 717 markers into 18 linkage groups of a synthetic map. The total map distance was 2144 cM, and the average marker density was 1 per 3 cM. In contrast to Lashermes *et al.* (2001) [64], they revealed significantly less meiotic recombination in the interspecific hybrid male parent than in the female parent.

Issues in genetic mapping of tropical trees

The first problem is distorted segregation of markers. It is well known for plants and is detected by underrepresented allelic classes presumably due to a dysfunction of the relevant gametes [67]. In outbred trees, marker distortion has been reported nearly in all mapping efforts and, in some cases, accounts for a considerably high ratio, for instance, 44% of RFLP and SSR markers in coffee DH population [64] and 40.6% of EST markers in *Pinus taeda* F_2 progeny [68]. A variety of factors may contribute to the phenomenon, such as the presence of semi-lethal genes and self-incompatibility loci as well as the selection process during seed and seedling development.

The second problem is lack of large numbers of suitable codominant markers, especially those with high transferability across pedigrees and species. It would be extremely desirable to be able to transfer markers and map information from marker-rich species to marker-poor species. However, the markers used for mapping in tropical trees have been PCR-based dominant markers with limited transferability within and across taxa. In contrast, microsatellites, which are highly variable and codominant, are transferable between fairly closely related species, but are of limited usefulness for comparative studies across subgenera or genera [57, 69]. RFLPs are more conserved and transferable but technically too difficult and time consuming to do in many labs. Markers conserved across different species within a taxon, such as coding-sequence-derived cDNA, EST or candidate gene markers, could serve as "anchor loci" to ultimately enable information to flow from map-rich to map-poor species and facilitate the comparative mapping even across distant taxa.

The third problem is relatively low density. Compared to high or ultra-high density genetic maps constructed in crops, e.g. rice *(Oryza sativa)* [10,11] and wheat (*Triticum aestivum*) [12], genetic maps of tropical trees are exclusively at a low density, containing up to a few hundred markers. High-density linkage maps with transferable markers potentially enable a reference map for a genus to be established and would find wide application in determining the organization and function of the genome. For some genera such as *Acacia* and *Eucalyptus* such maps should be an objective towards which tropical tree genetic mapping should be directed. For genera such as *Pinus* reference maps would be better based on major temperate species like loblolly or radiata pine. Moreover, long intervals between adjacent markers, say, 30cM and up, appear regularly on nearly all tropical tree genetic maps. Those sparse regions should be particularly targeted in the consequent efforts in map saturation, especially those known to be linked to quantitative trait loci (QTLs) in other species.

The fourth problem is little progress in comparative mapping with closely related species or such model plants as *Arabidopsis thaliana*, rice, and poplar, whose complete genome sequences are available. One reason for this could be the limitation of cross-species, closely or distantly, transferable markers as mentioned above. Also, the low map density constrains such comparative mapping to a limited scale. In this respect, the accumulation of EST databases of both tropical trees and other plants as well as the construction of ultra-high density genetic maps of tropical trees with such gene markers will allow comparison of genome organization with model plant species.

Genetic map based QTL detection in tropical trees

Much of the variation within populations or breeds is quantitative in nature. Examples of traits in trees include growth, yield, wood properties,

flowering time and stress responses. The genetic basis of such phenotypic variation lies in the combined effects of variation at several or many loci on chromosomes, and the individual loci that contribute to such variation have been termed quantitative trait loci or QTLs [66].

Traditionally, the effects of QTLs underlying a trait are evaluated as one phenotype in breeding programs and as such the number and size of separate QTLs are difficult to identify. Genetic maps based on fullsib pedigrees provide a suitable platform for QTL detection and dissection. QTLs are characterized by finding significant statistical associations between variation in phenotypes and variation in genotypes. The analyses can reveal the number, location and size of effect of QTL and even mode of action of alleles at each QTL. More efficient selection of a trait can then be achieved with markers associated with favourable QTL alleles.

Several considerations should be taken into account in QTL mapping. The first is map density. The closer a QTL is to a marker, the smaller the effect that QTL can have and still be detected statistically. Hence to some extent power of QTL mapping can be improved with high density molecular maps especially if it enables use of markers spaced at regular intervals along the genome. The second related consideration is significance thresholds for detection of QTL. With many non-independent markers statistical tests are not independent and higher probability thresholds for declaring a QTL effect are required. A result is a significant reduction in the chances of spurious QTLs being reported, but also probably reduces the chances of detecting QTLs with smaller effects. The third consideration is population sizes required in QTL mapping to detect and verify QTL. Large population sizes are required for correct identification of QTL, especially the size of effect of alleles at QTL [71]. Using a typical sample size ($n<500$), two or more genes closely linked will be usually detected as a single QTL and indicates that resolution of QTL in terms of genome location is limited. Such large experiments are difficult in resource terms in many tree species. The fourth consideration is heritability of the trait under investigation. The larger the environmental effect on the character (i.e. low heritability), the less likely a QTL will be detected. Estimates of heritability can be improved by controlling environmental error and adopting structure-fine population types, such as RILs, DHs, advanced backcrossing populations, or near-isogenic lines (NILs). For most tropical trees the availability of these pedigree types is unlikely and characterization of QTL for low heritability traits such as growth will remain a challenge. Finally the type of markers will be important to outcomes of QTL experiments with codominant markers enabling tracking of QTL alleles.

To date, a number of QTLs have been mapped on tropical tree genetic maps. Table 4 presents a summary of the QTLs detected in tropical tree species.

Table 4. Quantitative trait loci (QTLs) detected in tropical tree species.

Species	Trait [a]	No. [b] QTLs	% Var [c] explained	Reference
Eucalyptus grandis	FWS	6	41.6	[72]
	#Cutt	4	22.9	[72]
	%Root	1	8.5	[72]
	Breast-high circumference	2	>10.0	[73]
	Wood specific gravity	5	22.0	[73]
	Percentage dry weight of bark	2	>10.0	[73]
	18-month-old wood density	4	29.7	[74]
	26-month-old wood density	3	21.0	[74]
	38-month-old wood density	3	20.4	[74]
	18-month-old height/diameter ratio	1	6.2	[74]
	26-month-old height/diameter ratio	3	23.4	[74]
	38-month-old height/diameter ratio	2	15.6	[74]
	26-month-old stem growth	3	22.0	[74]
	38-month-old stem growth	3	17.7	[74]
	Monoterpene composition	6	>68.0	[76]
	Rust resistance	1	*NA*	[77]
E. tereticornis	MORT95	2	8.3	[75]
	MORT96	3	24.4	[75]
	ROOT95	5	24.2	[75]
	RCT95	3	13.9	[75]
	ROOT96	2	12.4	[75]
	RCT96	3	14.7	[75]
	PETR95	3	16.2	[75]
	PETR96	3	24.3	[75]
E. urophylla	FWS	4	25.2	[72]
	#Cutt	2	14.7	[72]
	%Root	3	26.3	[72]
	18-month-old wood density	1	10.5	[74]
	26-month-old wood density	2	15.3	[74]
	38-month-old wood density	1	6.0	[74]
	18-month-old height/diameter ratio	3	25.9	[74]
	26-month-old height/diameter ratio	1	6.9	[74]
	38-month-old height/diameter ratio	2	15.6	[74]
	18-month-old stem growth	2	26.2	[74]
	26-month-old stem growth	1	8.1	[74]
	38-month-old stem growth	3	22.1	[74]
Pinus caribaea	Average bark thickness	2	23.0	[78]
var. *hondurensis*	Average whorl spacing	2	34.0	[78]
	Average branch number per whorl	1	18.0	[78]
	Trunk height	3	51.0	[78]
	Whole core basic density	1	14.0	[78]
	Early wood ring width in the 3rd year	1	8.0	[79]
P. elliottii	Regularity of whorl spacing	1	17.0	[78]

Table 4. Continued

var. *elliottii*	Average branch diameter	2	33.0	[78]
	Average branch number per whorl	2	27.0	[78]
	Breast-high over bark diameter	1	16.0	[78]
	Breast-high under bark diameter	1	17.0	[78]
	Annual ring width in the 2nd year	1	8.0	[79]
	Dry wood mass index in the 3rd year	1	7.0	[79]
	Dry wood mass index in the 2nd year	1	6.0	[79]
	Average ring width (2-5 years)	1	10.0	[79]
	Mean dry wood index (2-5 years)	1	10.0	[79]
	Early wood ring width in the 5th year	1	8.0	[79]
	Early wood ring width in the 4th year	2	13.0	[79]
	Annual ring width in the 3rd year	1	7.0	[79]
	Early wood ring density in the 3rd year	1	7.0	[79]
	Mean early wood density (2-5 years)	1	8.0	[79]
Coffea arabica	Leaf rust	1	*NA*	[See text]
Hevea brasiliensis	Reaction type after blight inoculated	5	*NA*	[80]
	Lesion diameter after blight inoculated	4	*NA*	[80]

[a] FWS, fresh weight of micropropagated shoot clumps; #Cutt, number of operational stump sprout cuttings; %Root, percent rooting of cuttings; MORT95, ratio of dead/total cuttings in 1995 (the first year); MORT96, ratio of dead/total cuttings in 1996 (the second year); ROOT95, rooted/surviving cuttings in 1995 (the first year); ROOT96, ratio of rooted/surviving cuttings in 1996 (the second year); RCT95, ratio of rooted/total cuttings in 1995 (the first year); RCT96, ratio of rooted/total cuttings in 1996 (the second year); PETR95, ratio of surviving unrooted/total cuttings in 1995 (the second year); and PETR96, ratio of surviving unrooted/total cuttings in 1996 (the second year).

[b] Number of QTLs detected.

[c] Percentage of the phenotypic variation explained jointly; *NA*, not available.

QTLs in tropical eucalypts

Following the construction of linkage maps, several research groups have reported the identification of genomic regions that have a significant effect on the expression of economically important traits in tropical eucalypts. These traits include vegetative propagation ability (adventitious rooting, stump sprouting, and in vitro shoot multiplication) [72-74], volume growth, wood specific gravity, bark thickness, and stem form [75, 76]. QTLs for insect resistance and essential oil traits were mapped [77] and recently a major QTL for *Puccinia psidii* rust resistance with simple Mendelian inheritance was found and mapped in *E. grandis* [78].

Grattapaglia and colleagues (1995) detected ten QTLs for micropropagation response (measured as fresh weight of shoots, FWS), six for stump sprouting ability (measured as # stump sprout cutting, #Cutt), and four for rooting ability (measured as % rooting of cuttings, %Root) on genomes of

Eucalyptus grandis and *E. urophylla* (Table 4) [72]. They identified that the standardized gene substitution effects for the QTLs detected were typically between 0.46 and 2.1 phenotypic standard deviations (σ_p), while differences between the family mean and the favorable QTL genotype were between 0.25 and 1.07 σ_p. As the total genetic variation explained by the QTLs was large, that is, 89.0% for RWS, 67.1% for #Cutt, and 62.7% for %Root, the variation in these traits was mainly controlled by a relatively small number of major-effect QTLs. They also found that different traits seemed reasonably to share some QTLs in this experiment. In their mapping cross, *E. grandis* was responsible for most of the inherited variation in the ability to form shoots, while *E. urophylla* for most of the ability in rooting.

Grattapaglia and colleagues (1996) performed QTLs mapping of growth and wood quality in *E. grandis* using a maternal half-sib family and RAPD markers [73]. They found three QTLs related with volume growth (circumference at breast height, CBH) and five QTLs with wood specific gravity (WSG) (Table 4). They observed overlapping QTLs for different traits. For example, they detected QTLs for CBH, WSG, and percentage dry weight of bark (% bark) in the same interval between markers X1_1450 and R4_1300 on linkage group 5. They also observed significant digenic epistasis for volume growth. Again, their results demonstrate the existence of major genes involved in the expression of economically important traits related to forest productivity in *E. grandis* and have important implications for marker-assisted tree breeding.

Verhaegen and colleagues (1997) mapped a number of QTLs for different traits and different ages (18, 26, and 38 months) on genetic maps of *E. grandis* and *E. urophylla* [74]. By interval mapping, they detected 14 QTLs for age-specific wood density (PIL), 12 QTLs for age-specific stem-form (height:diameter ratio, HDR), and 12 QTLs for age-specific stem growth (VIG) in both species. They observed 68% of the QTLs being expressed at two ages, 32% being age-specific, and no QTL for all three ages. No significant marker × year interaction was found for the traits studied.

Marques and colleagues (1999) detected QTLs affecting vegetative propagation traits in *E. tereticornis* and *E. globulus* using AFLP genetic linkage maps [75]. For *E. tereticornis*, they detected a total of 18 QTLs putatively related with MORT (dead/total cuttings), ROOT (rooted/surviving cuttings), RCT (rooted/total cuttings) and PETR (surviving unrooted/total cuttings) in linkage groups 2, 4, 6, 7, 8, 10 and 11, but none with SPR (number of cuttings harvested) and STAB (stability of the adventitious rooting response). Single putative QTLs accounted for 2.79 – 17.03% of the phenotypic variance of a trait.

Shepherd and colleagues (1999) investigated the QTLs related with foliar oil (monoterpene) composition in *E. grandis* [76]. They clarified six putative QTLs in a single genomic region by interval mapping that could explain a

significant proportion (68–80%) of the phenotypic variation. They inferred that the single genomic region might harbor a gene or genes controlling the production of limonene, a predominant oil constituent, as heterozygotes at the QTL locus had in average a higher amount of limonene and lower amounts of the other four major monoterpenes.

Junghans and colleagues (2003) identified firstly a QTL controlling rust resistance in *Eucalyptus grandis* using RAPD markers [77]. They obtained 13 markers linked to the *Ppr1* gene (*Puccinia psidii* resistance gene 1), including five markers in repulsion phase, and constructed a 11.2-cM-long linkage group containing six markers and *Ppr1* gene. Among the linked markers, RAPD marker AT9/917 co-segregated with *Ppr1* without a single recombinant in 994 meioses. They suggested that the tightly linked marker should prove useful for marker-assisted introgression and will provide an initial lead for a positional cloning effort of this resistance allele.

QTL in tropical pines

Shepherd and colleagues (2002) identified putative QTLs with moderate additive effect for branch diameter, average number of branches per whorl per tree, average whorl spacing, and regularity of whorl spacing in a single *Pinus elliottii* var. *elliottii* × *P. caribaea* var. *hondurensis* cross [78]. They found no evidence of additive × additive epistasis or pleiotropy in their experiment. No marker-trait associations were detected for the average branch angle per whorl per tree. The genetic effects that they detected to be relatively larger for a number of branching traits were probably attributed to bias in the estimation of QTL magnitude and limited power to detect QTL due to the small sample size (89 individuals only). They also found that branch architecture traits exhibited considerable variation within the family with ranges of 4–6 standard deviations (SD) and tended to be less variable than height and diameter, and branching characters were largely independent of one another as well as growth, form and wood density properties and were not influenced by macro-environmental factors except for branch angle trait.

Shepherd and colleagues (2003) detected QTLs for physical wood properties and early growth traits in an interspecific hybrid between *Pinus elliottii* var. *elliottii* and *Pinus caribaea* var. *hondurensis* [79]. They detected a total of 12 putative QTLs by interval mapping across four genomic regions in *P. elliottii* var. *elliottii* and a single region in *P. caribaea* var. *hondurensis*, each of which could explain 6% - 10% of the total phenotypic variation. All putative QTLs originated from the *P. caribaea* parent except for early wood ring width in the 3rd year (ERW97). They observed clustering of QTLs on the pine genome, e.g. five putative QTLs for dry wood mass index (DWI) or ring width (RW) located to G1 of *P. elliottii*. They found that putative QTLs that influenced density and ring width did not colocate, suggesting independent

inheritance of these characters and being consistent with the lack of genetic correlation between wood density and diameter growth observed in quantitative studies in hybrid pines.

QTLs detected in coffee

Prakash and colleagues (2004) identified AFLP markers tightly linked to a leaf rust (pathogen *Hemileia vastatrix*) resistance gene *SH3* in *Coffea arabica* (http://www.asic-cafe.org/pdf/abstract/B207_2004.pdf). Totally, 21 markers were found associated with the *SH3* resistance gene and one marker co-segregated perfectly with *SH3*. Linkage analysis of the markers resulted in four linkage groups. All the markers linked to *SH3* were closely associated (6.3cM) and grouped together. These findings provide a starting point for further refinement of marker-trait associations.

QTLs detected in rubber tree

Lespinasse and colleagues (2000) mapped QTLs for resistance to South American leaf blight (SALB) (pathogen *Microcyclus ulei*) in rubber tree using a cross between a susceptible cultivated clone, PB260, and a resistant hybrid clone, RO38 [80]. They identified eight QTLs for resistance on the resistant parent RO38 map, and only one QTL on the susceptible parent PB260 map. Those QTLs were distributed in six regions on five among 18 chromosomes of rubber tree genome. They detected one common QTL on chromosome g13 controlling both the reaction type (RT) and the lesion diameter (LD) over the five strains tested, and two common QTLs for RT and LD, respectively, over four strains tested. Their results revealed that both partial and complete SABL resistances in rubber tree were multigenic traits, a genetic determinism considerably different from previous hypothesis.

Some considerations with QTL detection in tropical trees

One of the major objectives of QTL detection in plants is to facilitate marker-assisted selection (MAS), or so-called molecular breeding. For trees featuring in long generation turnover, this is extremely attractive regarding the great potential in reducing the breeding cycle and improving the selection efficiency. However, as Grattapaglia (2001) pointed out [81], the challenge for the application of MAS in forest trees is generally much more complex than crop plants as it presupposed the manipulation of several polygenic traits with different heritabilities in heterogeneous breeding populations, the incorporation of MAS in breeding schemes for large populations with altering frequencies of favorable alleles, and the inclusion of age × trait correlations and variable environments. Nevertheless, linkage equilibrium is the norm in forest trees and linkage phase between QTL and marker will not

extend from one pedigree to the whole breeding population. This suggests that application of MAS based on QTL alleles will not find wide application in breeding programs of tropical trees.

Like other trees, MAS-directed QTL detection in tropical trees is encountering several problems. Small population size, say, about 100 or 200 individuals, in most cases is a huge constraint affecting the characterization of QTL detection. Large populations are even more necessary for trees as mapping populations such as outcrossed two and three generation pedigrees have relatively low recombination events in comparison to multi-generation pedigrees. Moreover, comparative QTL mapping across two or more experimental populations has not been done in tropical trees, which is essential for tagging the superior QTL alleles and verification of QTL. Use of clonal replicates within and across field sites would increase accuracy of phenotypic data particularly for low heritability traits, reduce environmental variation and increase chances of correct QTL characterization. Additionally, map density can also be a factor influencing the resolution of QTL mapping and saturated maps are in urgent need for QTL mapping in tropical trees.

