

## Abréviations

|         |   |
|---------|---|
| - ADN   | Acide désoxyribonucléique   |
| - ADNc  | Acide désoxyribonucléique complémentaire  |
| - BET   | Bromure d'éthidium  |
| - BSA   | Bovine serum albumine   |
| - CAPS  | Cleaved amplified polymorphic sequence  |
| - CCR   | Cinnamoyl CoA reductase   |
| - CIRAD | Centre de Coopération Internationale en Recherche Agronomique pour le Développement |
| - cM    | Centimorgan   |
| - dNTP  | Désoxynucléotide triphosphate   |
| - EDTA  | Acide éthylène diamine tétracétique   |
| - EST   | Expressed sequence tag  |
| - GL    | Groupe de liaison   |
| - h     | Heure   |
| - HCl   | Acide chlorhydrique   |
| - LOD   | Log10 of the odds ratio   |
| - M     | Molaire   |
| - min   | Minute  |
| - ml    | Millilitre  |
| - mM    | Millimolaire  |
| - ng    | Nanogramme  |
| - pb    | Paire de bases  |
| - PCR   | Polymerase chain reaction   |
| - QTL   | Quantitative trait loci   |
| - RAPD  | Random amplified polymorphic DNA  |
| - RFLP  | Restriction fragment length polymorphism  |
| - SAM   | Sélection assistée par marqueurs  |
| - sec   | Seconde   |
| - SSCP  | Single strand conformation polymorphism   |
| - SSH   | Suppressive subtractive hybridization   |
| - SSR   | Single sequence repeat ou microsatellite  |
| - STS   | Sequence-tagged site  |
| - TBE   | Tris borate EDTA  |
| - UV    | Ultra violet  |
| - V     | Volt  |
| - µg    | Microgramme   |
| - µl    | Microlitre  |
| - µmol  | micromole   |

# Sommaire

|  |           |
|--|-----------|
| <b>1. INTRODUCTION</b> .....   | <b>1</b>  |
| <b>2. MATERIELS ET METHODES</b> .....  | <b>6</b>  |
| 1. MATERIEL VEGETAL.....   | 6         |
| 2. EXTRACTION D'ADN GENOMIQUE .....  | 6         |
| 3. CARACTERISTIQUES DES EST ET DES GENES ETUDIES .....                           | 7         |
| 4. MISE AU POINT ET CONDITIONS D'AMPLIFICATION PCR.....                          | 8         |
| 5. REVELATION DU POLYMORPHISME CHEZ LES GENITEURS .....                          | 9         |
| <i>Par SSCP</i> .....  | 9         |
| <i>Par PCR-RFLP</i> .....  | 10        |
| <i>Séquençage des produits PCR</i> .....   | 10        |
| 6. CARTOGRAPHIE GENETIQUE.....   | 11        |
| <b>3. RESULTATS</b> .....  | <b>13</b> |
| 1. AMPLIFICATION DES EST ET DES GENES .....                                      | 13        |
| 2. DETECTION DU POLYMORPHISME .....  | 14        |
| <i>Par SSCP</i> .....  | 14        |
| <i>Par utilisation d'enzymes de restriction</i> .....                            | 16        |
| <i>Séquençage des produits d'amplification</i> .....                             | 16        |
| 3. CARTOGRAPHIE GENETIQUE DES SEQUENCES .....                                    | 18        |
| <i>Localisation des séquences polymorphes uniquement chez E. urophylla</i> ..... | 20        |
| <i>Localisation des EST polymorphes uniquement chez E. grandis</i> .....         | 20        |
| <i>Localisation des séquences polymorphes chez les deux géniteurs</i> .....      | 20        |
| <b>4. DISCUSSION</b> .....   | <b>21</b> |
| 1. SSCP .....  | 21        |
| 2. LES CAUSES MOLECULAIRES DE LA DETECTION DE POLYMORPHISME EN SSCP.....         | 22        |
| 3. LES AUTRES APPORTS DU SEQUENÇAGE .....  | 24        |
| 4. LA TRANSFERABILITE DES EST .....  | 25        |
| 5. CARTOGRAPHIE.....   | 26        |
| <b>5. CONCLUSION ET PERSPECTIVES</b> .....                                       | <b>29</b> |

# 1. Introduction

Les eucalyptus sont les arbres les plus plantés dans le monde. Les estimations faites en 2000 (FAO, 2001) indiquaient que les plantations d'eucalyptus représentaient environ 18 millions hectares dans le monde.