The classical approach to QTL detection using large populations is laborious and makes it unrealistic to screen sufficiently large populations against a lot of markers and precisely locate the QTLs [82]. Alternative strategies have been proposed, such as BSA (bulked segregation analysis) based approach [83] and reverse QTL mapping (RQM) [84]. In BSA based approach, bulks could be built with selections (about 100 or 200 individuals per bulk) of extreme phenotypes from a large mapping population, e.g. more than 2000 individuals, and then screened against a large number of makers. The candidate markers that were identified on the bulk screening were subsequently analyzed on ~200 randomly chosen individuals of the mapping population. The usefulness of BSA based approach has been demonstrated in detecting QTLs related with erucic acid content in oilseed rape [82]. Another approach, RQM, relies on the selective phenotyping and genotyping with a limited number of individuals/markers in a two step procedure [84]. The first step is a classical QTL analysis on a fraction of a segregating population (e.g., 200 individuals) to identify the possible major QTLs for the trait of interest. The second step is to use markers flanking the major QTL and screen the entire population (e.g., 2000 individuals) to identify QTL isogenic recombinants (QNIRs): individuals that carry a recombination at one QTL region and bear identical homozygous genotypes at the other QTL. These QNIRs are then genotyped with sufficient markers at the recombinant QTL region to precisely map the recombination events. By increasing progeny or clones of the QNIRs, highly accurate phenotyping data can be obtained, and precise localization of the QTL gene could be reached within a sub-centimorgan interval.

Physical mapping in tropical tree species

Currently, a physical map is usually constructed for an organism with assignment of unique BAC clones or STSs to genetic maps readily constructed. A physical map is an intermediate level of resolution between genetic maps and the full genome sequence and represents a useful framework for map-based gene cloning, EST mapping, comparative genomics, and full genome sequencing. Regretfully, so far as we know, few physical maps have been published for tropical trees though some work is underway.

Genolyptus in Brazil is producing a physical map for *Eucalyptus grandis* using fluorescent fingerprints of BAC clones and BAC end sequencing (http://www.ieugc.up.ac.za/japan_meeting_report.pdf).

Lashermes and colleagues (2004) constructed a BAC library with a multi-disease resistance line of *Coffea arabica* (http://www.asic-cafe.org/pdf/abstract/B203_2004.pdf). The large insert DNA library contains 88 813 clones with an average insert size of 130 kb, and represents approximately eight *C. arabica* haploid genome equivalents. The undertaken mapping approach combined hybridization with mapped markers and BAC fingerprinting. They completed hybridization with both low-copy RFLP markers distributed on the 11 different chromosomes and probes corresponding to disease resistance gene analogs. BAC clones from subgenomes Ea and Ca were assembled into separate contigs. Accuracy of the map was verified using several approaches. These preliminary results represent the first step toward the construction of a physical map of the coffee genome.

The construction of physical maps will become integral to genomics studies in tropical tree species especially for tree genera that do not contain major commercial temperate species such as *Coffea*. Particularly, the construction of partial physical maps should be a preferable option for gene isolation as genomic projects in tropical trees advance. This could be extremely interesting for qualitative traits such as disease resistance and flowering traits.

Genome sequencing in *Eucalyptus*

To date, only one tropical tree species, *Eucalyptus camaldulensis*, has been involved in full genome sequencing. In 2004, Satoshi Tabata and colleagues of Kazusa DNA Research Institute in Japan initiated large-scale genome sequencing of *E. camaldulensis* in collaboration with Oji Paper, who donated genomic DNA of an *E. camaldulensis* clone called CPT1 for the effort (http://www.ieugc.up.ac.za/japan_meeting_report.pdf).

The aim of the sequencing project at Kazusa is to generate a platform from which to introduce genomic approaches to genetic improvement of *Eucalyptus* and not necessarily to produce high-quality sequence of the entire genome.

Kazusa has taken a very focused approach in which they are sequencing mostly genes and associated single-copy regions (e.g. promoters), and are avoiding highly repetitive DNA. The basic strategy for sequencing is a combination of whole-genome shotgun sequencing, BAC-end sequencing and targeted shotgun sequencing designed to maximize the efficiency of sequencing single-copy regions. Kazusa is sequencing the *E. camaldulensis* genome in two phases. The first phase involves a modified shotgun sequencing approach which is designed to provide sequence of most of the euchromatic regions of the genome. The second phase will involve clone-by-clone sequencing of specific gene regions of interest.

The current rate of data collection is approximately 200 000 runs per month, producing sequence data equivalent to over 1 × coverage of the entire genome per year. So far Kazusa has analyzed 1 343 708 files and obtained 1 072 309 data files with total length 681.7 Mb (approx 1× genome equivalent). Kazusa has been able to fully annotate the *E. camaldulensis* chloroplast genome (160,303 bp) and has identified ten highly repetitive sequence classes in the nuclear genome, including 4.5S rRNA and 5S rRNA. This constitutes an estimated 4.27% of the nuclear genome, which is considered to be quite low compared to other plant species. Table 5 summarizes the progress in *Eucalyptus* genome sequencing project at Kazusa by July 2005.

Table 5. Summary of *E. camaldulensis* genome sequencing progress at Kazusa DNA Research Institute by July 2005.

Sequence Type	No of sequences	Status	Release date
BAC End Sequences (BES)	112 500	Finished	End 2005
Whole Genome Shotgun (WGS)	1 072 309	Finished	Not planned
Selected BAC Mixture (SBX)	224 061	200 000/month (1× genome/year)	Not planned

Conclusions and perspectives

In conclusion, considerable progresses have been made to date in genome mapping of tropical tree species. Genetic maps have been constructed for a number of species, e.g. tropical eucalypts, tropical pines, tropical acacias, coffees, and rubber trees, on basis of which QTL detection has been carried out for a variety of agronomically important traits. Physical map construction is underway for eucalypt and coffee. Full genome sequencing is in progress in *Eucalyptus camaldulensis*. All these advances will form a necessary basis for future structural and functional genome exploration of tree genomes and one day allow the artificial design of superior varieties in breeding programs.

Nevertheless, several problems exist in genome mapping of tropical trees, which should be the focus of future research. Marker density in genetic maps needs to be increased, especially with cross-species transferable markers, such as SSR and EST-derived markers. QTL mapping and association mapping for commercially important traits is required based on sound experimental strategies. Comparative mapping within and between species should be the priority of research in map construction. Physical maps based on the construction of BAC libraries, will play a crucial role in anchoring genetic markers to overlapping BAC clones and help to facilitate map-based cloning of genes. Though full genome sequencing will be impossible for most tropical tree species of interest, partial genome sequencing targeting on QTL- or gene-harboring regions could be a feasible option.

Acknowledgments

The authors thank Dr Gavin Moran for his critical reading and comments as well as the National Natural Science Foundation of China (Grant No. 30371173) and the Ministry of Education of China (An initiating fellowship for fellows returning from abroad) for financial support.

References

1. Bernardi, G. 1995, Annu. Rev. Genet., 29, 445.
2. Chawla, H.S. 2002, Introduction to Plant Biotechnology (2nd edition), Science Publishers, Inc., Enfield.
3. Lander, E.S. 1996, Science, 274, 536.
4. Collins, F., and Galas, D. 1993, Science, 262, 43.
5. Bridges, C.B. 1935, J. Hered., 26, 60.
6. Strauss, S.H., and Conkle, M.T. 1986, Theor. Appl. Genet., 72, 483.
7. Adams, W.T., Neale, D.B., Doerksen, A.H., and Smith, D.B. 2000, Silvae Genetica, 39, 153.
8. Grodzicker, T., Williams, J., Sharp, P., and Sambrook, J. 1974, Cold Spring Harbor Symp. Quant. Biol., 39, 439.
9. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G.., and Erlich, H. 1986, Cold Spring Harbor Symp. Quant. Biol., 51, 263.
10. Harushima, Y., Yano, M., Shomura, A., Sato, M., Shimano, T., Kuboki, Y., Yamamoto, T., Lin, S.Y., Antonio, B.A., Parco, A., Kajiya, H., Huang, N., Yamamoto, K., Nagamura, Y., Kurata, N., Khush, G.S., and Sasaki, T. 1998, Genetics, 148, 479.
11. Chen, M., Presting, G., Barbazuk, W.B., Goicoechea, J.L., Blackmon, B., Fang, G., Kim, H., Fresch, D., Yu, Y., Sun, S., Higingbottom, S., Phimphilai, J., Phimphilai, D., Thurmond, S., Gaudette, B., Li, P., Liu, J., Hatfield, J., Main, D., Farrar, K., Henderson, C., Barnett, L., Costa, R., Williams, B., Walser, S., Atkins, M., Hall, C., Budiman, M.A., Tomkins, J.P., Luo, M., Bancroft, I., Salse, J., Regad, F., Mohapatra, T., Singh, N.K., Tyagi, A.K., Soderlund, C.S., Dean, R.A., and Wing, R.A. 2002, Plant Cell, 14, 537.

12. Qi, L.L., Echalier, B., Chao, S., Lazo, G.R., Butler, G.E., Anderson, O.D., Akhunov, E.D., Dvoák, J., Linkiewicz, A.M., Ratnasiri, A., Dubcovsky, J., Bermudez-Kandianis, C.E., Greene, R.A., Kantety, R., La Rota, C.M., Munkvold, J.D., Sorrells, S.F., Sorrells, M.E., Dilbirligi, M., Sidhu, D., Erayman, M., Randhawa, H.S., Sandhu, D., Bondareva, S.N., Gill, K.S., Mahmoud, A.A., Ma, X.-F., Miftahudin, Gustafson, J.P., Conley, E.J., Nduati, V., Gonzalez-Hernandez, J.L., Anderson, J.A., Peng, J.H., Lapitan, N.L.V., Hossain, K.G., Kalavacharla, V., Kianian, S.F., Pathan, M.S., Zhang, D.S., Nguyen, H.T., Choi, E.-W., Fenton, R.D., Close, T.J., McGuire, P.E., Qualset, C.O., and Gill, B.S. 2004, Genetics, 168, 701.
13. The Arabidopsis Genome Initiative. 2000, Nature, 408, 796.
14. Yu, J., Hu, S., Wang, J., Wong, G.K.-S, Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., Cao, M., Liu, J., Sun, J., Tang, J., Chen, Y., Huang, X., Lin, Wei., Ye, Chen., Tong, W., Cong, L., Geng, J., Han, Y., Li, L., Li, W., Hu, G., Huang, X., Li, X., Li, J., Liu, Z., Li, L., Liu, J., Qi, Q., Liu, J., Li, L., Li, T., Wang, X., Lu, H., Wu, T., Zhu, M., Ni, P., Han, H., Dong, W., Ren, X., Feng, X., Cui, P., Li, X., Wang, H., Xu, X., Zhai, W., Xu, Z., Zhang, J., He, S., Zhang, J., Xu, J., Zhang, K., Zheng, X., Dong, J., Zeng, W., Tao, L., Ye, J., Tan, J., Ren, X., Chen, X., He, J., Liu, D., Tian, W., Tian, C., Xia, H., Bao, Q., Li, G., Gao, H., Cao, T., Wang, J., Zhao, W., Li, P., Chen, W., Wang, X., Zhang, Y., Hu, J., Wang, J., Liu, S., Yang, J., Zhang, G., Xiong, Y., Li, Z., Mao, L., Zhou, C., Zhu, Z., Chen, R., Hao, B., Zheng, W., Chen, S., Guo, W., Li, G., Liu, S., Tao, M., Wang, J., Zhu, L., Yuan, L., and Yang, H. 2000. Science, 296, 79.
15. International Rice Genome Sequencing Project. 2005, Nature, 436, 793.
16. Grattapaglia, D., and Sederoff, R. 1994, Genetics, 137, 837.
17. Butcher, P.A., and Moran, G.F. 2000, Theor. Appl. Genet., 101, 594.
18. Jeffrey, A.J., Wilson, V., and Thein, S.L. 1985, Nature, 314, 67.
19. Glaubitz, J.C., and Moran, G.F. 2000, Forest Conservation Genetics: Principles and Practice, A.Young, T. Boyle, and D. Boshier (eds), CSIRO Publishing, Melbourne, 39.
20. Williams, J.G., Kubelick, A.R., Livak, K.J., Rafalski, J.A., and Tingey, S.V. 1990, Nucleic Acids Res., 18, 6531.
21. Schlotterer, C., Amos, B., and Tautz, D. 1991, Nature, 354, 63.
22. Zietkiewicz, E., Rafalski, A., and Labuda, D. 1994, Genomics, 20, 176.
23. Vos, P., Hogers, R., Bleeker, M., Reijans, M., van der Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. 1995, Nucleic Acids Res., 23, 4407.
24. Ottewell, K.M., Donnellan, S.C., Moran, G.F., and Paton, D.C. 2005, J. Hered., 96, 445.
25. Butcher, P.A., Decroocq, S., Gray, Y., and Moran, G.F. 2000, Theor. Appl. Genet., 101, 594.
26. Echt, C.S., and May-Marquard, P. 1997, Genome, 40, 9.
27. Scotti, I., Magni, F., Fink, R., Powell, W., Binelli, G., and Hedley, P.E. 2000, Genome, 43, 41.
28. Liewlaksaneeyanawin, C., Ritland, C.E., El-Kassaby, Y.A., and Ritland, K. 2004, Theor. Appl. Genet., 109, 361.
29. Bhat, P.R., Krishnakumar, V., Hendre, P.S., Rajendrakumar, P., Varshney, R.K., and Aggarwal, R.K. 2005, Mol. Ecol. Notes, 5, 80.

30. Ceresini, P.L., Silva, C.L.S.P., Missio, R.F., Souza, E.C., Fischer, C. N., Guillherme, I. R., Gregorio, I., da Silva, E.H.T., Cicarelli, R.M.B., da Silva M.T.A., Garcia, J.F., Avelar G.A., Neto, L.R.P., Marcon, A.R., Junior, M.B., and Marini D.C. 2005, Genetics Mol. Biol., 28, 589.
31. Gonzalez, I.L., Chambers, C., Gorski, J.L., Stambolian, D., Schmickel, R.D., and Sylvester, J.E. 1990, J. Mol. Bio., 212, 27.
32. Nikiforov, T.T., Rendle, R.B., Goelat, P., Rogers, Y.-H., Kotewicz, M.L., Anderson, S., Trainor, G.L., and Knapp, M.R. 1994, Nucleic Acid. Res., 22, 4167.
33. Thumma, B.R., Nolan, M.F., Evans, R., and Moran, G.F. 2005, Genetics, 171, 1257.
34. Paran, I., and Michelmore, R.W. 1993, Theor. Appl. Genet., 85, 985.
35. Tang, X., Tashiro, H., Eki, T., Murakami, Y., Soeda, E., Sakakura, T., Watkins, P.C., and Yokoyama, K. 1992, Genomics, 14, 185.
36. Kere, J., Nagaraja, R., Mumm, S., Ciccodicola, A., D'Urso, M., and Schlessinger, D. 1992, Genomics, 14, 241.
37. Beier, D.R., Dushkin, H., and Sussman, D.J. 1992, Proc. Natl. Acad. Sci. USA, 1992, 89, 9102.
38. Myers, R.M., Maniatis, T., and Lerman, L.S. 1987, Methods Enzymol., 155, 501.
39. Akopyanz, N., Bukanov, N.O., Westblom, T.U., Kresovich, S., and Berg, D.E. 1992, Nucleic Acid. Res., 20, 6221.
40. Cervera, M.T., Sewell, M.M., Faivre-Rampant, P., Storme, V., and Boerjan, W. 2004, Molecular Genetics and Breeding of Forest Trees, S. Kumar, and M. Fladung (eds.), Food Products Press, New York, 387.
41. Tulsieram, L.K., Glaubitz, J.C., Kiss, G., and Carlson, J.E. 1992, Biotechnology, 10, 686.
42. Paillard, M., Lashermes, P., and Petiard, V. 1996, Theor. Appl. Genet., 93, 41.
43. Byrne, M., Murrell, J.C., Allen, B., and Moran, G.F. 1995, Theor. Appl. Genet., 91, 869.
44. Tani, N., Takahashi, T., Iwata, H., Mukai, Y., Ujino-Ihara, T., Matsumoto, A., Yoshimura, K., Yoshimaru, H., Murai, M., Nagasaka, K., and Tsumura, Y. 2003, Genetics, 165, 1551.
45. Gan, S., Shi, J., Bai, J., Wu, K., and Wu, J. 2003, Scientia Silvae Sinicae, 39(2), 162.
46. Marshall, P., M, M.C., Lisieczko, Z., and Landry, B.S. 1994, Theor. Appl. Genet., 89, 853.
47. Maliepaard, C., Jansen, J., and Van Ooijen, J.W. 1997, Genet Res., 70, 237.
48. Lincoln, S., Daly, M., and Lander, E. 1992, Whitehead Institute Technical Report.
49. Manly, K.F., and Elliot, R.W. 1991, Mam. Genome, 1, 123.
50. Nelson, J.C. 2005, The Handbook of Plant Genome Mapping. Genetic and Physical Mapping, K. Meksem, and G. Kahl (eds), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 53.
51. Stam, P. 1993, Plant J., 8, 739.
52. Ling, S. 1999, PhD thesis, University of California, Berkeley, California.
53. Butcher, P.A., Williams, E.R., Whitaker, D., Ling, S., Speed, T.P., and Moran, G.F. 2002, Theor. Appl. Genet., 104, 1185.
54. Verhaegen, D., and Plomion, C. 1996, Genome, 39, 1051.
55. Marques, C.M., Araújo, J.A., Ferreira J.G., Whetten, R, O'Malley, D.M., Liu, B.-H., and Sederoff, R. 1998, Theor. Appl. Genet., 96, 727.

56. Agrama, H.A., George, T.L., and Salah, S.F. 2002, Silvae Genet., 51, 201.
57. Brondani, R.P.V., Brondani, C., and Grattapaglia, D. 2002, Mol. Genet. Genomics, 267, 338.
58. Myburg, A.A., Griffin, A.R., Sederoff, R.R., and Whetten, R.W. 2003, Theor. Appl.Genet., 107, 1028.
59. Gan, S., Shi, J., Li, M., Wu, K., Wu, J., and Bai, J. 2003, Genetica, 118, 59.
60. Nelson, C.D., Nance, W.L., and Doudrick, R.L. 1993, Theor. Appl. Genet., 87, 145.
61. Kubisiak, T.L., Nelson, C.D., Nance, W.L., and Stine, M. 1995, Theor. Appl. Genet., 90, 1119.
62. Shepherd, M., Cross, M., Dieters, M.J., and Henry, R. 2003, Theor. Appl. Genet., 106, 1409.
63. Ky C.-L., Barre, P., Lorieux, M., Trouslot, P., Akaffou, S., Louarn, J., Charrier, A., Hamon, S., and Noirot, M. 2000, Theor. Appl. Genet., 101, 669.
64. Lashermes, P., Combes, M.C., Prakash, N.S., Trouslot, P., Lorieux, M., and Charrier, A. 2001, Genome, 44, 589.
65. Pearl, H.M., Nagai, C., Moore, P.H., Steiger, D.L., Osgood, R.V., and Ming, R. 2004, Theor. Appl. Genet., 108, 829.
66. Lespinasse, D., Rodier-Goud, M., Grivet, L., Leconte, A., Legnate, H., Seguin, M. 2000, Theor. Appl. Genet., 100, 127.
67. Xu, Y, Zhu, L., Xiao, J., Huang, N., and McCouch, S.R. 1997, Mol. Gen. Genet., 253, 535.
68. Temesgen, B., Brown, G.R., Harry, D.E., Kinlaw, C.S., Sewell, M.M., and Neale, D.V. 2001, Theor. Appl. Genet., 102, 664.
69. Butcher, P.A., Glaubitz, J.C., and Moran, G.F. 2000, Forest Genetics Res. Info., 27, 34.
70. Geldermann, H. 1975, Theor. Appl. Genet., 46, 319.
71. Beavis, W.D. 1998, Molecular dissection of complex traits, A.H. Paterson (ed.), CRC Press, Boca Raton, 145.
72. Grattapaglia, D., Bertolucci, F.L.G., and Sederoff, R. 1995, Theor. Appl. Genet., 90, 933.
73. Grattapaglia, D., Bertolucci, F.L.G., Penchel, R., and Sederoff, R. 1996, Genetics, 144, 1205.
74. Verhaegen, D., Plomion, C., Gion, J.-M., Poitel, M., Costa, P., and Kremer, A. 1997, Theor. Appl. Genet., 95, 597.
75. Marques, C.M., Vasquez-Kool, J., Carocha, V.J., Ferreira, J.G., O'Malley, D.M., Liu, B.-H., and Sederoff, R. 1999, Theor. Appl. Genet., 99, 936.
76. Shepherd, M., Chaparr, J.X., and Teasdale, R. 1999, Theor. Appl. Genet., 99, 1207.
77. Junghans, D.T., Alfenas, A.C., Brommonschenkel, S.H., Oda, S., Mello, E.J., and Grattapaglia, D. 2003, Theor. Appl. Genet., 108, 175.
78. Shepherd, M., Cross, M., Dieters, M.J., and Henry, R. 2002, Ann. For. Sci., 52, 617.
79. Shepherd, M., Cross, M., Dieters, M.J., Harding, K., Kain, D., and Henry, R. 2003, Can. J. For. Res., 33, 1923.
80. Lespinasse, D., Grive, L., Troispoux, V., Rodier-Goud, M., Pinar, F., Seguin, M. 2000, Theor. Appl. Genet., 100, 975.

81. Grattapaglia, D. 2001, Proceedings of the IUFRO International Symposium "Developing the Eucalypt of the Future", Valdivia, chile, September 10-15, 2001.
82. Peleman, J.D., Sorensen, A.P., and van der Voort J.R. 2005, The Handbook of Plant Genome Mapping. Genetic and Physical Mapping, K. Meksem, and G. Kahl (eds), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 109.
83. Michelmore, R.W., Paran, I. and Kesseli, R.V. 1991, Proc. Natl. Acad. Sci. USA, 88, 9828.
84. Wye, C., van der Voort, J.R., and Peleman, J. 2000, Plant & Animal Genome Conference VIII, San Diego, poster #538.

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

Molecular Biology of Tropical Plants 2006 105-141 ISBN 81-308-0138-8
Editor Claudine Franche

Tropical Legume Nodulating Bacteria

N'zoue, A., Domergue, O , Moulin, L., Avarre, J -C and de Lajudie, P
Laboratoire des Symbioses Tropicales et Mediterranéennes (L S T M)
UMR 113 IRD/CIRAD/AGRO-M/UM-II USC INRA1242 Campus
International de Baillarguet TA 10 / J 34398 Montpellier Cedex 5 France

Abstract

Legumes contribute to soil fertility and to 33% of global human food needs and up to 80% in developing countries This success is mainly due to the symbiosis they establish with nitrogen-fixing bacteria This symbiosis is characterised by (1) its specificity between plants and bacteria (2) a plant can be nodulated by several different bacterial species (3) the nitrogen-fixing efficiency depends on both partners The techniques used for bacterial characterisation have greatly evolved during the past twenty years and include increasing molecular techniques In the same time taxonomical and diversity studies on Legume Nodulating Bacteria ("LNB") have multiplied,

Correspondence/Reprint request Dr de Lajudie P Laboratoire des Symbioses Tropicales et Mediterraneennes (L S T M) UMR 113 IRD/CIRAD/AGRO M/UM II USC INRA1242 Campus International de Baillarguet TA 10 / J 34398 Montpellier Cedex 5 France E mail P De Lajudie@mpl ird fr

in particular on rhizobia associated to understudied plants and regions so far. Numerous new species have been proposed, and at present LNBs include over 13 genera and 50 valid species. Phylogenetically LNBs belong to six families inside the sub-phyla α and β of the Proteobacteria phylum. They are intertwined with other bacterial species, several of which being involved in other kinds of interactions with plants (pathogenicity, growth promotion) or with mammals (pathogenicity). Here we present the recent advances in LNB taxonomy, focusing on bacteria associated to tropical legumes, emphasizing on unexpected functions associated to them, stem nodulation, free-living nitrogen-fixation, photosynthesis, endophytic association with non-legumes, biodegradative properties, opening new perspectives as fundamental models for research and for applications in future.

Introduction

In many parts of the world, especially in the tropics, humans must face arable soil needs and durable food security ensurement due to demographical pressure and deep socio-economical changes. In this perspective, the considerable, but understudied biodiversity constitutes a resource to be discovered and exploited. There are about 5000 validly described prokaryotic species but it is generally agreed that the ratio of cultivable to microscopically detectable prokaryotes is 0.3 - 1% [1]. In particular, bacteria are abundant in soils - up to 10^9 cells per gram [2] and diverse - a minimum of 4000-7000 different genomes per gram [3], many of them interacting with plants. Some of them can live endophytically and some are elected by legumes to develop nitrogen-fixing symbioses. During these specific interactions, the bacteria enter root tissues via root hairs or directly by crack entry via wounded tissues and induce nodule formation on roots and/or shoots. Inside the nodule, they fix nitrogen for the benefit of the plant. Biological nitrogen fixation (BNF) is particularly important in the tropics, where acid and degraded soils with poor nutrient contents represent a major limitation to agriculture. Developing countries, especially tropical ones, may most directly benefit from BNF for sustainable agriculture and environment maintenance, to combat world hunger by producing cheaper proteins for animal and human consumption.

Although legume biodiversity is mostly concentrated in tropical regions [4], biologists are just beginning to characterize the symbiotic relationships between root-nodule bacteria and legumes in these environments. Early studies focused on cultivated plants, but during the past 20 years, several authors reported isolation of rhizobia from previously uninvestigated wild legumes in different parts of the world, especially in tropical ecosystems [5]. The plant family Fabaceae is subdivided into three subfamilies: Mimosoideae, Caesalpinioideae and Papilionoideae, and not all of them are nodulated. While few species within the more primitive subfamily Caesalpinioideae (23%) nodulate,

nodulation is predominant among species within the subfamilies Mimosoideae and Papilionoideae (90-97% respectively), considered to have evolved from the Caesalpinioideae [6-8].

Rhizobial strains from the tropics are stored in several international collections like ATCC and USDA (USA), CNPBS/EMBRAPA/INPA (Brazil), ORS/STM (France), BCCM^{TM}/LMG (Belgium), CFN (Mexico), CIAT (Colombia), HAMBI (Finland), MIRCENs. However, the majority of the rhizobial biodiversity probably remains to be discovered since a maximum of 23% of the legume species have been examined for nodulation, i.e. 3856 species from which 3397 (88%) do nodulate [9]. Symbiotic rhizobia from around 11200 leguminous species are completely unknown around the world.

For a century, since their discovery in 1889, all Legume Nodulating Bacteria (LNB) were classified in the genus *Rhizobium*, and their species were created following the criterium of their plant species of isolation. A first distinction was made with the introduction of the genus *Bradyrhizobium* for slow-growing rhizobial strains differing in a number of distinctive features [10, 11]. By this time it was acknowledged that the taxonomy of " rhizobia " had to follow the rules of the general bacterial taxonomy, the so-called polyphasic taxonomy, which, through a combination of phenotypic and genotypic techniques with different resolutions at the strain, species, genus or family levels, integrate several molecular, genetic and plylogenetic data in the species description. DNA-DNA hybridisation is crucial for the determination of species, and 16S rDNA sequencing play an important role in the genus attribution to bacteria and the evaluation of the phylogenetic relationships between organisms [12].

Over the last few years, the increasing amount of research focused on bacteria that nodulate stems or roots of legumes has demonstrated the unexpected large biodiversity among legume nodulating bacteria at the levels of strain, species, genus, family and class. Many of the recently described "new rhizobia" have thus been isolated from tropical legumes. Atypical nitrogen-fixing symbioses from natural tropical ecosystems are characterised by their adaptation to a wide range of environmental and climatic conditions, varying from deserts to waterlogged areas and from savannahs to tropical rain forests. In each of these contrasted environments, organic soil nitrogen appears as one main limiting factor for plant growth, and biological nitrogen fixation thus represents the major input of N in the system. In contrast with temperate ecosystems, this high requirement for nitrogen has entailed a greater diversity of symbioses between rhizobia and legumes [13].

In this chapter, we will consider modern bacterial taxonomy, and describe recent advances concerning bacteria associated to plants, especially legume nodulating bacteria.

General bacterial taxonomy

The concept of bacterial species has progressively evolved over the last century [14-17]. The evolutionary processes create patterns of biodiversity [18], and the underlying basis of systematics is evolution [19]. There is a consensus that the small subunit of ribosomal DNA (SSU rDNA)-based phylogenies are largely consistent with the evolutionary history of the organisms, since the groups formed using this approach are often confirmed by other data. The 16S rDNA is thus generally accepted as the ultimate molecular chronometer. The intragenomic heterogeneity between multiple 16S rDNA operons is rather limited and is unlikely to have a profound effect on the classification of taxa. Specifically, whole-genome sequences of rhizobia show that *M. loti*, *S. meliloti* and *A. tumefaciens* have 2, 3 and 4 identical copies of rDNA, respectively [20]. There is a general consensus that a bacterial species should be monophyletic or at least approximately so (over 97% of 16S rDNA sequence homology), which means that most of the DNA of the members of a species should be derived from a single common ancestor. A single isolate is declared the "type strain" and all isolates sufficiently similar to it are included in the same species. The similarity is expected to be polyphasic, i.e. based on as many features as possible, but in practice two techniques have emerged as the main arbiters: total DNA hybridisation and phylogenetic reconstruction from SSU rDNA sequences. Members of a species should form stable heteroduplexes and display at least 70% of DNA-DNA relatedness value, with a melting temperature (Tm) difference of less than 5°C. This indicates that they share similar sequences for most of their genome. In contrast, members of a genus should form a well-supported clade in the SSU rDNA phylogeny [21].

Legume Nodulating Bacteria (LNB) taxonomy

The general terms of "rhizobium" or "rhizobia" refer to nitrogen-fixing bacteria capable of living in symbiosis with legume plants. They are taxonomically very diverse, and over the last twenty years, their classification has undergone great changes, especially because of the discovery of new LNB species from uninvestigated plants, mainly from the tropical and subtropical regions. From one genus including four species in 1981, the classification now includes at least 13 genera comprising more than 50 species, and the number continues to increase. The current systematics and nomenclature of rhizobial bacteria can be found at the following URLs: http://www.bacterio.cict.fr/, http://www.cme.msu.edu/bergeys/page7 and http://www.rhizobia.co.nz. In 2004, Zakhia *et al.* [22] proposed the term LNB (Legume Nodulating Bacteria) to avoid confusion between the general term of rhizobium and the genus name.

At present all LNB described so far belong to the Proteobacteria and represent at least 9 monophyletic groups (Table 1). The majority of them belong to genera of the α-class of Proteobacteria, namely *Rhizobium, Mesorhizobium, Sinorhizobium, Allorhizobium, Bradyrhizobium, Azorhizobium.* Recently LNB were surprisingly discovered in genera from the α-class of Proteobacteria, i.e. *Methylobacterium* [23] [24], *Devosia* [25], *Blastobacter* [26], *Ensifer* [27], *Ochrobactrum* [28] and *Phyllobacterium* (Mantelin S., unpublished). LNB were also discovered in the β–class of Proteobacteria, namely in *Burkholderia* and *Ralstonia* [29, 30]. Moreover, Benhizia *et al.* [31] reported the association between the legume *Hedysarum* and strains belonging to the γ–class of Proteobacteria, i.e. *Pantoea agglomerans, Enterobacter kobei, Enterobacter cloacae, Leclercia adecarboxylata, Escherichia vulneris* and *Pseudomonas* sp., although their nodulating ability was not clearly demonstrated. All LNB are phylogenetically intertwined with *a priori* non-nodulating genera like the soil bacteria *Mycoplana* (branching soil bacteria), animal pathogens *Brucella, Burkholderia, Bartonella* and *Afipia*, plant pathogens like *Agrobacterium* and *Ralstonia*, photosynthetic bacteria like *Rhodopseudomonas* and chemoautotrophs like *Xanthobacter*. However, in the past, several genera phylogenetically close to LNBs like *Agrobacterium, Ensifer, Ochrobactrum and Phyllobacterium* (leaf and rhizosphere colonists), first described as non-nodulating, were recently reported to include nodulating strains or to be able to acquire nodulation capacity upon acquisition of symbiotic genes by lateral transfer [32-34]. Similar findings are most probably to be expected in the near future in other genera closely related to LNBs and this may rise safety concerns. *Bartonella* are intracellular parasites of red blood cells and endothelial cells. They are re-emerging human pathogens associated to cat scratch disease and human endocarditis. Demba-Diallo [35] performed PCR amplication of 16S rDNA on microbial DNA from soil samples collected under *Acacia tortilis* subsp. raddiana in Senegal. They identified γ-Proteobacteria (35%), Firmicutes (24%), α–Proteobacteria, β–Proteobacteria, Acidobacteria and Actinobacteria, varying according to the season. It would be interesting to isolate and test such bacteria for their nodulating ability or their capacity to acquire it, in order to conclude on the potential of such soil bacteria to become LNBs.

Unexpected recently discovered LNB accompanied with new functions and opening to new applications

Since a previous review [36], Young *et al.* [37] proposed the inclusion of *Agrobacterium* and *Allorhizobium* into *Rhizobium*, but this remains controversial.

Several new LNB species have also been proposed, and will be described in more details. *Ensifer adhaerens* was proposed by Casida [38] as a soil bacterium capable to adhere to and lyse other soil bacteria. It was mainly described on the basis of phenotypic features. It is neither nutritionally fastidious nor an obligate predator. *Ensifer adhaerens* was only recently recognised as phylogenetically intertwined with *Sinorhizobium* species and Rogel *et al.* [39] demonstrated that it can effectively nodulate *Phaseolus vulgaris* and *Leucaena leucocephala* when provided with symbiotic plasmids of *R. tropici*. *E. adhaerens* nodule isolates were later described [27], but reclassification of *E. adhaerens* as *Sinorhizobium adhaerens* [27] is controversial due to taxonomic rules [40]. *S. morelense*, isolated from *Leucaena leucocephala* [41], is phylogenetically intertwined with *Ensifer adhaerens* strains, and was proposed to be a later heterotypic synonym of *Ensifer adhaerens* [40]. *Bradyrhizobium betae* sp. nov. was proposed as an endophyte isolated from roots of *Beta vulgaris* affected by tumor-like deformations, but was not shown to be the causal agent of the symptoms [42]. *Azorhizobium johannense* sp. nov. was proposed for strains nodulating *Sesbania virgata* in Brazil [43]. Rivas *et al.* [25, 44] characterized LNB strains from *Neptunia natans* in India as *Devosia neptuniae*. Since then Vannini *et al.* [45] interestingly identified *Devosia* sp. nov. strains as endosymbionts inhabiting the cytoplasm of the marine ciliated protozoon *Euplotes magnicirratus*.

Phylogenetic relatedness of certain LNB with bacteria of same or separate genera known for their original properties may give some indication on the biological significance and may predict unexpected potentialities in new LNBs. LNB characterisation is thus a prerequisite for optimal valorisation of natural nitrogen-fixing symbioses in the tropics. Behind the diversity of bacteria, lies the diversity of the adaptative mechanisms, molecules and genes... Rhizobia were mainly known for nodulation so far, but in recent years, other unexpected and useful functions were discovered, like stem nodulation, free living nitrogen fixation, metabolic diversity, methylotrophy, photosynthesis, endosymbiosis with non-legume plants, and biodegradation of recalcitrant compounds. Stem nodulation is associated with the tropical plant genera *Aeschynomene, Sesbania, Neptunia* and *Discolobium*. Stem nodulation and free living nitrogen fixation were reviewed by Boivin *et al.* [46] and will not be developed here. The metabolic diversity appears as an advantage for colonizing plants adapted to special environments, and especially to legumes producing unusual metabolites.

Methylotrophy

Crotalaria (Papilionoideae) is a well represented plant genus in Africa with 500 species, herbs and shrubs, majoritarily nodulated by *Bradyrhizobium*

sp. strains [47] [48] *Crotalaria* are useful in agriculture for green manuring, nematode population control and mycorhization enhancing [49]. *Methylobacterium* LNB were first found as nodule symbionts of *Crotalaria glaucoides, C. perrottettii, C. podocarpa* and *C. spherocarpa* in Senegal and the name *M. nodulans* was proposed for the new species consisting of this bacterial group [23, 24, 48]. *Methylobacterium* LNB strains were then discovered in South Africa in nodules of *Lotononis bainesii* [50], and in other plants of the *Lotononis, Indigofera* and *Calpurnia* genera [49] [51]. *Nod*A gene of *M. nodulans* is most similar to that of *Bradyrhizobium. Methylobacterium* members are ubiquitous in nature and have been detected in soil, dust, freshwater, lake sediments, on leaf surfaces and nodules, in rice grains, air and hospital environments. *Methylobacterium* could have acquired the capacity to nodulate *Crotalaria* due to their trophic functions associated to methylotrophy. Indeed, *Methylobacterium* members are often able to use and produce various complex organic chemicals, aromatic and aliphatic long chain complexes, "suies" and others from car pollution [52, 53]. Alkaloids (toxic secondary metabolites) may play a role in symbiotic specificity and nematode population control [24] [49]. *Lotononis* and *Crotalaria* synthesize methylated C1 molecules which, almost exclusively can be degraded by *Methylobacterium. M. nodulans* is a facultative methylotroph capable to grow on C1 components like formate, formaldehyde and methanol as sole Carbon source. *M. nodulans* strains express the methanol dehydrogenase (*mxaF*) genes in the apical zone of the nodule. Symbiosis with *mxaF*- mutants (loss of methylotrophy) is correlated with a plant growth decrease of 30%, indicating that methylotrophy is required for efficient symbiosis. However, not all *Methylobacterium* LNB are methylotrophs, and methylotrophy is rare in other rhizobia. This feature was only reported for a group of *Mesorhizobium* able to use methanol as sole carbon source [54].