Sur 700 espèces environ que compte le genre *Eucalyptus* originaire d'Australie et des Iles de la Sonde, seule une dizaine sont plantées dans les régions d'introduction dont *E. urophylla*, *E. grandis*, *E. gunnii* et *E. globulus* appartenant toutes au même sous genre (*Symphyomyrtus*). Les exigences édaphiques de ce genre sont très variables d'une espèce à l'autre. Ainsi trouve-t-on des espèces adaptées aux conditions climatiques tropicales et équatoriales (*E. urophylla*) et d'autres adaptées aux conditions tempérées apparentées à celle des hauts plateaux de Tasmanie (*E. gunnii*, *E. globulus*) expliquant une aire d'implantation entre les latitudes 40° nord et 45° sud. D'autres paramètres ont contribué à l'engouement général pour ces essences : une croissance juvénile forte et continue, une rotation variant de 7 à 10 ans quand il faut attendre 20 ans pour le peuplier, la possibilité d'une multiplication végétative par bouturage, la propriété de rejeter après coupe permettant plusieurs rotations sur la même plantation. La valorisation du bois d'eucalyptus s'effectue dans plusieurs domaines. D'après Vignerot et Bouvet (1997), 80% de la production de bois d'eucalyptus est destiné au bois d'énergie (bois de chauffe et charbon). La fabrication de pâte à papier représente également une part importante de son utilisation (1/3 de la production mondiale), son rendement papetier étant très supérieur aux autres feuillus (Afocel, 2003). Dans son aire d'origine, l'eucalyptus est utilisé comme bois d'œuvre mais cette utilisation reste rare dans les pays d'introduction. En effet, le bois issu de plantations après coupe est sujet à de nombreuses déformations (fentes en bout, retraits...) liées aux contraintes de croissance notamment. D'autres valorisations non ligneuses sont à noter : les huiles essentielles (dont eucalyptol) que contiennent les feuilles sont valorisées en pharmacopée, le feuillage est également utilisé dans la fabrication de bouquets. Certaines espèces sont par ailleurs utilisées à titre ornemental dans les jardins et parcs.

Au Congo, les plantations industrielles d'eucalyptus bénéficient d'un programme d'amélioration génétique local. Ce programme initié en 1955 par le Centre Technique Forestier Tropical (CTFT, qui deviendra le CIRAD forêt) a débuté par l'introduction d'une quarantaine d'espèces d'eucalyptus. Le but était alors la valorisation des zones de savanes par le reboisement, pour la production locale de bois de chauffe, de bois de trituration et de charbon de bois. Dans les générations suivantes, des hybrides naturels sont apparus se révélant plus productifs que les espèces implantées. Le fort hétérosis observé chez ces hybrides a naturellement conduit à leur plantation à grande échelle et a suscité la mise en route d'un programme de croisements contrôlés. Parmi les combinaisons hybrides testées, deux se sont révélées particulièrement intéressantes : *E. urophylla* x *E. grandis* et *E. urophylla* x *E. pellita* (*E. urophylla* et *E. pellita* bien adaptées aux conditions congolaises, *E. grandis* mal adaptée mais intéressante pour sa productivité et sa qualité papetière). Ces hybrides artificiels ont permis une augmentation significative de la production de bois. En une trentaine d'années, la productivité des plantations est passée de 10 m<sup>3</sup>/ha/an à 20 m<sup>3</sup>/ha/an. Depuis 1989, un schéma de sélection récurrente réciproque (SSR) a été adopté. Son principe est d'améliorer conjointement deux populations constituées par *E. grandis* et *E. urophylla*. Un plan de croisement entre ces deux populations a été réalisé et a permis de tester plusieurs centaines de familles afin d'évaluer la valeur des géniteurs en croisement à partir de la valeur des descendants sur des critères de croissance, les mères étant *E. urophylla* et les pères étant *E. grandis*. Les arbres sélectionnés sont recombinaisonnés entre eux dans l'objectif de produire des hybrides combinant différents caractères présents séparément dans chacune des deux populations. A chaque étape, des hybrides sont sélectionnés et multipliés par bouturage horticole pour des plantations clonales.