Bioremediation

After the first reports of Moulin *et al.* [29] and Chen *et al.* [30], Vandamme *et al.* [55] showed that nodule isolates from the tropical legumes *Aspalathus, Machaerium, Alysicarpus* and *Mimosa* belong to four distinct *Burkholderia* species: *B. caribensis, B. cepacia* genomovar VI, *B. tuberum* and *B. phymatum*, indicating that the capacity to nodulate legume plants has spread among multiple *Burkholderia* species. *Burkholderia* cepacia genomovar VI was later accomodated as *Burkholderia dolosa* sp. nov. by Vermis *et al.* [56]. *Burkholderia cepacia*-like organisms, although originally known as plant pathogens, have attracted much interest for the agricultural industry as natural promoters of plant growth and biological control agents, as well as for bioremediation. They occupy diverse ecological niches, which may raise safety problem when considering plant crop field inoculation. *Burkholderia* are able

to degrade methylterbutylester, toluene, hexane, ethanol, acetone, petrol. *Burkholderia* constitutes a phylogenetically well defined group, but functionally very diverse. Several *Burkholderia* have been completely sequenced. The genus *Burkholderia* includes N2-fixing bacteria that are associated with plants. Some are associated to rhizospheres, rhizoplanes and internal tissues of maize, sugarcane, sorghum and coffee plants [57].

During the last decade, *B. cepacia* has caused great concern, since it has been recognised as an opportunistic human pathogen, especially among cystic fibrosis patients for whom this multi-resistant bacterium is a major pathogen. *B. cepacia* isolates belong to at least nine distinct genomic species (genomovars), referred to as the *B. cepacia* complex. So far, there are no phenotypic, genomic or taxonomic grounds to differentiate clinical and environmental members of this complex [58]. Roots and rhizospheres of various economically important crops including corn, maize, rice, pea, sunflower and radish can be colonized by *B. cepacia*-like organisms, some of which producing a variety of antimicrobial compounds that are active against soil pathogens [58].

R. taiwanensis represent 93% of the *Mimosa pudica and diplotricha* isolates in Taiwan, showing a high degree of specificity between plant and bacterial partners [59]. *Ralstonia eutropha, R. oxalatica, R. basilensis, R. campinensis, R. metallidurans*, environmental organisms, all displaying potential ability in bioremediation of soils and waters polluted by heavy metals and chloride organic compounds, were reclassified with *R. taïwanensis* in *Wautersia* gen. nov. by Vaneechoutte [60] and more recently in Cupriavidus. *Ralstonia taïwanensis* was also found in clinical samples, as an opportunistic human pathogen. *Ralstonia* are known for their ability to survive in oligotrophic environments with plant pathogens.

Photosynthesis

Photosynthetic *Bradyrhizobium* were isolated from stem nodules of *Aeschynomene sensitiva* and *A. indica*, belonging to cross inoculation group III [61]. The *nod*A gene of photosynthetic *Bradyrhizobium* is dissimilar from others [62]. Photosynthetic *Bradyrhizobium* synthetize bacteriochlorophyll A as well as carotenoids [63, 64]. Although aerobes, they are phylogenetically close to *Rhodopseudomonas palustris*, a purple photosynthetic bacterium able, in anaerobiosis, to use CO_2 as the sole carbon source through photosynthesis. They constitute a separate branch among other *Bradyrhizobium* [65, 66] and they could represent at least two separate genospecies [67]. Photosynthetic *Bradyrhizobium* have the unique ability among rhizobia to utilize energy from light, inorganic or organic compounds allowing growth and survival under a wide range of conditions. The photosynthetic genes are expressed in the stem nodule at different stages during symbiosis [68, 69]. Phytochrome induces

photosynthesis under far red light (752 nm). Photons are capted by carotenoids, bacteriochlorophyll A, phytochrome and canthaxanthin. Signal is interpreted and transmitted to activate the bacterial photosystem. Bacteriochlorophyll A is oxydated and electrons are transmitted to the cytochrom chain, resulting in an H+ gradient and production of ATP. This is a case of anoxygenic photosynthesis, without any O_2 production and C fixation. Mutations in photosynthetic *(puf)* genes result in less efficient and lower number of root nodules and poor plant growth [68]. It was later discovered that photosynthetic *Bradyrhizobium* are also endophytic of rice roots and strongly enhance the growth of some rice varieties (14% to 32% of heigth gain, depending on the rice variety). This was observed on *O. breviligulata*, an ancestral rice, and *O. glaberima*, a cultivated rice living with *A. sensitiva* and *A. indica* in temporary ponds in Africa. Bacteria colonize the whole rhizoplane and penetrate intercellularly to the 5th cellular layer [70]. This constitutes a kind of primitive infection as observed on *Aeschynomene*. Since no nitrogen fixation could be demonstrated in rice, the observed plant growth promoting effect upon bacterial inoculation must be due to another factor. The production of canthaxanthin by photosynthetic *Bradyrhizobium* may be of industrial importance. Canthaxanthin, a β-carotene, has a high antioxydative power, protecting cells against photooxydative lesions. It is used in cosmetics and food industry. Canthaxanthin constitutes 85% of the carotenoids produced by the bacterial cells.

In conclusion, the occurence of the photosynthetic ability in rhizobia raises the question of the adaptation of *Bradyrhizobium* to light, aerial life in stem nodules and adaptation of photosynthesis to symbiosis. Rice may constitute a greater reservoir for photosynthetic *Bradyrhizobium* than *Aeschynomene* in nature. Pods are microaerophylic biotopes favorable for photosynthesis, which is an advantage for free living, for plant colonization, inside the nodule to increase nitrogen fixation (less dependance to ATP from the plant).

Endophytic life with non legumes

Microorganisms interact with plants because plants offer a wide diversity of habitats including the phyllosphere (aerial plant part), the rhizosphere (zone of influence of the root system), and the endosphere (internal transport system). Plant-associated microorganisms play essential roles in agricultural and food safety, and contribute to the environmental equilibrium [71]. Many bacterial endophytes have been reported in tropical plants, for instance *Acetobacter diazotrophicus* in sugar cane; *Azoarcus* in xylem of Kallar grass (which are abundant and diffuse systemically); *Azospirillum* in maize and rice; *Herbaspirillum seropedicae* in sugar cane and rice; the nitrogen-fixing *Pseudomonas stutzeri* in rice. Some of them are source of safety concern, like *Serratia marcescens* (enterobacteria, non pigmented, urease +) that can be both

a rice endophyte and a potential human pathogen. *Klebsiella pneumoniae* and *Raoultella planticola* strains have been isolated from diverse plants including rice and maize. Pathogenic *Klebsiella* have been found to successfully colonize potato and lettuce, and plants may be considered to be reservoirs of human opportunistic *Klebsiella.* Inside plants, bacteria encounter new niches where adapted clones may be selected and some endophytic bacteria may represent new species [72]. *K. variicola* was recently described for clinical and plant-associated isolates (banana, rice, sugar cane, maize) on the basis of phylogenetic analysis derived from the sequences of *rpo*B, *gyr*A, *mdh*, *inf*B, *pho*E and *nif*H genes [73].

Some LNB have also been reported to be involved in endophytic relationships with non-legume plants; as already mentioned, photosynthetic *Bradyrhizobium* sp. strains can both induce nitrogen-fixing stem nodules on *Aeschynomene* spp. and be natural endophytes of rice roots in Africa [70]; *Azorhizobium caulinodans* has been found in the rhizosphere [74] and in the roots [75] of rice in Asia. *R. leguminosarum* engages in beneficial endophytic associations with rice, maize and wheat [76-79]; *R. etli* is a natural maize endophyte [80] while *R. tropici* was also found to be a competitive endophyte when inoculated to rice [73]. Some LNB were also found in sorghum and millet. Endophytic *Rhizobium* sp. were isolated from surface-sterilized banana plants [81]. Bananas were recently reported to be associated to *Burkholderia* and *Ochrobactrum* [82], two genera including LNB members [28, 29, 31].

Plant growth promoters

Phyllobacterium genus was originally proposed for bacteria that develop within leaf nodules of tropical ornamental plants [83] with one species, *P. myrsinacearum* [84]. *Phyllobacterium* strains are frequently isolated from different environments. They are associated with various plants, either rhizospheric of *Picea abies* and *Lotus* [85, 86], endophytic in *Zea mays, Gossypium hirsutum* and *Trifolium pratense* [87-90], or in tight connection with roots in *Saccharum officinarum, Beta vulgaris* and *Brassica napus* [91-93]. Several *Phyllobacterium* sp. have a PGP effect directly on root hair and secondary roots, or indirectly via biocontrol, hormone production (Indole Acetic Acid), ethylene reduction, mineral nutrition, nitrate absorption, ion transportation, P solubilizing, or protection against pathogens [93, 94]. In early studies, Van Veen *et al.* [32] reported crown gall tumor and root nodule formation by *Phyllobacterium myrsinacearum* after the introduction of an *Agrobacterium* Ti plasmid or a *Rhizobium* Sym plasmid. Recently several groups characterised LNB strains as *Phyllobacterium* [95] (Zakhia *et al.*, unpublished; Mantelin *et al.*, unpublished).

These observations suggest that LNB may have dynamic lifestyles, alternatively behaving as a soil heterotroph, as a depollutant, as a Plant-Growth-Promoting endophyte and as a legume symbiont. LNB must have

mechanisms to endophytically colonize internal plant tissues, and to survive plant defense reactions, as well as to adequatedly use plant nutrients without damaging the plant. Some bacteria may have a specificity of trophism adaptation near a plant. Whether part of these mechanisms are common pathways towards colonization of legumes and non-legumes is unknown. The biological significance of the taxonomic diversity of tropical legume bacterial symbionts is still pending. This diversity is probably largely underestimated, since new LNB species are regularly discovered, as was recently the case for *Blastobacter denitrificans* [26], *Devosia neptuniae* [25], *Ensifer adhaerans* and even more surprisingly, some beta-Proteobacteria like *Ralstonia taiwanensis* [30] and *Burkholderia* [29] with 4 species: *B. caribensis, B. dolosa, B. tuberum, B. phymatum* [55, 56]. Similar to other *Wautersia* spp., the LNB *Wautersia (Ralstonia, Cupriavidus) taiwanensis* strains [60] may have - or easily acquire - bioremediation qualities for soil and water depollution of heavy metals or chloride organic compounds. One can imagine that in future, legume symbiosis would both ensure detoxification-revegetalisation of polluted soils and act as reservoir for biodegradating bacteria in the environment.

It is likely that many more valuable functions remain to be discovered, most probably governed by a multitude of genes useful for adaptation to the wide variety of ecological niches, edaphic and climatic conditions. These genes could be exploited to improve competitivity of strains under the diverse environmental conditions where the nitrogen-fixing symbioses are to be introduced. Strains of sewage water *Zooglea ramigera* were found to be relatives of the rhizobia group [96]. *Pseudoaminobacter, Chelatobacter* and *Aminobacter*, very useful in biotechnology for bioremediation, show close relationships with *Mesorhizobium* spp. [97, 98]. *Ensifer* isolates are similar to *Sinorhizobium* [27], and *Brucella* constitutes a cluster related to *Rhizobium tropici* [39].

In conclusion, taxonomy helps development of suitable genetic tools, prediction of functions, and facilitates evolutionary studies.

Host specificity

Studies on rhizobium-legume symbioses have long reported the specificity between plant and bacteria. In recent years, they have also described the great taxonomic diversity of the rhizobia. Many rhizobium strains can nodulate a wide range of different plant species and conversely, some plant species are able to host a varying number of different bacterial species [81, 99-101]. Different legumes grown in the same geographic region may nodulate with a single rhizobial species [102-104]. The host spectrum of a rhizobium can be either "narrow" or "large". *Medicago* is considered as a rather specific plant as it is nodulated by a limited number of bacterial species. In contrast, *Phaseolus vulgaris* and *Glycine* display wide spectra of nodulating bacteria. All the nodule isolates of *P. vulgaris* in nature are so far affiliated to the genus

Rhizobium and belong to *R. etli, R. tropici, R. leguminosarum* bv. phaseoli, *R. gallicum* and *R. giardinii. P. vulgaris* is a promiscuous host like other species in the *Phaseolae* genus, such as *Macroptilium* and *Vigna*. Experimental assays in the laboratory under axenic conditions showed that it may be nodulated by many more bacterial species, i.e. 28 different species, belonging to *Rhizobium, Sinorhizobium, Mesorhizobium, Bradyrhizobium* and *Azorhizobium* [81]. However *Phaseolus* is usually considered as a poor nitrogen fixer and problems of successful inoculation, nodulation and nitrogen fixation have not been overcome yet. Benhizia *et al.* [31] stated that the cultivated legumes may have undergone a constant selection pressure by humans, which in turn may explain their observed high specificity of interaction with bacteria. This can be opposed to the non domesticated wild legumes, from which a number of diverse bacteria were isolated in the past few years, corresponding to primary plant-bacteria interactions. This may apply for many tropical legumes, like *Leucaena, Phaseolus, Glycine, Sesbania, Acacia...* (Table 1). The distinction between rhizobia and the bulk of bacteria that grow endophytically within plants is their ability to synthesize Nod factors. Nodulation (*nod*) genes govern the biosynthesis of Nod factors and are unique to rhizobia. Around 60 different nodulation genes have been described so far (*nod, nol, noe*), and they can be classified into two categories [105]: the common genes (*nod* MABC for the synthesis of Nod factor (NF) skeleton, *nod*IJ for NF export, *nod*D involved in NF regulation), and some more specific genes that exist in variable combinations depending on strains (*nod* PQ or *nod* HLZUS involved in the skeleton decoration). *nod*A gene is a key gene for symbiosis establishment, it is present as a single copy and has a constant size (590-660 bp) in all described rhizobia to date [62]. The corresponding protein is a key enzyme for NF synthesis, transferring an acyl chain on a chitin oligomer, resulting in a biologically active molecule. *nod*A gene sequence is informative on the symbiotic characteristics of the rhizobium. It may be used to predict the type of Nod factor and the host specificity [106, 107], as its sequence is usually more related to the bacterial biovar than to its own species.

Nitrogen fixation

Nitrogen-fixing effectiveness depends on the strain. Boivin *et al.* [108] showed that several *Sinorhizobium* and *Azorhizobium* strains can nodulate all *Sesbania* species tested, but *Azorhizobium* strains are effective on *Sesbania rostrata* only. However, as nodulation and nitrogen fixation functions are genetically controlled, it is possible to improve it by selection of both partners. In particular, rhizobia can be selected for their efficiency, their competitiveness among natural populations, and their adaptation to the ecosystem where they are introduced. Experiments performed in Brazil have shown that rates of N2 fixation with soybean can exceed 300 kg of N2 ha^{-1}, accounting for 69 to 94% of

total plant N [109]; furthermore there are also benefits due to the release of N to the next crop. In comparison it represents 50% in USA. The success of this symbiosis in Brazil results from breeding programs that aimed at identifying both plant and *Bradyrhizobium* genotypes carrying a higher capacity of N2 fixation [109, 110]. Soybean is considered as one of the oldest crops in the world, with reports of its cultivation in China dating from around 2500 BC. In Brazil, it was introduced in 1882, but large scale cultivation began in the early 1960s. Brazilian soils were originally devoid of soybean bradyrhizobia. Strains were brought from the USA and most soils now contain a naturalized population of soybean bradyrhizobia, among which predominate three most competitive *B. elkanii* strains. Introduction of new strains that fix N2 more efficiently may thus be very difficult [111]. However Santos *et al.* [110] detected a great variability among *Bradyrhizobium* strains consecutive to their adaptation to the soil, and variants showing higher rates of N2 fixation and a better competitiveness than the parental genotypes were isolated. This suggested that it was possible to select variant strains that can contribute to an improved plant N nutrition status. *Sinorhizobium* spp. strains isolated from soybean nodules in Brazil can fix as much N2 as the *B. japonicum*/*B. elkanii* strains carried in Brazilian commercial inoculants, but they are less competitive [112].

Two superior strains PRF 81 and H 12 were identified on common bean (*Phaseolus vulgaris*) cultivated in low fertility soils, despite the high soil rhizobial population, and the high competitiveness of this population. The obtained yields with these two strains were 5-fold higher [113].

In Senegal Sy and N'doye (personal communication) conducted inoculation experiments using *Alysicarpus glumaceus*, *Alysicarpus ovalifolius* and *Tephrosia purpurea* to improve fallow soil fertility. Significant biomass production was observed after a six-month culture, but response to inoculation was variable. Compared to non-inoculated controls, dry matter production was about three times higher in the case of *Alysicarpus glumaceus* (5 tons/ha), and about two times higher for *Alysicarpus ovalifolius* (3 tons/ha). However, inoculation of *Tephrosia purpurea* did not entail any effect, suggesting the presence of competitive but inefficient indigeneous rhizobia. Soil analysis indicated an important increase in the nitrogen content.

In Republic of Guinea, inoculation of 9 plant species of the genera *Albizia*, *Milletia* and *Erythrophleum* had a highly significant effect. Inoculation tests on *Milletia rhodantha* and *Milletia zechiana* in nurseries using the different types of symbionts showed that tree growth can be enhanced by 50 to 100% depending on the strain [5, 114].

Horizontal transfer

There are clear observations for subsets of functions (involved in metabolism and others) spread among diverse organisms, and lateral gene

transfers play an important role in bacterial adaptation to environment. High degree of genome plasticity in LNBs is known for long [33] and this is confirmed by recent ecological, genetic and genomic studies. Several rhizobial population survey reports evidenced the presence of large numbers of non-symbiotic rhizobial strains in soils [115-117]. These strains can often become effective symbionts upon acquisition of symbiotic genes [115, 116, 118]. Non symbiotic rhizobia persist in soils in the absence of legume plants, and upon introduction of legumes, they may acquire symbiotic genes from inoculant strains [119].

Early studies already pointed out that differences in plasmid content may explain to a good extent the different behaviors of *Rhizobium* and *Agrobacterium* as symbionts or pathogens [34]. Symbiotic plasmid loss and gain is a continuous and dynamic process in rhizobia. The acquisition of genetic information for becoming a pathogen or a symbiont seems to be a very recent event for some lineages of rhizobia. Different *Rhizobium* species containing tumor inducing (Ti) plasmids from *Agrobacterium tumefaciens* induce tumors, although these tumors are smaller in size. On the other hand, *Agrobacterium tumefaciens* and *Phyllobacterium* sp. strains containing symbiotic plasmids from *Rhizobium* form nitrogen-fixing nodules on the corresponding host legume [32, 34].

Although rhizobia are polyphyletic, their symbiotic genes are rather conserved. Phylogenetic studies on nodulation genes show that they evolved independently compared to housekeeping genes, suggesting a different origin and a probable acquisition via lateral transfer. The *nod* A, B, C genes probably originate from outside, since their G+C content is significantly lower than the average G+C content of rhizobia [120]. *nod* and *nif* genes are generally located close to each other. Laguerre *et al.* [121] showed that *nod*C and *nif*H phylogenies are generally similar, suggesting their co-transfer in diverse organisms. However, cases of incongruence were also detected, suggesting that genetic rearrangements occurred in the time-course of evolution, and that lateral genetic transfer across rhizobial species plays a role in diversification and structuration of the natural rhizobial populations.

Phylogenetic studies suggest that rhizobia diverged well before the existence of legumes and probably before the appearance of angiosperms [122]. Indeed fast-growing rhizobia diverged around 200 to 300 MYA, whereas divergence between fast-growing rhizobia and slow-growing rhizobia occurred around 500 MYA. These times are earlier than the split between monocots and dicots (156-171 MYA) and the separation of brassicas and legumes (125-136 MYA). Therefore nodulation capacity is thought to have been acquired after bacterial divergence and horizontal spread among different genera. This hypothesis is strongly supported by the congruence of phylogenetic trees of bacterial nodulation gene *nod* A and rbc L plant gene,

and by the observed similarity between *nod* genes from α and β rhizobia, despite their taxonomic distance. Most probably "β-rhizobia" evolved from nitrogen-fixing β-proteobacteria through multiple lateral *nod* gene transfers [59].

Adaptation of rhizobia to legumes during evolution and nodulation specificity involves 1) recruitment of genes (by transfer or duplication as a set of grouped genes on plasmid, symbiotic islands) and 2) subsequent evolution of these genes (i.e. allelic variation modulating the gene function). In *M. loti* chromosome, a 500-kb region, called SYM island, integrated inside a t-RNA phenylalanine gene, carries all necessary genetic information for nodulation, nitrogen fixation, and transfer [118, 119]. *Bradyrhizobium* genome harbours a total of 167 genes coding for transposases (including104 components of insertion sequences), indicating a potentially high rate of genetic recombination [123]. Phylogenetic analyses of symbiotic nodulation genes (*nod*A, *nod*Z and *noe*I), housekeeping genes, 16S rRNA and *dna*K genes support vertical and lateral gene co-transfer within the *Bradyrhizobium* genus [62].

One hypothesis to explain nodulation acquisition is that nodulation genes were recruited from fungi. Most fungi produce chitin as part of their cell wall and therefore have chitin synthases, which are similar to *nod* C. Interestingly, one of the endomycorrhizal fungi which can infect plant roots using a pathway that shares common steps with nodulation, was found to contain a *Burkholderia* strain that harbored nitrogen fixation genes [124].

Overview of tropical Legume Nodule Bacteria studies around the world

The first rhizobial isolations on tropical trees were reported one century ago, but major advances were achieved by Allen and Allen [125, 126] who isolated large collections of rhizobial strains, all slow-growing, from 72 species of bush and tree species in Hawaï. Strains were classified in cross-inoculation groups based on infectiveness and effectiveness. Until 1964, trees were thought to be nodulated by slow-growing rhizobia only [127]. Later, Trinick [128-130] reported that fast-growing strains could also nodulate some trees of the genera *Leucaena, Mimosa, Sesbania and Acacia*. Since then, many studies around the world have shown that tropical tree legumes host a large rhizobial diversity ([5-7, 99, 131-144] and others...).

Africa

In Senegal, extensive studies were performed on LNB isolated from many wild legume plants, trees (*Faidherbia albida, Acacia* spp., *Pterocarpus* spp), water-logged herbs (*Sesbania spp., Aeschynomene spp., Neptunia natans*), dry land annual and perenial species naturally growing in different pedoclimatic regions and some cultivated crops (*Phaseolus, Vigna, Arachis*...). A large

diversity of LNB was described among these plants, which were also found nodulated by several different bacterial species (Table 2). Some were found to belong to described species (*R. tropici, R. etli, B. elkanii*), but the majority of them represented new species in *Sinorhizobium, Mesorhizobium, Allorhizobium, Azorhizobium, Bradyrhizobium*, and also surprisingly *Methylobacterium* and *Burkholderia.*

Dreyfus and Dommergues [145] isolated fast-growing strains nodulating *Acacia* spp. and distinguished three host-specificity groups among *Acacia* species, that were nodulated by *Bradyrhizobium* strains, *Rhizobium* strains or both. Dreyfus *et al.* [146] then discovered *Azorhizobium caulinodans* nodulating stems and roots of *Sesbania rostrata.* This questioned the dogma according to which tropical legumes were nodulated by broad host range slow-growing rhizobia. Polyphasic taxonomical characterisation of 80 fast-growing *"Rhizobium"* strains isolated from different *Acacia* spp. (mainly *A. senegal* and *A. tortilis* subsp. raddiana) and from several *Sesbania* spp. (mainly *S. rostrata, S. grandiflora, S. pubescens*) led to the revision of *Sinorhizobium* [147] and the description of *Sinorhizobium terangae* and *Sinorhizobium saheli* [99, 148]. Nodulation tests performed on different *Sesbania* spp. (*S. rostrata, S. grandiflora, S. pubescens*), *Acacia* spp. (*A. raddiana, A. senegal, A. seyal*), *Leucaena leucocephala* and *Neptunia natans* showed that the host spectrum is not a distinctive feature of the species. Two biovars « acaciae » and « sesbaniae » were defined for each of the two species *S. terangae* and *S. saheli* [108]. *S. terangae* biovar sesbaniae and *S. saheli* biovar sesbaniae nodulate both roots and stems of *S. rostrata.* Nod factors of Sinorhizobium spp. biovar sesbaniae strains are identical to those of *A. caulinodans* [106, 149]. In addition to the latter, [142] brought indications for the presence of *Sinorhizobium arboris* nodulating *Acacia* spp. in Senegal.

Mesorhizobium plurifarium [139] was proposed for a group of tree strains in Africa (mainly from *Acacia* spp. in Senegal and Sudan) and Brazil. *M. plurifarium* is phenotypically and genotypically separate from known species, but with a certain degree of internal heterogeneity. It is however phylogenetically (16S rDNA) homogeneous and groups in *Mesorhizobium,* in the vicinity of *M. huakuii.*

Rhizobial strains associated to *Acacia tortilis* subsp. raddiana were studied in different regions, biotopes and depth in soil in Senegal, Tunisia, and Mauritania. They were found taxonomically diverse among *Mesorhizobium* and *Sinorhizobium* but with homogenous symbiotic characteristics [150].

Faidherbia albida was found to be nodulated by a great variety of strains belonging to several *Bradyrhizobium* genospecies [151] present down to the water table at -32 m depth [138].

Two other tree species, *Pterocarpus lucens* and *Pterocarpus erineaceus*, both having economical and forest importances in Senegal, are nodulated by strains belonging to *Mesorhizobium (M. plurifarium), Rhizobium* (several groups),

Bradyrhizobium elkanii, *B. japonicum* and *Bradyrhizobium* spp. genospecies IV & VII [143, 144].

Strains isolated from *Phaseolus vulgaris* cultivated in Senegal were identified as *R. etli* and *R. tropici* b [152].

Doignon-Bourcier *et al.* [153, 154] isolated and characterised 71 slow-growing bacterial strains from nodules of 27 native leguminous plant species in Senegal (West-Africa). These strains were further identified as representing several *Bradyrhizobium* genospecies [151]. These plants belonged to the genera *Abrus, Alysicarpus, Bryaspis, Chamaecrista, Cassia, Crotalaria, Desmodium, Eriosema, Indigofera, Moghania, Rhynchosia, Sesbania, Tephrosia,* and *Zornia* which play an ecological role with an agronomic potential for arid regions.

Symbionts nodulating *Neptunia natans* in Senegal belong to *Sinorhizobium*, *Mesorhizobium* and *Allorhizobium undicola*, a species phylogenetically related to *Agrobacterium vitis* [155] and unpublished).

In Mali, integration of the legume *Dolichos lablab* c.v. Highworth in the production system is a true solution for a sustainable soil fertility and an increased fodder production to compensate the shortening of the fallow period without manuring. Yattara [156] studied 32 nodule isolates from *Dolichos lablab* in Mali and in Senegal. Their nitrogen-fixing potential and effect upon plant inoculation depended on the strains. Its symbionts were found taxonomically diverse and corresponded to several species of the genera *Rhizobium, Sinorhizobium, Bradyrhizobium, Agrobacterium*, and *Burkholderia*. Six of the latter have a 16S rDNA similar to *Burkholderia fungorum*.

In Sudan, Zhang *et al.* [132] described the phenotypic diversity of nodule isolates from *Acacia senegal* and *Prosopis* chilensis. This diversity was further confirmed by molecular techniques [132, 157] and two new species were proposed in *Sinorhizobium, S. arboris* and *S. kostiense* [142].

In Ethiopia Wolde-Meskel *et al.* [158-161] reported that 75% of the strains nodulating woody legumes form several groups not related to reference species. In particular, this is the first report on the characterization of the symbionts of *Milletia ferrugenea, Acacia abyssinica* and *Albizia gummifera*.

In South Africa most of the LNB corresponding to 19 endemic *Cyclopia* species are alphaproteobacteria (*Bradyrhizobium* spp., *R. tropici*) but 7 strains are *Burkholderia tuberum* and *Burkholderia* spp. [162]. As already mentioned, *Methylobacterium* LNB strains were identified as symbionts of *Lotononis bainesii* [50], and of other species in *Lotononis, Indigofera, Calpurnia* genera [51].

Diabate *et al.* [5] studied the occurence of nodulation in 156 leguminous species - mostly mature trees - growing in six natural rain forest areas in Guinea.

Madagascar

In Madagascar, *Dalbergia* (Rosewood) is nodulated by *Mesorhizobium, Rhizobium, Bradyrhizobium, Azorhizobium, Ralstonia* and *Burkholderia*.

Numerous *Azorhizobium* sp. strains were isolated from nodules, but their nodulation ability was not demonstrated (Rasolomampianina *et al.*, in preparation).

Canary islands

Vinuesa *et al.* [163] characterised nodulating bacteria associated with several endemic legume species (*Adenocarpus* spp., *Chamaecytisus proliferus, Lupinus* spp., *Spartocytisus supranubius, Teline* spp.) from the Canary Islands and described several *Bradyrhizobium* new genospecies, among which one was proposed as *Bradyrhizobium canariense* sp. nov.

Central and South America

A number of 1294 different Leguminosae species have been reported in Amazon. From the most intact to the most disturbed forests frequencies of Leguminosae in relation to the total number of species are usually high [9]. A large number of strains have been isolated from several forest species in Brazil. Considering collections made by CNPAB-EMBRAPA (Rio de Janeiro), INPA (Amazonas), and UFLA (Minas Gerais), there are about 4000 strains isolated from species belonging to around 60 native genera of Leguminosae, i.e. half of the estimated number of the native nodulating genera, and from around 10 exotic genera. Slow-growing, *Bradyrhizobium*-like strains were isolated from 89% of Caesalpinioideae genera and from 70 to 76% of genera in Mimosoideae and Papilionoideae, respectively [136, 137]. Strains with intermediate growth were also isolated from forest species, but with a much lower frequency in native genera (11% Caesalpinioideae, 30% Mimosoideae and 15% Papilionoideae). Fast-growers were found in symbiosis with 44% Caesalpinioideae, 67% Mimosoideae and 39% Papilionoideae [43]. Several authors [131], [6, 7, 135-137] investigated the nodulated Amazonian legume tree species and evidenced the diversity of the associated rhizobia of almost 200 tree and lianas species. 16S rDNA sequencing of 44 strains isolated from 29 legume tree species that represent 13 tribes including all three subfamilies of the Leguminosae showed their relationships with *Rhizobium, Sinorhizobium, Mesorhizobium, Bradyrhizobium* and *Azorhizobium* [137].

In Puerto Rico, Zurdo-Pineiro *et al.* [164] identified the fast-growing rhizobia nodulating tropical legumes *Sesbania, Caliandra, Poitea, Piptadenia, Neptunia* and *Mimosa* as *R. gallicum* and *R. tropici.*

In Costa Rica, Parker [165] studied the rRNA and dnaK relationships of the *Bradyrhizobium* sp. nodule bacteria from four Papilionoid legume trees *Andira inermis, Dalbergia retusa, Platymiscium pinnatum, Lonchocarpus atropurpureus,* and showed lateral transfer bertween *B. japonicum B. elkanii* and other *Bradyrhizobium* sp. strains.

In the Carrabean region, Muller *et al.* [166] observed that the genetic diversity of the populations of *Pterocarpus officinalis* (Jacq.) - *Bradyrhizobium* spp. are lower in islands compared to continent.

In Barro Colorado Island, Panama, Parker [167] characterized 96 *Bradyrhizobium* isolates from 10 legume host species in six genera (*Centrosema, Desmodium, Dioclea, Inga, Machaerium* and *Vigna*). They evidenced significant mosaic structure across the rRNA region, indicating that lateral gene transfer events have played a role in the evolution of symbiotic bacteria in this environment.

Asia

Great diversity has been detected among the rhizobia isolated from legumes in China. More than twenty rhizobial species within the genera *Rhzobium, Sinorhizobium, Mesorhizobium, Bradyrhizobium, Burkholderia* and *Ralstonia* have been described of recorded among LNBs in China. However most of these rhizobial species, except *R. hainanense*, *Burkholderia* and *Ralstonia*, were found in the temperate zones of China (see Table 2). In the tropical region of Hainan province of China Gao *et al.* [102] described 63 isolates corresponding to 21 tropical legume species, and evidenced 3 slow-growing and 3 fast-growing groups among them. *Sinorhizobium fredii* and *Bradyrhizobium* sp. are predominant LNBs associated to soybean [168] and to common bean plants [169] in the subtropical zones of China.

Acacia mangium is an important tree legume in tropical and subtropical developing countries. *A. mangium* is used as a pioneer plant for reforestation and has many uses such as fuel wood, timber, wind protection and animal fodder. *A. mangium* is broadly used in industrial plantations for pulp production because of its good silvicultural ability on degraded soils due to symbiotic nitrogen fixation [170]. In Indonesia *A. mangium* is nodulated by *B. elkanii* strains [171]. In the Philippines and in Thailand, it is nodulated by *Bradyrhizobium, Rhizobium* and *Ochrobactrum* [28]. *Ochrobactrum* sp. strain also nodulates *A. albida* and *Paraserianthes falcataria.*

By phylogenetic analysis of 128 strains from 23 legume hosts in Korea, Kwon *et al.* [172] identified several new groups in *Bradyrhizobium, Mesorhizobium, Rhizobium* and *Sinorhizobium.*

Transcontinental studies

Munive [114] described the genetic diversity of 119 nodule isolates from 27 tree species in the tropical humid forests of Guyana, Guinea and Madagascar, representing 9 tribes in the 3 subfamilies of Leguminosae (18 genera in Papillonaceae, 9 genera in Mimosaceae, 2 genera in Caesalpiniaceae). This is the first description of the nodulation of 20 species and 4 genera, *Chidlowia, Samanea, Calopogonium* and *Chadsia*. Partial sequencing of the

Table 1. Classification of the Nitrogen-fixing Bacteria Symbionts of Legumes.

Species	Host plants	References
Class Alpha Proteobacteria		
Order *Rhizobiales*		
Family *Rhizobiaceae*		
Genus *Rhizobium*		
R. leguminosarum		[11, 193]
biovar viciae	*Pisum sativum, Vicia, Lathyrus, Lens*	[11, 193]
biovar trifolii	*Trifolium pratense, Trifolium spp.*	[11, 193]
biovar phaseoli	*Phaseolus vulgaris L., P. angustifolius, P. multiflorus*	[11, 193]
R. tropici		
Type II A	*P. vulgaris L., Leucaena, Amorpha fruticosa*	[194]
Type II B	*P. vulgaris L., Leucaena*	[194]
R. elti biovar phaseoli	*Phaseolus vulgaris, Leucaena*	[195, 196]
R. etli biovar mimosae	*Mimosa affinis*	[197]
R. hainanense	*Desmodium sinuatum, Desmodium gyroides, Desmodium triquetrum, Desmodium heterophyllum, Acacia sinicus, Arachis hypogaea, Centrosema pubescens, Macroptilium lathyroides, Stylosanthes guianensis, Tephrosia candida, Uraria crinita, Zornia diphylla*	[198]
R. gallicum		[199]
biovar gallicum	*Phaseolus vulgaris L., Leucaena leucocephala, Macroptilium atropurpureum, Onobrychis viciifolia*	[199]
biovar phaseoli	*Phaseolus vulgaris L.*	[199]
R. mongolense	*Medicago ruthenica, Phaseolus vulgaris*	[200]
R. galegae		[201]
biovar orientalis	*Galega orientalis*	[202]
biovar officinalis	*Galega officinalis*	[202]
	Astragalus cruciatus, Argyrolobium uniflorum, Anthyllis henoniana, Lotus creticus, Medicago spp.	[22]
R. giardinii		[199]
biovar giardinii	*Phaseolus vulgaris L., Leucaena leucocephala, Macroptilium atropurpureum*	[199]
biovar phaseoli	*Phaseolus vulgaris L.*	[199]
R. huautlense	*Sesbania herbacea*	[203]
R. indigoferae	*Indigofera amblyantha, I. carlesii, I. potanini*	[204]
R. sullae	*Hedysarum coronarium*	[205]
R. loessense	*Astragalus, Lespedeza*	[206]
R. yanglingense	*Coronilla varia, Amphicarpaea trisperma, Gueldenstaedtia multiflora*	[207]
Genus *Sinorhizobium*		[99, 147]
S. meliloti	*Medigaco, Melilotus, Trigonella*	[11, 99, 208]
biovar acaciae		[150]
S. fredii		[99, 209]
chemovar fredii	*Glycine max*	[209]
chemovar siensis	*Glycine max*	[209]