L'amélioration des essences forestières est confrontée aux délais d'attente et aux contraintes de place (Plomion *et al.*, 1996). En effet, avant de pouvoir évaluer les individus sur les caractères de croissance et de qualité du bois il faut attendre, dans le cas de l'eucalyptus, 3 ans (Bouvet, 1995, Osorio *et al.*, 2002). Avant de récolter des graines ou d'effectuer des croisements, la maturité sexuelle des arbres (de 3 à 5 ans chez l'eucalyptus) doit être acquise. De plus, les caractères intéressants pour le sélectionneur (croissance, adaptation, qualité du bois) sont soumis à de forts effets environnementaux, ils sont également gouvernés par plusieurs gènes et l'héritabilité de ces caractères est généralement faible à moyenne variant de 0,10 à 0,40 (Kremer, 1994).

Dans l'optique d'accroître les gains génétiques et de développer plus rapidement des variétés d'*Eucalyptus* à forte productivité adaptées aux applications industrielles, deux stratégies sont envisageables. D'une part la transformation génétique par introduction de gènes favorisant tel ou tel caractère, combinée à la production en masse de matériel par la multiplication *in vitro*, permet de construire puis de produire rapidement des génotypes répondant aux attentes des industriels. Toutefois, cette voie se développe peu en raison de l'opinion publique face aux OGM. L'autre voie consiste à exploiter la variabilité génétique naturelle des gènes impliqués dans le contrôle de caractères d'intérêt. La sélection assistée par marqueurs (SAM) a été développée dans cette optique et passe par l'appui de marqueurs anonymes mais de plus en plus par le positionnement des gènes.

Depuis quelques années, avec le développement d'outils d'analyses de plus en plus précis, le marquage moléculaire ainsi que la connaissance structurale et fonctionnelle du génome de nombreuses plantes ont connu un essor considérable. Les eucalyptus, de par leur importance économique, n'ont pas échappé à ce mouvement. En ce sens, plusieurs cartes génétiques d'*Eucalyptus sp.* ont été construites durant la dernière décennie reposant essentiellement sur des marqueurs RAPD (random amplified polymorphic fragment) et AFLP (amplified fragment length polymorphism) (Grattapaglia et Sederoff, 1994, et Plomion, 1996, Marques *et al.*, 1998) en utilisant la stratégie du double pseudo-testcross sur une population F1 interspécifique ou sur une combinaison de marqueurs RAPD et RFPL (restriction fragment polymorphism length) (Byrne *et al.*, 1994). Ces cartes génétiques ont fourni une couverture suffisante du génome pour détecter des zones d'intérêts à caractères qualitatif ou quantitatif (Verhaegen *et al.*, 1997). L'emploi de marqueurs dominants et anonymes (Brondani *et al.*, 2002) a limité l'utilisation de ces cartes au croisement dont elles sont issues et un enrichissement progressif de ces cartes en marqueurs codominants qui sont transférables et comparables est en cours (Gion *et al.*, 2000, Thamarus *et al.*, 2002, Marques *et al.*, 2002). Parmi ces marqueurs codominants, il existe deux approches différentes :

- ✓ D'une part, l'utilisation de microsatellites ou SSR dont le polymorphisme correspond au nombre de répétitions de motifs nucléotidiques de 2 à 5 bases. Ces marqueurs sont anonymes, très variables et uniformément répartis sur le génome des plantes.
- ✓ L'approche gène candidat et la constitution de banques de génome exprimé constituent une source importante de marqueurs codominants. Pour leur part, ils correspondent à des zones du génome dont dépendent les caractères phénotypiques.

Avec l'accroissement constant des banques de données génomiques, de plus en plus de séquences de fonctions connues ou hypothétiques sont disponibles chez de nombreuses espèces annuelles ou pérennes. Cette information représente une source potentielle considérable de gènes candidats. Ainsi, plusieurs possibilités sont envisageables pour la cartographie de ces gènes. Le gène cible a déjà été identifié chez l'espèce étudiée ou une espèce proche, il peut devenir une source de marqueurs RFLP, CAPS ou être utilisé comme tel et sa variabilité en population peut être exploitée. Si aucune séquence du gène n'est disponible chez l'espèce étudiée, la séquence consensus des homologues éventuellement disponibles dans les bases de données peut permettre la définition d'amorces et isoler le gène ciblé. Cependant, les processus biologiques sont encore loin d'être connus surtout chez les arbres forestiers et pour mieux aborder ces aspects, de nombreuses banques d'ADNc spécialisées ont été construites (formation des racines, formation du xylème, spécifiques à la feuille...).