Table 1. Continued

S. sahelense	*Sesbania* spp.	[99]
biovar acaciae	*Acacia* spp.	[107]
biovar sesbaniae	*Sesbania* spp.	[107]
S. terangae		[99, 148]
biovar acaciae	*Acacia* spp.	[106]
biovar sesbaniae	*Sesbania* spp.	[106]
S. medicae	*Medicago* spp.	[210]
S. kostiense	*Acacia, Prosopis*	[211]
S. morelense	*Leucaena leucocephala*	[41]
S. americanum	*Acacia*	[212]
S. arboris	*Acacia, Prosopis*	[211]
S. kummerowiae	*Kummerowia stipulacea*	[204]
Ensifer adhaerens	*Sesbania, Medicago*	[38]
(Sinorhizobium adhaerens)	*Sesbania grandiflora, Leucaena leucocephala, Pithecellobium dulce, Medicagosativa*	[213]
Genus *Allorhizobium*		[139]
A. undicola	*Neptunia natans, Acacia senegal, A. seyal, A; tortilis, Lotus arabicus, Faidherbia albida*	[139]
Family *Phyllobacteriaceae*		
Genus *Mesorhizobium*		
M. loti	*Lotus corniculatus, L. tenuis, L. japonicum, L. krylovii, L. filicalius, L. schoelleri, Anthyllis spp., Lupinus spp.*	[214]
M. huakuii	*Astragalus sinicus, Acacia* spp.	[215]
M. ciceri	*Cicer arietinum*	[100]
M. tianshanense	*Glycyrrhiza pallidiflora, G. uralensis, Sophora alopecuroides, Glycine max, Swainsonia salsula, Halimodendron holodendron, Caragana polourensis*	[103]
M. mediterraneum	*Cicer arietinum*	[216]
M. plurifarium	*Acacia senegal, A. seyal, A. tortilis, Leucaena leucocephala, L. diversifolia, Prosopisjuliflora, Chamaecrista ensiformis*	[139]
M. amorphae	*Amorpha fruticosa*	[197]
M. chacoense	*Prosopis alba*	[217]
Genus *Phyllobacterium*		
Phyllobacterium sp.		[95]
Family *Methylobacteriaceae*		
Genus *Methylobacterium*	*Crotalaria podocarpa, C. Perottetti,*	[24, 47]
M. nodulans	*C. glaucoides.*	
Family Brucellaceae		
Genus *Ochrobactrum*	*A. mangium, Faidherbia albida, Paraserianthes falcataria*	[28]
Family Hyphomicrobiaceae	*Neptunia natans*	[25, 44]
Genus *Devosia*		
"Devosia neptuniae"		

Table 1. Continued

Family *Hyphomicrobiaceae*		
Genus *Azorhizobium*	*Sesbania rostrata*	[146]
A. caulinodans	*Sesbania rostrata*	[146]
A. johannense	*Sesbania virgata*	[9]
Azorhizobium sp.	*Sesbania rostrata*	[218]
Family *Bradyrhizobiaceae*		[10]
Genus *Bradyrhizobium*		
B. japonicum	*Glycine max, Glycine soja, Macroptilium atropurpureum*	[11, 219]
B. elkanii	*Glycine max, Vigna spp., Macroptilium atropurpureum*	[220]
B. liaoningense	*Glycine max, Glycine soja*	[101]
B. yuanmingense	*Lespedeza cuneata.*	[221]
B. betae	*Beta vulgaris*	[42]
B. canariense	*Genisteae & Loteae* plants	[192]
Bradyrhizobium sp.	*Vigna, Lupinus, Mimosa*	[10]
	Faidherbia, Acacia, 27 herb legumes	[66, 138, 151]
	Aeschynomene	[61, 222]
Genus *Blastobacter*		
B. denitrificans	*Aeschynomene indica*	[26]
Class Beta Proteobacteria		
Order *Burkholderiales*		
Family *Burkholderiaceae*		
Genus *Burkholderia*		
B. caribensis	*Mimosa pudica, M. diplotricha.*	[55]
B. cepacia genomovar VI (*B. dolosa*)	*Alysicarpus glumaceus*	[56]
B. tuberum	*Aspalatus carnosa*	[55]
B. phymatum	*Machaerium lunatum*	[55]
Genus *Ralstonia*		
R. taiwanensis	*Mimosa pudica, M. diplotricha.*	[30]
Class Gamma-Proteobacteria		
Order *Enterobacteriales*		
Pantoea agglomerans	*Hedysarum carnosum, H. spinosissimum subsp. capitatum, H. pallidum*	[31]
Enterobacter kobei	*Hedysarum carnosum, H. spinosissimum subsp. capitatum, H. pallidum*	[31]
Enterobacter cloacae	*Hedysarum carnosum, H. spinosissimum subsp. capitatum, H. pallidum*	[31]
Leclercia adecarboxylata	*Hedysarum carnosum, H. spinosissimum subsp. capitatum, H. pallidum*	[31]
Escherichia vulneris	*Hedysarum carnosum, H. spinosissimum subsp. capitatum, H. pallidum*	[31]
Order *Pseudomonadales*		
Pseudomonas sp.	*Hedysarum carnosum, H. spinosissimum subsp. capitatum, H. pallidum*	[31]

16S-23S rDNA intergenic spacer revealed that 90 new slow-growing isolates represent several species in *Bradyrhizobium* spp., and 29 fast-growing isolates belong to *Rhizobium* (Guyana and Madagascar), *Mesorhizobium, Sinorhizobium, Azorhizobium* (Madagascar) and *Burkholderia* (Guinea).

Calliandra calothyrsus, Gliricidia sepium, Leucaena leucocephala and *Sesbania sesban* are among the three most important fast-growing legume tree species used in tropical agroforestry. Their success can be explained by their ability to nodulate with a wide range of rhizobia that are indigenous to soils grographically spread across tropical regions all over the continents. Indeed, Bala *et al.* [173] characterised the natural populations of their corresponding nodulating bacteria in soils from nine sites across tropical areas of three continents. They found them distributed among *Rhizobium, Mesorhizobium, Sinorhizobium* and *Agrobacterium.* Specificity for nodulation and N_2 fixation greatly varied depending on the association of both partners. Symbionts of all four legumes exhibited a wide range of promiscuity and symbiotic effectiveness with isolates of *S. sesban* having the narrowest host range. *Calliandra calothyrsus, Gliricidia sepium, Leucaena leucocephala* rhizobial isolates were able to effectively cross nodulate each others' hosts as well as a number of other species [174]. In another study by Räsänen *et al.* [175], half of the 250 nodule isolates from *Calliandra calothyrsus* in its native regions of Central America (Mexico, Guatemala, Honduras, Nicaragua and Costa Rica) and in areas in which the tree has been introduced (Cameroon, Kenya and New Caledonia) were *R. tropici.* Other isolates were *S. meliloti, Agrobacterium* sp. *S. saheli, R. etli*; two isolates were *Steno-trophomonas* sp. and *Enterobacter* sp. *R. tropici* was predominant everywhere except in Cameroon where it is *S. meliloti. nod*A genes of *Calliandra* strains were found different from those of *Mesorhizobium, Sinorhizobium* and *Rhizobium* reference strains [175].

McInroy *et al.* [176] characterized rhizobia from different African acacias and other tropical woody legumes from Kenya, Zimbabwe, Sudan and Honduras using Biolog and partial 16S rRNA sequencing.

Genomics of LNB

Martinez-Romero *et al.* [33] reviewed early advances in *Rhizobium* genomics. Bacterial complete genome sequences are now being published almost every week. By the time of the writing of this paper, the complete genome structures of 160 bacteria have been determined and genome sequencing of several hundred other species is in progress [177].

The first complete sequence of a broad host-range symbiotic plasmid (pNGR234a) of *Sinorhizobium* sp. strain NGR 234 was reported by Freiberg *et al.* [178]. Since then, the nucleotide sequences of entire genomes of several LNBs, *Mesorhizobium loti, Sinorhizobium meliloti, Bradyrhizobium japonicum* and *Rhizobium etli* have been released [123, 179-181]. Based on phylogenetic

analysis of four chromosomal loci, Turner *et al.* [182] assigned the sequenced *M. loti* strain to *M. huakuii* biovar *loti*. Pathogenic *Agrobacterium, Ralstonia* and *Burkholderia* strains, photosynthetic *Bradyrhizobium* strains and representatives of genera and species intertwinned with LNB have also been sequenced, like *Brucella* and *Bartonella* for instance.

Sequenced prokaryotic genomes currently range in size from 490885 bp to 9 105 828 bp for *Bradyrhizobium japonicum* genome, which is the largest sequenced bacterial genome until now [123], meaning a 20-fold size difference between microbial genomes [177]. In particular, within the alpha proteobacteria, variations in genome size and structure are tremendous. The genome (1 to almost 10 megabases) often consists of multiple circular or linear replicons, but little is known about the underlying mechanisms that cause this diversity. There is a possible correlation between the lifestyle of the organism and its genome content. Plant-associated bacteria such as *Mesorhizobium loti* have undergone extreme genome expansion (up to a few thousand genes), whereas the shift to intracellular environment and vector-mediated transmission have resulted in extreme genome reduction (for example *Buchnera* endosymbionts of aphids or *Bartonella* species transmitted by fleas). *M. loti* genome consists of one chromosome (7.0 Mb) and two plasmids (0.35 Mbp and 0.2 Mbp). *R. etli* genome (6,53 Mb) is divided into one circular chromosome and six large plasmids [181]. *S. meliloti* genome is 6.7 Mb large and essentially composed of 3 replicons (6200 genes), one chromosome and two megaplasmids, pSYMA and pSYMB. Unlike the chromosome, they carry the replication and stability genes (rep ABC) that are common to most rhizobial plasmids. The chromosome (3.65 Mb) show little reiteration, and carries 3 regions showing a lower GC content. Fifty-nine per cent of the genes are of known functions, housekeeping ones, amino acid or peptide transporters, degradation genes, ose metabolism, nucleotide cyclases, homologs to animal and plant virulence determinants. pSYM A (1.35 Mb) has a lower GC content and was acquired more recently; it carries genes involved in symbiosis, N & C metabolism, transport, stress, resistance but no essential gene. In contrast pSYM B (1.68 Mb), like the chromosome, carries essential genes. It confers competence for saprophytic life: catabolism of small components from soil or rhizosphere, N metabolism in different forms, uptake of solutes of the ATP binding family (20%), biosynthesis of polysaccharides and exopolysaccharides (14 %). 20-40% of the genes have no known function and a number of them are unique (no homolog in other organisms).

Complete genome sequences can now be compared for a growing number of rhizobia and related alpha proteobacteria. There is a basic core of 600 genes that are found in all these genomes. They are mostly located on the main chromosome, mostly high in G+C content, and mostly support the same phylogenetic tree. In contrast, several other genes are shared among several

species but they do not support this consensus phylogenetic tree, indicating that they probably originate from horizontal transfer. Some of these are on the chromosome, but many are on plasmids, like nodulation and fixation genes. Besides the chromosome, several species have a very large second replicon, like pSymB in *S. meliloti*, linear chromosome in *Agrobacterium tumefaciens*, chromosome II in *Brucella melitensis*, with a G+C content similar to the chromosome, indicating they are long-term residents of the species. pSymB bears genes that are not shared across species or whose phylogeny suggests a history of horizontal transfer. For instance phylogeny of nif genes indicates an ancient horizontal transfer. The *Brucella* genome share common gene and mechanisms with *S. meliloti* among which *bac* a, transmembrane transport protein necessary for symbiosis. Similar to LNBs, *Brucella* bacteria live intracellularly, surrounded by endocytosis.

Extensive orthology and nucleotide colinearity between the genomes of the symbiont *Sinorhizobium meliloti* and the pathogen *Agrobacterium tumefaciens* suggest a recent evolutionary divergence [183]. Their similarities include metabolic, transport, and regulatory systems that promote survival in the highly competitive rhizosphere.

S. meliloti and *M. loti* genomes show numerous differences regarding their content and organization. Thirty five per cent of *M. loti* genes have no ortholog in *S. meliloti*, in spite of their symbiotic and taxonomic relatedness. However, both genomes are large (6.7 and 7.6 Mb respectively) and many genes known to be required for the symbiosis are grouped in clusters. In *M. loti*, symbiotic genes are located on a chromosomal symbiosis island of 611 kb, whereas in *S. meliloti* most of them are located on either of two large plasmids pSym A or pSymB. This location of symbiosis genes on islands or plasmids supports the idea that these regions have the potential to be horizontally transferred. Nodulation capacity transfer was observed as a very efficient mechanism in field experiments [118]. Excision and integration of the symbiosis island out of/in the chromosome occurs at a t-RNA phenylalanine gene [184]. Over 50% of the genes on the 536 kb symbiosis island of *Sinorhizobium* sp. NGR 234 strain have no ortholog in *S. meliloti*. In *B. japonicum*, a symbiosis island with low GC content (59.4%) was identified. It includes 60% of the genomic transposase genes and is flanked by a val-tRNA gene at one end. In addition, 34% of the 8317 potential protein-encoding genes in *B. japonicum* match genes in *S. meliloti* and *M. loti*, 632 of which are genes of unknown function [123]. Nine and seven per cent of *B. japonicum* genes are commonly found in either *M. loti* or *S. meliloti* genomes, respectively. Fifty per cent are unique to *B. japonicum*. This indicates that significant portions of the gene components in the genomes are unique to the species [123].

Complete genomic sequences provide not only the information needed to perform functional analysis of the genes but also, by comparative genomics,

new insights into gene function, gene evolution, and genome evolution. As we learn more about genomes, we need to reassess the criteria by which we partition the bacterial world into genera and species. Do bacterial species correspond to natural biological units or to human attempts to partition a continuum? Bacterial genomes are composite. The basic genome is composed of housekeeping genes that are needed under almost all growth conditions. They are carried on the chromosome, their organisation is reasonably stable, and they are inherited vertically from mother to daughter cell. In contrast, the accessory genome consists of selfish genes offering adaptations to special circumstances; they are carried on plasmids, islands, transposons and phages, they undergo frequent rearrangements and are often transferred horizontally between cells [185]. The recent development of complete genome sequencing allows construction of phylogenies for hundreds of genes, most of which agree with the SSU rRNA phylogeny. Accessory genome comprises genes that are intermitently advantageous and not uniformly distributed among individuals of a species, though they may be shared between species. The accessory genes encode important phenotypes often of economic interest to medicine (pathogenicity islands, antibiotic resistance) and agriculture. In rhizobia, the basic genome is high in G+C %, different from accessory genome which confers ecological specialisations, varies in content, and is subjected to horizontal transfers via plasmids or genomic islands. In *Mesorhizobium*, except for *glnII* which is transferred from *Rhizobium* to *Mesorhizobium*, the basic housekeeping genes generally share a consistent phylogeny [182, 186]. In contrast, the accessory genome, including symbiotic genes, undergoes detectable transfers within and between species. Accessory DNA makes up to 10-25% of the DNA in the four sequenced rhizobial genomes *M. loti*, *S. meliloti* and *Agrobacterium tumefaciens*. Comparison of whole sequences of more than one strain per species will reveal the common core of genes which defines the species. It will also identify the associated accessory gene pool that provides a species with its adaptations to different niches and determines the variety of key properties such as symbioses or diseases with which the species is associated [182]. Several other LNBs are in the process of being sequenced. This will bring a considerable amount of data and will undoubtedly help understanding the molecular basis of how a given bacterium becomes a pathogen or a symbiont. Whole genome assays will constitute a revolution in bacteriology in general, and in taxonomy in particular. New tools for taxonomical studies are already developing from genomics, like Multilocus Sequence Typing (MLST) and microarrays enabling to test each strain for a multitude of criteria in a single experiment. For instance the european-funded BACDIVERS project (http://lmg.ugent.be/bacdivers) aims to design arrays dedicated to either identification of symbiotic specificity or environment adaptation (salinity, drought). Plant tests are limiting because seeds,

germination and plant growth conditions for many legume species are not available. Microarrays dedicated to symbiotic functions will help to overcome this bottleneck.

Future challenges in rhizobium taxonomy

The recent findings on legume symbiotic bacteria open multiple perspectives for studying their biology, mechanisms, evolution, prediction of possible properties, and also potential applications. Discovery of symbioses involving new rhizobia (*Azorhizobium caulinodans*, photosynthetic *Bradyrhizobium*, *Methylobacterium*, *Burkholderia*, *Ralstonia*, *Cupriavidus*) has brought new models to progress in fondamental comprehension of intimate functioning of these interactions, pre-requisite of applications. They suggest that much remains to be discovered concerning the diversity of rhizobia, since only a small proportion of legumes (less than 20%) has been examined for symbiosis so far, especially in tropical areas. A renewed vision and new tools must be developed to detect and identify the new types of rhizobia.

The future challenges for taxonomy are to include more genes in characterization studies, to develop bioinformatic tools for integration of the different kinds of data for exponential increasing number of strains, and to coordinate the databases of different research groups in the world. The concept of bacterial species should evolve and include new knowledge coming from whole genome sequencing programs (genes, genome structure and organisation...) and from bacterial population genetics [19]. Cohan [187] proposed a clonal definition of the species, based on an ecological niche. This definition argues that during the past half-century of bacterial systematics, species demarcation has not been guided by a theory-based concept of species. This author proposes to apply to bacteria the universal concept of species developed for eukaryote systematics, consisting of a group of organisms whose divergence is capped by a force of cohesion and irreversible. Here different species are ecologically distinct. In the case of bacteria, these universal populations are held not by species but by ecotypes. Dykhuisen and Green [188], proposed to define species limits by applying a biological species concept, where a species is defined as a group of strains that share a common pool of genes (by lateral transfer and recombination). This can be determined by comparison of the sequences of multiple loci in a group of strains thought to belong to the same species. Young (personnal communication) proposes to refer to basic genome (housekeeping genes) for genus and species delineation and to consider accessory genes to define biovars. The SSU phylogeny had a major influence on our current perception of evolutionary relationships among bacteria and rhizobia in particular [189], but other genes are now currently used to confirm the relationships among them. In *Agrobacterium* and *Rhizobium*, *atp*D and *rec*A phylogenies often but not always agree with 16S

rDNA-based one, [186]. Multilocus sequence typing (MLST) is based on the well-tested principles of MLEE but assigns the alleles at each locus directly by nucleotide sequencing, which reveals all of the variation at a locus. In MLST 450-500 bp internal fragments of 5-7 housekeeping genes are determined for each isolate. Coenye and Vandamme [190] observed correlation between the results obtained with various approaches including the comparative sequences of 16S rDNA and nine housekeeping genes, the fraction of shared putative orthologous protein-encoding genes, conservation of gene order, dinucleotide relative abundance and codon usage among 11 genomes of species of lactic acid bacteria.Vinuesa *et al.* [191] proposed to use population genetic structure to define species limits in the *Rhizobiaceae*. [192] used a combination of population genetics and several phylogenetic inference methods to delineate *Bradyrhizobium* species.

New data from genomics will enable great advances in our understanding of all mechanisms involved in the success of establishment and functioning of nitrogen fixing symbioses in the diversity of environmental conditions where they exist. The increasing amount of complete bacterial genome sequences opens the possibility to compare whole genomes in their compositions and structures, and to design new tools to test hundreds or thousands of genes for special functions at a time [19]. Comparative genomics of the diverse LNBs with their non LNB phylogenetic neighbours (different species of the same genera like *Phyllobacterium, Burkholderia, Ralstonia* or of different genera like *Brucella*) may lead to the minimal genetic information for a bacterium to switch to different lifestyles, soil heterotroph, plant symbiont or pathogen... Microarrays will enable both phenotypic and genotypic characterizations instead of neutral information (non functional) generated by molecular fingerprints.