La qualité du bois est un caractère complexe et hétérogène, variant entre génotype mais également au sein même de l'arbre. Le bois (ou xylème secondaire) se forme en cinq grandes étapes incluant la division cellulaire, l'expansion cellulaire (élongation longitudinale et radiale), la formation des parois primaires et secondaires, la mort cellulaire programmée et la formation de l'aubier (Plomion *et al.*, 2001). Le développement de la biologie moléculaire a permis d'appréhender certains mécanismes moléculaires qui gouvernent ce processus même si de nombreux travaux restent à faire. Cette connaissance a permis de mettre en lumière certaines protéines et enzymes intervenant dans la formation du bois dont on peut en déduire les séquences codantes du génome constituant d'excellents gènes candidats pour la sélection et la colocalisation gènes/QTL. A titre d'exemple, le gène codant pour la Cinnamoyl CoA réductase (CCR), enzyme de la voie de biosynthèse des lignines, colocalise avec un QTL de teneur en lignines (Gion *et al.*, 2000). De plus l'apport et les retombées des programmes de séquençage total de génomes d'arbres forestiers en cours (peuplier, eucalyptus) vont permettre l'obtention de nouvelles informations sur le fonctionnement des arbres. Plusieurs banques ont été réalisées pour le génome exprimé (banques d'EST et d'ADNc développées chez de nombreux genres dont l'eucalyptus) (Byrne *et al.*, 1994). Avec le développement de la technique dite SSH (suppressive soustractive hybridation (Diatchenko *et al.*, 1996), qui permet l'enrichissement de la banque en gènes surexprimés dans une condition par rapport à une autre, de nouvelles banques plus ciblées sont apparues. Ainsi, Bossinger et Leitch (2000) ont créé une banque d'ADNc enrichie en gènes issus du cambium d'*E. globulus*.

Pour cartographier de telles EST, plusieurs techniques sont utilisables. La plus simple est la PCR-RFLP ou CAPS (cleaved amplified polymorphic sequence, Tragoonrung *et al.*, 1992). Elle consiste en une digestion des produits amplifiés avec une enzyme de restriction spécifique et une visualisation directe des produits obtenus sur gel d'agarose. D'autres techniques comme la DGGE (denaturing gradient gel electrophoresis, Myers *et al.*, 1985) ou la TGGE (temperature gradient gel electrophoresis, Riesner *et al.*, 1992) sont basées sur le principe de comparaison de la stabilité de l'ADN sous des conditions de gels spécifiques. Ces techniques sont simples a priori mais demandent des conditions très contrôlées lors de la migration sur gel. Une autre alternative consiste à révéler du polymorphisme sur des fragments amplifiés par la SSCP (single strand conformation polymorphism, Orita *et al.*, 1989). Cette approche est basée sur les différences de conformation des simples brins d'ADN liées à leur séquence et variant pour une seule mutation. Ces conformations peuvent montrer une mobilité différente sur un gel en conditions non dénaturantes. Nous pouvons également envisager la solution du séquençage de ces EST permettant de déterminer précisément les variations de séquence. Toutefois, cette approche, trop onéreuse, est peu réaliste pour la cartographie qui nécessite plusieurs centaines d'individus (De Vienne, 1995).

Dans ce contexte, un partenariat est né entre le CNRS de Toulouse, le CIRAD Forêt (Montpellier) et deux entreprises privées du secteur papetier, RAIZ (Portugal) et ENCE (Espagne). Le CNRS a utilisé la technique SSH à partir d'*E. gunnii* pour obtenir une banque d'EST enrichie en gènes surexprimés dans le xylème par rapport à la feuille (Paux *et al.*, sous presse). Ce projet a permis d'isoler 250 séquences non homologues dont 57% n'ont pas de fonction connue. Toutes ces séquences ont été réparties entre les trois autres partenaires dont le travail va de la définition d'amorces à la cartographie et la détection de QTL pour leurs croisements élites respectifs. Ce stage s'inscrit dans cette approche gène candidat pour la qualité du bois menée par le CIRAD Forêt dont la première étape consiste à cartographier une partie des EST dont les amorces ont été prédéfinies et testées sur *E. globulus* par RAIZ. Pour cela, il a fallu adapter les conditions d'amplification au laboratoire de génétique du CIRAD, révéler du polymorphisme par différentes techniques dont principalement la SSCP et cartographier les EST polymorphes sur les cartes génétiques déjà existantes d'*E. grandis* et d'*E. urophylla* à partir d'une population F1 de 200 individus.

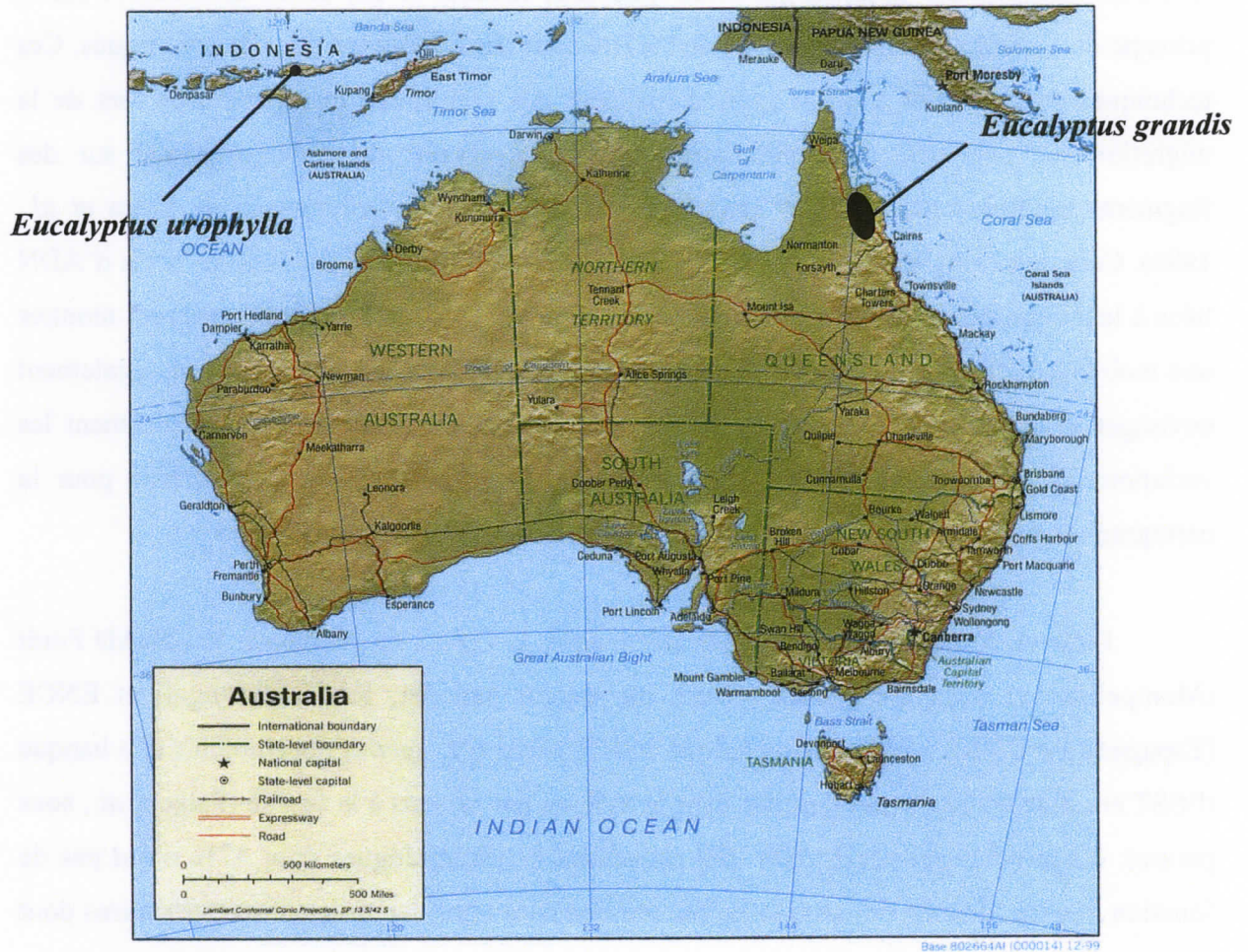


Figure 1 : provenance des deux géniteurs (*E. urophylla* et *E. grandis*) du croisement élite utilisé dans cette étude.