References

1. Schleifer, K.H. 2004, Syst. Appl. Microbiol., 27, 3.
2. Ranjard, L., Brothier, E., and Nazaret, S. 2000, Appl. Environ. Microbiol., 66, 5334.
3. Torsvik, V., Goksoyr, J., and Daae, F.L. 1990, Appl. Environ. Microbiol., 56, 782.
4. Polhill, R.M., and Raven, P.H. 1981, U.K.: Royal Botanic Gardens, Kew.
5. Diabate, M., Munive, A., de Faria, S.M., Ba, A., Dreyfus, B., and Galiana, A. 2005, New Phytol, 166, 231.
6. de Faria, S.M., Franco, A.A., de Jesus, R.M., de Manandro, M., Baitello, J.B., Mucci, E.S.F., Döbereiner, J., and Sprent, J.I. 1984, New Phytol., 98, 317.
7. de Faria, S.M., de Lima, H.C., Franco, A.A., Mucci, E.S.F., and Sprent, J.I. 1987, Plant Soil, 99, 347.
8. de Faria, S.M., Lewis, G.P., Sprent, J.I., and Sutherland, J.M. 1989, New Phytol., 111, 607.
9. Moreira, F.M.S. 1999, Current Plant Science and Biotechnology in Agriculture, M.H. F.O. Pedrosa, M.G. Yates and W.E. Newton (Eds.), Kluwer Acad. Publ., 181.

10. Jordan, D.C. 1982, Int. J. Syst. Bacteriol., 32, 136.
11. Jordan, D.C. 1984, Bergey's Manual of Systematic Bacteriology 321AL., a.H.J.C. N.R. Krieg (Ed.), The Williams and Wilkins Co., Baltimore, 234.
12. Graham, P.H., Sadowsky, M.J., Keyser, H.H., Barnet, Y.M., Bradley, R.S., Cooper, J.E., De Ley, D.J., Jarvis, B.D.W., Roslycky, E.B., Strijdom, B.W., and Young, J.P.W. 1991, Int. J. Syst. Bacteriol., 41, 582.
13. Dreyfus, B. 2004, 11th AABNF Conference, Dakar (Senegal).
14. Breed, R.S., Murray, E.G.D. and Smith, N.R., ed. s. edition. 1957, Baltimore.: The Williams & Wilkins Co.
15. Colwell, R.R. 1970, J. Bacteriol., 104, 410.
16. Staley, J.T. and Krieg, N.J. 1984, Bergey's manual of systematic bacteriology, N.R.K.a.J.G.H. (Eds.), William and Wilkins, Baltimore, 1.
17. Wayne, L.G., Colwell, R.R., Grimont, P.A.D., Kandler, P., Krichevsky, M.I., Moore, L.H., Moore, W.E.C., Murray, R.G.E., Stackebrandt, E., Starr, M.P., and Trüper, H.G. 1987, Int. J. Syst. Bacteriol., 37, 463.
18. Hey, J. 2001, Trends Ecol. Evol., 16, 326.
19. Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A., Kampfer, P., Maiden, M.C., Nesme, X., Rossello-Mora, R., Swings, J., Truper, H.G., Vauterin, L., Ward, A.C., and Whitman, W.B. 2002, Int. J. Syst. Evol. Microbiol., 52, 1043.
20. Coenye, T., and Vandamme, P. 2003, FEMS Microbiol. Lett., 228, 45.
21. Vandamme, P., Pot, B., Gillis, M., de Vos, P., Kersters, K., and Swings, J. 1996, Microbiol. Rev., 60(2), 407.
22. Zakhia, F., Jeder, H., Domergue, O., Willems, A., Cleyet-Marel, J.C., Gillis, M., Dreyfus, B., and de Lajudie, P. 2004, Syst. Appl. Microbiol., 27, 380.
23. Sy, A., Giraud, E., Jourand, P., Garcia, N., Willems, A., de Lajudie, P., Prin, Y., Neyra, M., Gillis, M., Boivin-Masson, C., and Dreyfus, B. 2001, J. Bacteriol., 183, 214.
24. Jourand, P., Giraud, E., Bena, G., Sy, A., Willems, A., Gillis, M., Dreyfus, B., and de Lajudie, P. 2004, Int. J. Syst. Evol. Microbiol., 54, 2269.
25. Rivas, R., Velazquez, E., Willems, A., Vizcaino, N., Subba-Rao, N.S., Mateos, P.F., Gillis, M., Dazzo, F.B., and Martinez-Molina, E. 2002, Appl. Environ. Microbiol., 68, 5217-22.
26. van Berkum, P., and Eardly, B.D. 2002, Appl. Environ. Microbiol., 68, 1132.
27. Willems, A., Fernandez-Lopez, M., Munoz-Adelantado, E., Goris, J., De Vos, P., Martinez-Romero, E., Toro, N., and Gillis, M. 2003, Int. J. Syst. Evol. Microbiol., 53, 1207.
28. Ngom, A., Nakagawa, Y., Sawada, H., Tsukahara, J., Wakabayashi, S., Uchiumi, T., Nuntagij, A., Kotepong, S., Suzuki, A., Higashi, S., and Abe, M. 2004, J. Gen. Appl. Microbiol., 50, 17.
29. Moulin, L., Munive, A., Dreyfus, B., and Boivin-Masson, C. 2001, Nature, 411, 948.
30. Chen, W.M., Laevens, S., Lee, T.M., Coenye, T., De Vos, P., Mergeay, M., and Vandamme, P. 2001, Int. J. Syst. Evol. Microbiol., 51, 1729.
31. Benhizia, Y., Benhizia, H., Benguedouar, A., Muresu, R., Giacomini, A., and Squartini, A. 2004, Syst. Appl. Microbiol., 27, 462.
32. Van Veen, R.J.M., den Dulk-Ras, H., Bisseling, T., Schilperoort, R.A., and Hooykaas, P.J.J., Mol. Plant-Microbe Interaction, 1, 231.

33. Martinez-Romero, E., Romero, D., and Palacios, R. 1990, Plant Sci., 1, 59.
34. Martinez-Romero, E. 1994, Plant Soil, 161, 11.
35. Demba-Diallo, M., Martens, M., Vloemans, N., Cousin, S., Vandekerckhove, T., Neyra, M., de Lajudie, P., Willems, A., Gillis, M., Vyverman, W., and Van Der Gucht, K. 2004, System. Appl. Microbiol., 27, 238.
36. Zakhia, F., and de Lajudie, P. 2001, Agronomie, 21, 569.
37. Young, J.M., Kuykendall, L.D., Martinez-Romero, E., Kerr, A., and Sawada, H. 2001, Int. J. Syst. Evol. Microbiol., 51, 89.
38. Casida, L.E.J. 1982, Int. J. Syst. Bacteriol., 32, 339.
39. Rogel, M.A., Hernandez-Lucas, I., Kuykendall, L.D., Balkwill, D.L., and Martinez-Romero, E. 2001, Appl. Environ. Microbiol., 67, 3264.
40. Young, J.M. 2003, Int. J. Syst. Evol. Microbiol., 53, 2107.
41. Wang, E.T., Tan, Z.Y., Willems, A., Fernandez-Lopez, M., Reinhold-Hurek, B., and Martinez-Romero, E. 2002, Int. J. Syst. Evol. Microbiol., 52, 1687.
42. Rivas, R., Willems, A., Palomo, J.L., Garcia-Benavides, P., Mateos, P.F., Martinez-Molina, E., Gillis, M., and Velazquez, E. 2004, Int. J. Syst. Evol. Microbiol., 54, 1271.
43. Moreira, F.M.S., Carvalho, Y., Gonçalves, M., Haukka, K., Young, J.P.W., Faria, S.M., Franco, A.A., Cruz, L.M., and Pedrosa, F.O. 1999, Molecules to crop productivity, M.H. F.O. Pedrosa, M.G. Yates and W.E. Newton (Eds.), Current Plant Science and Biotechnology in Agriculture, Kluwer Acad. Publ., 197.
44. Rivas, R., Willems, A., Subba-Rao, N.S., Mateos, P.F., Dazzo, F.B., Kroppenstedt, R.M., Martinez-Molina, E., Gillis, M., and Velazquez, E. 2003, Syst. Appl. Microbiol., 26, 47-53.
45. Vannini, C., Rosati, G., Verni, F., and Petroni, G. 2004, Int. J. Syst. Evol. Microbiol., 54, 1151.
46. Boivin, C., Ndoye, I., Lortet, G., Ndiaye, A., de Lajudie, P., and Dreyfus, B. 1997, Crit. Rev. Plant Sci., 16, 1.
47. Sy, A., Giraud, E., Samba, R., de Lajudie, P., Gillis, M., and Dreyfus, B. 2001, Can. J. Microbiol., 47, 503.
48. Samba, R.T., de Lajudie, P., Gillis, M., Neyra, M., Spencer-Barreto, M.M., and Dreyfus, B. 1999, Symbiosis, 27, 259.
49. Germani, G., and Plenchette, C. 2004, Plant Soil, 266, 333.
50. Jaftha, J.B., Strijdom, B.W., and Steyn, P.L. 2002, Syst. Appl. Microbiol., 25, 440.
51. Yates, R.J., Howieson, J.G., Ardley, J.K., Nistelberger, H., Reeve, W.G., Mutch, L., O'Hara, G.W., and Law, I.J. 2004, 6Sixth European Nitrogen Fixation Conference, Toulouse (France).
52. Lidstrom, M.E., and Stirling, D.I. 1990, Ann. Rev. Microbiol., 44, 27.
53. Green, P.A.1992, The Prokaryotes, H.G. Balows (Ed.), England, 2342.
54. Tan, Z.Y., Wang, E.T., Peng, G.X., Zhu, M.E., Martinez-Romero, E., and Chen, W.X. 1999, Int. J. Syst. Bacteriol., 49, 1457.
55. Vandamme, P., Goris, J., Chen, W.M., de Vos, P., and Willems, A. 2002, Syst. Appl. Microbiol., 25, 507.
56. Vermis, K., Coenye, T., LiPuma, J.J., Mahenthiralingam, E., Nelis, H.J., and Vandamme, P. 2004, Int. J. Syst. Evol. Microbiol., 54, 689.
57. Caballero-Mellado, J., Martinez-Aguilar, L., Paredes-Valdez, G., and Santos, E. 2004, Int. J. Syst. Evol. Microbiol., 54, 1165.

58. Vandamme, P., and Mahenthiralingam, E. 2003, J. Nematology, 35, 208.
59. Chen, W.M., Moulin, L., Bontemps, C., Vandamme, P., Bena, G., and Boivin-Masson, C. 2003, J. Bacteriol., 185, 7266.
60. Vaneechoutte, M., Kampfer, P., De Baere, T., Falsen, E., and Verschraegen, G. 2004, Int. J. Syst. Evol. Microbiol., 54, 317.
61. Alazard, D. 1985, Appl. Environ. Microbiol., 50, 732.
62. Moulin, L., Bena, G., Boivin-Masson, C., and Stepkowski, T. 2004, Mol. Phylogenet. Evol., 30, 720.
63. Eaglesham, A.R.J., Ellis, J.M., Evans, W.R., Fleischman, D.E., Hungria, M., and Hardy, R.W.F. 1990, The first photosynthetic N2-fixing Rhizobium: characteristics, in Nitrogen fixation: Achievements and Objectives, L.E.R. P.M. Gresshoff, G. Stacey and W.L. Newton (Eds.), Chapman and Hall, New York, 805.
64. Lorquin, J., Molouba, F., and Dreyfus, B. 1997, Appl. Environ. Microbiol., 63, 1551.
65. Molouba, F., Lorquin, J., Willems, A., Hoste, B., Giraud, E., Dreyfus, B., Gillis, M., de Lajudie, P., and Masson-Boivin, C. 1999, Appl. Environ. Microbiol., 65, 3084.
66. Willems, A., Coopman, R., and Gillis, M. 2001, Int. J. Syst. Evol. Microbiol., 51,111.
67. Willems, A., Doignon-Bourcier, F., Goris, J., Coopman, R., de Lajudie, P., and Gillis, M. 2001b, Int. J Syst. Evol. Microbiol., 51, 1315.
68. Hannibal, L., Lorquin, J., D'Ortoli, N.A., Garcia, N., Chaintreuil, C., Masson-Boivin, C., Dreyfus, B., and Giraud, E. 2000, J. Bacteriol., 182, 3850.
69. Giraud, E., Fardoux, J., Fourrier, N., Hannibal, L., Genty, B., Bouyer, P., Dreyfus, B., and Vermeglio, A. 2002, Nature, 417, 202.
70. Chaintreuil, C., Giraud, E., Prin, Y., Lorquin, J., Ba, A., Gillis, M., de Lajudie, P., and Dreyfus, B. 2000, Appl. Environ. Microbiol., 66, 5437.
71. Montesinos, E. 2003, Int. Microbiol., 6, 221.
72. Rosenblueth, M., Martinez, L., Silva, J., and Martinez-Romero, E. 2004, Syst. Appl. Microbiol., 27, 27.
73. Rosenblueth, M., and Martinez-Romero, E. 2004, Arch. Microbiol., 181, 337.
74. Ladha, J.K., Garcia, M., Miyan, S., Padre, A., and Watanabe, I. 1989, Appl. Environ. Microbiol., 55, 454.
75. Engelhard, M., Hurek, T., and Reinhold-Hurek, B. 2000, Environ. Microbiol., 2, 131.
76. Yanni, Y.G., Rizk, R.Y., Corich, V., Squartini, A., Ninke, K., Phillip-Hollongsworth, S., Orgambide, G., de Bruijn, F., Stoltzfus, J., Buckley, D., Schmidt, T.M., Mateos, P.F., Ladha, J.K., and Dazzo, F.B. 1997, Plant Soil, 194, 99.
77. Yanni, Y.G., Rizk, R.Y., Abd El-Fattah, F.K., Squartini, A., Corich, V., Giacomini, A., de Bruijn, F., Rademaker, J., Maya-Flores, J., and Ostrom, P. 2001, Aust. J. Plant Physiol., 28, 845.
78. Biswas, J.C., Ladha, J.K., Dazzo, F.B., Yanni, Y.G., and Rolfe, B.G. 2000, Agron. J., 92, 880.
79. Dazzo, F.B., Joseph, A.R., Gomaa, A.B., Yanni, Y.G., and Robertson, G.P. 2003, Symbiosis, 35, 147.
80. Gutierrez-Zamora, M.L., and Martinez-Romero, E. 2001, J. Biotechnol., 91, 117.
81. Martinez-Romero, E. 2003, Plant Soil, 252, 11.
82. Cruz, L., Maltempi de Souza, E., Weber, O.B., Baldani, J.I., Döbereiner, J., and P.F., O. 2001, Appl. Environ. Microbiol., 67, 2375.

83. Knösel, D.H. 2001, Bergey's manual of systematic bacteriology, vol. 1, N.R. Krieg and J.G. Holt (Eds.), Baltimore, U.S.A., 254.
84. Mergaert, J., Cnockaert, M.C., and Swings, J. 2002, Int. J. Syst. Evol. Microbiol., 520, 1828.
85. Elo, S., Maunuksela, L., Salkinoja-Salonen, M., Smolander, A., and Haahtela, K. 2000, FEMS Microbiol. Ecol., 31, 143.
86. Oger, P.M., Mansouri, H., Nesme, X., and Dessaux, Y. 2003, Microb. Ecol., 47, 96.
87. McInroy, J.A. and Kloepper, J.W. 1995, Plant Soil, 173, 337.
88. Hallmann, J., Kloepper, J.W., and Rodriguez-Kabana, R. 1997, Can. J. Microbiol., 43, 411.
89. Sturz, A.V., Christie, B.R., Matheson, B.G., and Nowak, J. 1997, Biol. Fertil. Soils, 25, 13.
90. Sturz, A.V., Christie, B.R., and Matheson, B.G. 1998, Can. J. Microbiol., 44, 162.
91. Lambert, B., Joos, H., Dierickx, S., Vantomme, R., Swings, J., Kersters, K. and van Montagu, M. 1990, Appl. Environ. Microbiol., 56, 1093.
92. Lilley, A.K., Fry, J.C., Bailey, M.J. and Day, M.J. 1996, FEMS Microbiol. Ecol., 21, 231.
93. Bertrand, H., Nalin, R., Bally, R., and Cleyet-Marel, J.-C. 2001, Biol. Fertil. Soils, 33, 152.
94. Larcher, M., Muller, B., Mantelin, S., Rapior, S., and Cleyet-Marel, J.-C. 2003, New Phytol., 160, 119.
95. Valverde, A., Willems, A., Vizcaino, N., Fernandez-Santos, F., Igual, J.M., Mateos, P.F., Martinez-Molina, E., and Velazquez, E. 2004, 6th European Nitrogen Fixation Conference, Toulouse (France).
96. Rosello-Mora, R.A., Ludwig, W., and Schleifer, K.H. 1993, FEMS Microbiol. Lett., 114, 129.
97. Auling, G., Busse, H.J., Egli, T., El-Banna, T., and Stackebrandt, E. 1993, System. Appl. Microbiol., 16, 104.
98. Kampfer, P., Muller, C., Mau, M., Neef, A., Auling, G., Busse, H.J., Osborn, A.M., and Stolz, A. 1999, Int. J. Syst. Bacteriol., 49, 887.
99. de Lajudie, P., Willems, A., Pot, B., Dewettinck, D., Maestrojuan, G., Neyra, M., Collins, M.D., Dreyfus, B., Kersters, K., and Gillis, M. 1994, Int. J. Syst. Bacteriol., 44, 715.
100. Nour, S.M., Fernandez, M.P., Normand, P., and Cleyet-Marel, J.C. 1994, Int. J. Syst. Bacteriol., 44, 511.
101. Xu, L.M., Ge, C., Cui, Z., Li, J., and Fan, H. 1995, Int. J. Syst. Bacteriol., 45, 706.
102. Gao, J.L., Sun, J.G., Li, Y., Wang, E.T., and Chen, W.X. 1994, Int. J. Syst. Bacteriol., 44, 151.
103. Chen, W.X., Wang, E.T., Wang, S.Y., Ly, Y.B., Chen, X.Q., and Li, Y. 1995, Int. J. Syst. Bacteriol., 45, 153.
104. Parker, M.A. 1999, Appl. Environ. Microbiol., 65, 4914.
105. Perret, X., Staehelin, C., and Broughton, W.J. 2000, Microbiol. Mol. Biol. Rev., 64, 180.
106. Lortet, G., Mear, N., Lorquin, J., Dreyfus, B., de Lajudie, P., Rosenberg, C., and Boivin, C. 1996, Mol. Plant-Microbe Interact., 9, 736.
107. Boivin, C., and Giraud, E. 1999, Highlights of Nitrogen Fixation Research, E. Martínez-Romero and G. Hernández (Eds.), Kluwer Academic/ Plenum Publishers, New York, 295.