## 2. Matériels et méthodes

### 1. Matériel végétal

Les travaux de cartographie génétique ont été réalisés avec une famille d'hybrides interspécifiques *E. urophylla* \* *E. grandis* produite dans le cadre du programme d'amélioration génétique des eucalyptus au Congo. Le géniteur *E. urophylla* 14.144 (géniteur femelle) est originaire du Mont Lewotobi de l'île de Florès située au nord-ouest de l'Australie tandis que le géniteur *E. grandis* 9.21 (géniteur mâle) provient d'Atherton située au nord-est de l'état du Queensland (Australie, figure 1). Cette descendance élite totalise 201 individus pleins frères et constitue la base des différents travaux de cartographie et de détection de QTL (Verhaegen et Plomion, 1996 ; Verhaegen *et al.*, 1997 ; Gion *et al.*, 2000 ; Carrouché 2003).

### 2. Extraction d'ADN génomique

L'extraction est réalisée selon un protocole modifié de Saghäi-Marouf *et al.* (1984), à partir de jeunes feuilles prélevées sur les arbres adultes et séchées à l'air libre. Pour chaque individu, 300mg de feuilles sont broyées avec une pincée de polyvinylpyrrolidone, le tout sous azote liquide. La poudre obtenue est incubée sous agitation durant 1 heure (h) à 65°C avec 9 ml de tampon d'extraction stérile (Tris HCl 0,1 M pH 8, EDTA 0,02 M, ATMAB 20g/l, NaCl 1,4M) et 0,1% mercaptoéthanol. Après refroidissement des tubes à l'air libre, 4,5 ml de dichlorométhane sont rajoutés. Les extraits sont mis à agitation douce pendant 10 minutes (min) puis centrifugés 10 min à 4000 T/min. Le surnageant est prélevé et l'étape précédente est reconduite tant que les extraits ne sont pas limpides (2 à 5 fois en général). L'ADN génomique est précipité à froid par ajout de 6 ml d'isopropanol glacé, placé à -20°C pendant 1 h puis centrifugé à 4000T/min pendant 10 min. Le surnageant est éliminé et les culots sont mis à sécher au Speed Vac durant 30 min à 40°C. Ensuite, ils sont repris sous agitation dans 1 ml de tampon de stockage (EDTA 10<sup>-4</sup> M, NH<sub>4</sub>OAc 0,01M) toute la nuit. Le lendemain, les extraits sont précipités par ajout de 50µl de NaCl 5M et 2,5ml d'éthanol absolu et centrifugés à 4000T/min pendant 10 min. Une fois le surnageant éliminé, 5 ml de solution de lavage 1

Tableau 1 : Description des EST utilisées. Les indications données ci dessous sont une partie des informations transmises au CIRAD par la compagnie RAIZ (tableau complet en annexe)

| Nom séquence | Fonction probable (BLAST homology)                                | Séquences des amorces sens (5' → 3') | Séquences des amorces antisens (5' → 3') |
|--------------|---|--------------------------------------|--|
| GF#27        | xyloglucan endo-1, 4-beta-D-glucanase                             | TTATGTGGATGAAGTGCCC                  | AACTTGAGCCTCCCAGTGT                      |
| GF#86        | alpha-tubulin   | CTTCCACCCTGAGCAACTC                  | AAAGCACAGCAACATCGGT                      |
| 08A09        | No Hits   | GTACGCGGGGATATGACAAG                 | GCCCGTTAATAGAAAAGCAC                     |
| 08B02        | hypothetical protein [Methanosarcina mazei Goe1]                  | ATTTACGGGTGACCAAAGA                  | ATGCGGCTCAAGGTGTATCA                     |
| 08B06        | similar to DNA repair protein RAD50, putative                     | TTTACGAGCCTCTCCTTGA                  | AAGCAGCAGCTATTGTTTCC                     |
| 08C02        | outer capsid protein  | GAAATACTGCTTCTCCAAG                  | CACAAGATGCCTCAACAATG                     |
| 08C05        | Contains similarity to  | AGAAGGGCGGTGATTTAGGA                 | GACCCATCAAGATGTCTCCAGT                   |
| 08C08        | peptidyl-prolyl cis-trans isomerase                               | ACTCACCCAACCTCAACCAA                 | CGTTTGCCACGATGTATTTT                     |
| 08C10        | coat protein [Cucumber mosaic virus]                              | TTGATGTGCCGTATTCTTTT                 | TTGTTGCCCTTTACTACTCAC                    |
| 08E04        | Plasmodium falciparum trophozoite antigen                         | GCCTGCACTTACAATACCGC                 | GATTACTACACGGGCTTCACAA                   |
| 08E06        | cellulase (EC 3.2.1.4) T21L14.7 - Arabidopsis thaliana P0529E05.5 | CAATGGGCTGTTGAGTTCT                  | GTGTTGGAGGGTGAGGGTTA                     |
| 08E11        | [Oryza sativa (japonica cultivar-group)]                          | GGGGCCGAAACTCGCTTCTC                 | ATGGCATCACGCTCTGGTCT                     |
| 08E12        | hypothetical protein - Arabidopsis thaliana                       | TTCAATAATCCCTGGTCAAT                 | CCAAACGCACCTTAATAGTC                     |
| 08F01        | hypothetical protein F11C1.10 - Arabidopsis thaliana              | ATAGAACACGCGTCCATACT                 | GCATTTTGATCGCTTTGATT                     |
| 08F03        | No Hits   | GAGGAGGAGCAGAAAAGCAGG                | AAATCAGCATGAAAGCGGTA                     |
| 08F08        | ceo protein [Arabidopsis thaliana]                                | ACCGCCGATTTACTCTTACT                 | TCCGATAATCGACTATTCCA                     |
| 08F09        | No Hits   | GGGCACTTTCAGGCACCACC                 | GCATCGCATCGTATCGGGAG                     |
| 08F09        | hypothetical protein; 44869-47459 [Arabidopsis thaliana]          | CGAAGCGAGGACTGAAGAAG                 | CGACCGATGATTTGATAGCC                     |
| 08G01        | hypothetical protein [Plasmodium yoelii yoelii]                   | AAAGTTTCCCTGTATCTTCTCA               | CGTGTATTTAGTTTGCTTCTC                    |
| 08G05        | gamma-aminobutyric acid (GABA) A receptor, beta 1 [Homo sapiens]  | GGGAGTAGATGGTGACCTGGAA               | GCGGGAAGCATGAAAGCAAG                     |
| 08G06        | putative harpin-induced protein [Arabidopsis thaliana]            | TGAAGCAAATCGAGTTGTTT                 | TTTACGCCAAAACATGTCAA                     |
| 08G08        | No Hits   | TTTTGATGATACGGCTGGAA                 | GAGGGACTCTGTGGTTCTTG                     |
| 08G09        | Putative transposase number 3 for insertion sequence ISRM26       | CGGGGAAACCTACTTGCTAA                 | GCATAAGTGAACACCAGGAT                     |
| 08H01        | conserved hypothetical protein [Streptococcus pyogenes M1 GAS]    | TGAAGTTCAGCGGCAAGAAG                 | TTGGAGGAAAACAGCTAACCAAT                  |
| 08H04        | unknown [Glycine max]   | AATTTATTTGTTTCGAGGGTG                | CAACGCATTTCGTGAGTTAGT                    |
| 08H07        | hypothetical protein [Azotobacter vinelandii]                     | AAGCATCGGTTGCCAAGGAA                 | GCTGCTGTGCCAGGTGAAG                      |
| 08H11        | unknown [Arabidopsis thaliana]                                    | ATGACAAATAAGTCTATGGAAA               | TACCGCATCCTCTACAAGAA                     |
| 09A11        | unknown protein [Arabidopsis thaliana]                            | GCCGAGCTGATGCAATAACT                 | GACCTTCTGGGTCTTGAAA                      |
| 09C05        | No Hits   | GTACATTGCCAAGGGGATAA                 | CAGGTCTTCCGAATACAGA                      |
| 09C08        | No Hits   | TGGACGATGAGGACCAGAAG                 | GGTGAACCATAATCAGGGAG                     |
| 09C10        | Unknown protein [Oryza sativa (japonica cultivar-group)]          | ATCGAGGAGGCATACGACC                  | CGGCGGAGGCAGAATAGAC                      |
| 09D01        | Transformer serine/arginine-rich ribonucleoprotein, putative      | GAAGGATGGGGTTCACAAGG                 | CAACATGGATCAACCGAAAA                     |
| 09D03        | subtilisin-like proteinase  | GTACGGTTTGCATGATTGTT                 | TTAATCGCAATCCACAACAT                     |
| 08D04        | No Hits   | AGCTGGTTATTCTTCAGAA                  | TAAGTGAAGCAACCGAGTG                      |
| 09E01        | No Hits   | AAAGCACAAAGGGAGAAGAAA                | CTCAGAATTAGATGCCAAGA                     |
| 09F08        | VI  | CGAAATGATTATGGGAAAC                  | ATCACTTCTATGGACCGTGT                     |
| 09H01        | MAP protein kinase [Arabidopsis thaliana]                         | ACTTTGCTTGTGGTTATTTT                 | AATTCCTTCCGACTTGACTT                     |
| 09H02        | hypothetical protein XP_243418 [Rattus norvegicus]                | GGACAGAGCTTCAGCAGATT                 | CCACTTTACTCCAAGCAATT                     |
| 09H05        | putative translation initiation factor                            | GGCATAGTCGATCACATTAG                 | AGTCGCCTGAGATAAACTGG                     |
| 09H06        | cell surface protein precursor [Lactobacillus plantarum WCFS1]    |                                      |  |