108. Boivin, C., Ndoye, I., Molouba, F., de Lajudie, P., Dupuy, N., and Dreyfus, B.L. 1997, Appl. Environ. Microbiol., 63,1040.
109. Hungria, M., Boddey, L.H., Santos, M.A., and Vargas, M.A.T. 1998, Biol. Fertil. Soils, 27, 393.
110. Santos, M.A., Vargas, M.A., and Hungria, M. 1999, FEMS Microbiol. Ecol., 1999, 30, 261.
111. Hungria, M., and Bohrer, T.R.J. 2000, Biol. Fertil. Soils, 31, 45.
112. Hungria, M., Campo, R.J., Chueire, L.M.O., Grange, L., and Megias, M. 2001, Biol. Fertil. Soils, 33, 387.
113. Hungria, M., and de Faria, S.M. 2004, 11th AABNF Conference, Dakar (Senegal).
114. Munive, J.A. 2002, Claude Bernard-Lyon 1, Lyon (France).
115. Segovia, L., Pinero, D., Palacios, R., and Martinez-Romero, E. 1991, Appl. Environ. Microbiol., 57, 426.
116. Sivakumaran, S., Lockhart, P.J., and Jarvis, B.D. 1997, Can. J. Microbiol., 43, 164.
117. Laguerre, G., Bardin, M., and Amarger, N. 1993, Can. J. Microbiol., 39, 1142.
118. Sullivan, J.T., Patrick, H.N., Lowther, W.L., Scott, D.B., and Ronson, C.W. 1995, Proc. Natl. Acad. Sci. U.S.A., 92, 8985.
119. Sullivan, J.T., Eardly, B.D., van Berkum, P., and Ronson, C.W. 1996, Appl. Environ. Microbiol., 62, 2818.
120. Downie, J.A., and Young, J.P. 2001, Nature, 412, 597.
121. Laguerre, G., Nour, S.M., Macheret, V., Sanjuan, J., Drouin, P., and Amarger, N. 2001, Microbiology, 147, 981.
122. Turner, S.L., and Young, J.P. 2000, Mol. Biol. Evol., 17, 309.
123. Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M., and Tabata, S. 2002, DNA Res., 2002, 9, 189.
124. Minerdi, D., Fani, R., Gallo, R., Boarino, A., and Bonfante, P. 2001, Appl. Environ. Microbiol., 67, 725.
125. Allen, O.N., and Allen, E.K. 1936, Soil Sci., 42, 61.
126. Allen, O.N., and Allen, E.K. 1939, Soil Sci., 47, 63.
127. Graham, P.H., and Parker, C.A. 1964, Plant Soil, 20, 383.
128. Trinick, M.J. 1965, Aust. J. Sci., 27, 263.
129. Trinick, M.J. 1968, Exp. Agric., 4, 243.
130. Trinick, M.J. 1980, J. Appl. Bacteriol., 49, 39.
131. Magalhaes, F.M.M., and Silva, M.F. 1986, Acta Amazonica, 16/17, 7.
132. Zhang, X., Harper, R., Karisto, M., and Lindström, K. 1991, Int. J. Syst. Bacteriol., 41, 104-113.
133. Barnet, Y.M., and Catt, P.C. 1991, Plant and Soil, 135, 109.
134. Oyaizu, H., Matsumoto, S., Minamisawa, K., and Gamou, T. 1993, J. Gen. Appl. Microbiol., 39, 339.
135. Moreira, F.M.S., Silva, M.F., and Faria, S.M. 1992, New Phytol., 121, 563-570.
136. Moreira, F.M.S., Gillis, M., Pot, B., Kersters, K., and Franco, A.A. 1993, System. Appl. Microbiol., 16, 135.
137. Moreira, F.M.S., Haukka, K., and Young, J.P.W. 1998, Mol. Ecol., 7, 889.

138. Dupuy, N., Willems, A., Pot, B., Dewettinck, D., Vandenbruaene, I., Maestrojuan, G., Dreyfus, B., Kersters, K., Collins, M.D., and Gillis, M. 1994, Int. J. Syst. Bacteriol., 44, 461.
139. de Lajudie, P., Willems, A., Nick, G., Moreira, F., Molouba, F., Hoste, B., Torck, U., Neyra, M., Collins, M.D., Lindstrom, K., Dreyfus, B., and Gillis, M. 1998, Int. J. Syst. Bacteriol., 48, 369-82.
140. Odee, D.W., Sutherland, J.M., Kimiti, J.M., and Sprent, J. 1995, Plant Soil, 173, 211.
141. Haukka, K., Lindström, K., and Young, J.P.W. 1996, System. Appl. Microbiol., 19, 352.
142. Nick, G., Jussila, M., Hoste, B., Niemi, R.M., Kaijalainen, S., de Lajudie, P., Gillis, M., De Bruijn, F.J., and Lindström, K. 1999, System. Appl. Microbiol., 22, 287.
143. Sylla, S.N., Samba, R.T., Neyra, M., Ndoye, I., Giraud, E., Willems, A., de Lajudie, P., and Dreyfus, B. 2002, Syst. Appl. Microbiol., 25, 572.
144. Sylla, S.N. 2003, Dakar University: Dakar (Senegal).
145. Dreyfus, B.L., and Dommergues, Y. 1981, Appl. Environ. Microbiol., 41, 97-99.
146. Dreyfus, B.L., Garcia, J.L., and Gillis, M. 1988, Int. J. Syst. Bacteriol., 38, 89.
147. Chen, W.X., Yan, G.H., and Li, J.L. 1988, Int. J. Syst. Bacteriol., 38, 392.
148. Trüper, H.G., and de Clari, L. 1997, Int. J. Syst. Bacteriol, 47, 908.
149. Lorquin, J., Lortet, G., Ferro, M., Mear, N., Dreyfus, B., Prome, J.C., and Boivin, C. 1997, Mol. Plant Microbe Interact., 10, 879.
150. Ba, S., Willems, A., de Lajudie, P., Roche, P., Jeder, H., Quatrini, P., Neyra, M., Ferro, M., Prome, J.C., Gillis, M., Boivin-Masson, C., and Lorquin, J. 2002, Syst. Appl. Microbiol., 25, 130.
151. Willems, A., Doignon-Bourcier, F., Goris, J., Coopman, R., de Lajudie, P., De Vos, P., and Gillis, M. 51, Int. J. Syst. Evol. Microbiol., 51, 1315.
152. Diouf, A., de Lajudie, P., Neyra, M., Kersters, K., Gillis, M., Martinez-Romero, E., and Gueye, M. 2000, Int. J. Syst. Evol. Microbiol., 50, 159.
153. Doignon-Bourcier, F., Sy, A., Willems, A., Torck, U., Dreyfus, B., Gillis, M., and de Lajudie, P. 1999, Syst. Appl. Microbiol., 22, 647.
154. Doignon-Bourcier, F., Willems, A., Coopman, R., Laguerre, G., Gillis, M., and de Lajudie, P. 2000, Appl. Environ. Microbiol., 66, 3987.
155. de Lajudie, P., Laurent-Fulele, E., Willems, A., Torck, U., Coopman, R., Collins, M.D., Kersters, K., Dreyfus, B., and Gillis, M. 1998, Int. J. Syst. Bacteriol., 48, 1277.
156. Yattara, I.I. 2000, option Ecologie Microbienne, Université du Mali, Institut Supérieur de Formation et de Recherche Appliquée, Bamako, Mali.
157. Haukka, K., and Lindström, K. 1994, FEMS Microbiol. Lett., 119, 215.
158. Wolde-Meskel, E., Terefework, Z., Lindström, K., and Frostegard, A., 2004.
159. Wolde-meskel, E., Terefework, Z., Lindström, K., and Frostegard, A. 2004, Soil Biol. Biochemistry, 36, 2013.
160. Wolde-meskel, E., Terefework, Z., Lindström, K., and Frostegard, A. 2004, System. Appl. Microbiol., 27, 603.
161. Wolde-Meskel, E., Terefework, Z., Lindström, K., and Frostegard, A. 2006, Int. J. Syst. Evol. Microbiol., in press.
162. Kock, M., and Steyn, P.L. 2004, Sixth European Nitrogen Fixation Conference, Toulouse (France).

163. Vinuesa, P., Silva, C., Werner, D., and Martinez-Romero, E. 2005, Mol. Phylogenet. Evol., 34, 29.
164. Zurdo-Pineiro, J.L., Velazquez, E., Lorite, M.J., Brelles-Marino, G., Schroder, E.C., Bedmar, E.J., Mateos, P.F., and Martinez-Molina, E. 2004, Syst. Appl. Microbiol., 27, 469.
165. Parker, M.A. 2004, Syst. Appl. Microbiol., 27, 334.
166. Muller, F., Bouvet, J.M., Bena, G., Vaillant, A., Prin, Y., Le Roux, C., Rousteau, A., Dreyfus, B., and Ba, A. 2004, 11th AABNF Conference, Dakar (Senegal).
167. Parker, M.A. 2003, Mol. Ecol., 12, 2447.
168. Dowdle, S.F., and Bohlool, B.B. 1985, 50, 1171.
169. Han, S.Z., Wang, E.T., and Chan, W.X. 2006, System. Appl. Microbiol., in press.
170. Galiana, A., Gnahousa, M., Chaumont, J., Lesueur, D., and Prin, Y. 1998, Agroforestry Systems, 40, 297.
171. Nuswantara, F., Fujie, M., Sukiman, H.I., Yamashita, M., Yamada, T., and Murooka, Y. 1997, J. Ferment. Bioeng., 84, 511.
172. Kwon, S.W., Park, J.Y., Kim, J.S., Kang, J.W., Cho, Y.H., Lim, C.K., Parker, M.A., and Lee, G.B. 2005, Int. J. Syst. Evol. Microbiol., 55, 263.
173. Bala, A., Murphy, P., and Giller, K.E. 2003, Mol. Ecol., 12, 917.
174. Bala, A., and Giller, K.E. 2001, New Phytol., 149, 495.
175. Räsänen, L.A., Dresler-Nurmi, A., Turkki, E.M., Paulin, L., Andronov, E., Lesueur, D., and Lindström, K. 2004, Sixth European Nitrogen Fixation Conference, Toulouse (France).
176. McInroy, S.G., Campbell, C.D., Haukka, K.E., Odee, D.W., Sprent, J.I., Wang, W.J., Young, J.P., and Sutherland, J.M. 1999, FEMS Microbiol. Lett., 170, 111.
177. Ussery, D.W. 2004, Microbiology, 150, 261.
178. Freiberg, C., Fellay, R., Bairoch, A., Broughton, W.J., Rosenthal, A., and Perret, X. 1997, Nature, 387, 394.
179. Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M., and Tabata, S. 2000, DNA Res., 7, 331.
180. Galibert, F., Finan, T.M., Long, S.R., Puhler, A., Abola, P., Ampe, F., Barloy-Hubler, F., Barnett, M.J., Becker, A., Boistard, P., Bothe, G., Boutry, M., Bowser, L., Buhrmester, J., Cadieu, E., Capela, D., Chain, P., Cowie, A., Davis, R.W., Dreano, S., Federspiel, N.A., Fisher, R.F., Gloux, S., Godrie, T., Goffeau, A., Golding, B., Gouzy, J., Gurjal, M., Hernandez-Lucas, I., Hong, A., Huizar, L., Hyman, R.W., Jones, T., Kahn, D., Kahn, M.L., Kalman, S., Keating, D.H., Kiss, E., Komp, C., Lelaure, V., Masuy, D., Palm, C., Peck, M.C., Pohl, T.M., Portetelle, D., Purnelle, B., Ramsperger, U., Surzycki, R., Thebault, P., Vandenbol, M., Vorholter, F.J., Weidner, S., Wells, D.H., Wong, K., Yeh, K.C., and Batut, J. 2001, Science, 293, 668.
181. Gonzalez, V., Bustos, P., Santamaria, R.I., Hernandez-Gonzales, I., Medrano, A., Chandra, S., Moreno, G., Ramirez, M.A., Gomez, R.E., Fernandez, J.L., Collado, J., and Davila, G. 2004, 6th European Nitrogen Fixation Conference, Toulouse (France).

182. Turner, S.L., Zhang, X.X., Li, F.D., and Young, J.P. 2002, Microbiology, 148, 3330.
183. Wood, D.W., Setubal, J.C., Kaul, R., Monks, D.E., Kitajima, J.P., Okura, V.K., Zhou, Y., Chen, L., Wood, G.E., Almeida, N.F., Jr., Woo, L., Chen, Y., Paulsen, I.T., Eisen, J.A., Karp, P.D., Bovee, D., Sr., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Kutyavin, T., Levy, R., Li, M.J., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Romero, P., Gordon, D., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z.Y., Dolan, M., Chumley, F., Tingey, S.V., Tomb, J.F., Gordon, M.P., Olson, M.V., and Nester, E.W. 2001, Science, 294, 2317.
184. Sullivan, J.T., and Ronson, C.W. 1998, Proc. Natl. Acad. Sci. U.S.A., 95, 5145.
185. Young, J.P.W. 1999, Nitrogen Fixation: from molecules to crop productivity, M.H. Pedrosa, M.G. Yates and W.E. Newton (Eds.), Kluwer Acad. Publ.
186. Gaunt, M.W., Turner, S.L., Rigottier-Gois, L., Lloyd-Macgilp, S.A., and Young, J.P. 2001, Int. J. Syst. Evol. Microbiol., 51, 2037.
187. Cohan, F.M. 2002, Annu. Rev. Microbiol., 56, 457.
188. Dykhuisen, D.E., and Green, L. 1991, J. Bacteriol., 173, 7257.
189. Willems, A., and Collins, M.D. 1993, Int. J. Syst. Bacteriol., 43, 305.
190. Conye, T., and Vandamme, P. 2003, Microbiology, 149, 3507.
191. Eardly, B., and Van Berkum, P.2005, Symbiosis, in press.
192. Vinuesa, P., Leon-Barrios, M., Silva, C., Willems, A., Jarabo-Lorenzo, A., Perez-Galdona, R., Werner, D., and Martinez-Romero, E. 2005, Int. J. Syst. Evol. Microbiol., 55, 569.
193. Frank, B. 1889, Ber. Deut. Bot. Ges., 7, 332.
194. Martinez-Romero, E., Segovia, L., Mercante, F.M., Franco, A.A., Graham, P., and Pardo, M.A. 1991, Int. J. Syst. Bacteriol., 41, 417.
195. Segovia, L., Young, J.P., and Martinez-Romero, E. 1993, Int. J. Syst. Bacteriol., 43, 374.
196. Hernandez-Lucas, I., Segovia, L., Martinez-Romero, E., and Pueppke, S.G. 1995, Appl. Environ. Microbiol., 61, 2775.
197. Wang, E.T., Rogel, M.A., Garcia-de los Santos, A., Martinez-Romero, J., Cevallos, M.A., and Martinez-Romero, E. 1999, Int. J. Syst. Bacteriol., 49, 1479.
198. Chen, W.X., Tan, Z.Y., Gao, J.L., Li, Y., and Wang, E.T. 1997, Int. J. Syst. Bacteriol., 47, 870.
199. Amarger, N., Macheret, V., and Laguerre, G. 1997, Int. J. Syst. Bacteriol., 47, 996.
200. van Berkum, P., Beyene, D., Bao, G., Campbell, T.A., and Eardly, B.D. 1998, Int. J. Syst. Bacteriol., 48, 13.
201. Lindström, K. 1989, Int. J. Syst. Bacteriol., 39, 365.
202. Radeva, G., Jurgens, G., Niemi, M., Nick, G., Suominen, L., and Lindstrom, K. 2001, Syst. Appl. Microbiol., 24, 192.
203. Wang, E.T., van Berkum, P., Beyene, D., Sui, X.H., Dorado, O., Chen, W.X., and Martinez-Romero, E. 1998, Int. J. Syst. Bacteriol., 48, 687.
204. Wei, G.H., Wang, E.T., Tan, Z.Y., Zhu, M.E., and Chen, W.X. 2002, Int. J. Syst. Evol. Microbiol., 52, 2231.
205. Squartini, A., Struffi, P., Doring, H., Selenska-Pobell, S., Tola, E., Giacomini, A., Vendramin, E., Velazquez, E., Mateos, P.F., Martinez-Molina, E., Dazzo, F.B., Casella, S., and Nuti, M.P. 2002, Int. J. Syst. Evol. Microbiol., 52, 1267.

206. Wei, G.H., Tan, Z.Y., Zhu, M.E., Wang, E.T., Han, S.Z., and Chen, W.X. 2003, Int. J. Syst. Evol. Microbiol., 53, 1575.
207. Tan, Z.Y., Kan, F.L., Peng, G.X., Wang, E.T., Reinhold-Hurek, B., and Chen, W.X., Int. J. Syst. Evol. Microbiol., 51, 909.
208. Dangeard, P.A. 1926, Paris: Le Botaniste, 270.
209. Scholla, M.H., and Elkan, G.H. 1984, Int. J. Syst. Bacteriol., 34, 484.
210. Rome, S., Fernandez, M.P., Brunel, B., Normand, P., and Cleyet-Marel, J.C. 1996, Int. J. Syst. Bacteriol., 46, 972-80.
211. Nick, G., de Lajudie, P., Eardly, B.D., Suomalainen, S., Paulin, L., Zhang, X., Gillis, M., and Lindstrom, K. 1999, Int. J. Syst. Bacteriol., 49, 1359.
212. Toledo, I., Lloret, L., and Martinez-Romero, E. 2003, Syst. Appl. Microbiol., 26, 54.
213. Willems, A., Munive, A., de Lajudie, P., and Gillis, M. 2003, Syst. Appl. Microbiol., 26, 203.
214. Jarvis, B.D.W., Pankhurst, C.E., and Patel, J.J. 1982, Int. J. Syst. Bacteriol., 32, 378.
215. Chen, W.X., Li, G.S., Qi, Y.L., Wang, E.T., Yuan, H.L., and Li, J.L. 1991, Int. J. Syst. Bacteriol., 41, 275.
216. Nour, S.M., Cleyet-Marel, J.C., Normand, P., and Fernandez, M.P. 1995, Int. J. Syst. Bacteriol., 45, 640.
217. Velazquez, E., Igual, J.M., Willems, A., Fernandez, M.P., Munoz, E., Mateos, P.F., Abril, A., Toro, N., Normand, P., Cervantes, E., Gillis, M., and Martinez-Molina, E. 2001, Int. J. Syst. Evol. Microbiol., 51, 1011.
218. Rinaudo, G., Orenga, S., Fernandez, M., Meugnier, H., and Bardin, R. 1991, Int. J. Syst. Bacteriol., 41, 114.
219. Kirchner, O. 1896, Beitr. Biol. Pflanz., 7, 213.
220. Kuykendall, L.M., Saxena, B., Devine, T.E., and Udell, S.E. 1992, Can. J. Microbiol., 38, 501.
221. Yao, Z.Y., Kan, F.L., Wang, E.T., Wei, G.H., and Chen, W.X. 2002, Int. J. Syst. Evol. Microbiol., 52, 2219.
222. Young, J.P.W. 1991, Biological Nitrogen Fixation, G. Stacey (Ed.), Chapman & Hall, New York, 43.

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.