(éthanol 75 %, NaOAc 0,2 M) est ajouté, la mise en contact dure 20 min puis les tubes sont centrifugés 10 min à 4000T/min. Le surnageant est à nouveau éliminé et 2 ml de solution de lavage 2 (éthanol 75%, NH<sub>4</sub>OAc 0,01M) sont ajoutés, puis les extraits sont centrifugés 10 min à 4000T/min. Les culots sont alors lavés 10 min par centrifugation avec 1 ml d'éthanol à 70%. Enfin, la pelote est séchée 30 min à 40°C au Speed Vac puis reprise dans un volume d'eau stérile compris entre 50µl et 2 ml selon sa taille et conservée à -20°C.

La concentration d'ADN est mesurée avec un fluorimètre (TKO 100 Minifluorometer). Pour chaque extrait, 2 µl d'ADN génomique sont ajoutés à 200 µl de mélange réactionnel (TNE 1X (Sigma<sup>TM</sup>), Hoeschst à 0,1 µg/ml). Une gamme étalon de 0 à 300 µg/ml est réalisée à partir d'une solution de thymus de veau pour réaliser la quantification. Les échantillons sont ensuite dilués à 3 ng/µl et stockés à -20°C (solution de travail).

### 3. *Caractéristiques des EST et des gènes étudiés*

Les EST utilisées ont été isolées d'une banque suppressive et soustractive où les gènes du xylème d' *E. gunnii* sont préférentiellement exprimés par rapport à la feuille (Paux *et al.*, sous presse) ; elles correspondent en majeure partie à des extrémités 3' UTR de gènes. Le tableau 1 liste les EST et les gènes de fonction utilisés dans cette étude. Certaines EST jouent un rôle directement ou indirectement dans la formation des parois (exemple de l'EST 08E04 codant pour une cellulase), d'autres ont des fonctions liées aux mécanismes de défense de la plante (exemple de l'EST 08G06 qui code pour une protéine induite par les harpins), certaines sont impliquées dans des mécanismes cellulaires plus généraux (exemple de l'EST 09H01 codant pour une MAP kinase), mais une part non négligeable n'a pas de fonction connue. Des amorces spécifiques aux séquences d'*E. gunnii* ont été définies, puis testées chez *E. globulus*.

Les 2 gènes candidats fonctionnels ont été sélectionnés à partir des bases de données publiques. Ils interviennent dans les étapes de différenciation des cellules du xylème :

- ✓ La xyloglucan endo-1,4-beta-D glucanase génère des oligosaccharides de xyloglucane à partir des hémicelluloses. Elle intervient dans le processus d'élongation cellulaire, première étape de la différenciation cellulaire, permettant un renforcement des connections entre les parois primaires et secondaires lors du dépôt de cette dernière. Trois gènes différents ont été clonés sur plusieurs espèces dont *E. gunnii* lors de l'isolation des EST de notre étude (Paux *et al.*, sous presse).
- ✓ L'alpha-tubulin est une protéine qui rentre dans la composition des microtubules. Le complexe microtubulaire oriente les microfibrilles de cellulose à travers les pores cellulaires et définit ainsi la structure des microfibrilles de la paroi primaire. La désorganisation du complexe entraîne un arrêt de l'élongation cellulaire. Ce gène a été cloné chez *E. globulus sp. bicostata* sous le nom EgTubA1.

#### **4. Mise au point et conditions d'amplification PCR**

Pour chaque couple d'amorces, la température d'hybridation proposée par RAIZ a été réajustée sur un Mastercycler Gradient Eppendorf en microplaque de 96 puits. L'ADN génomique, 15 ng pour chaque échantillon, est amplifié par PCR dans un volume total de 25 µl contenant 0,2 µM de chaque amorce, 1 ng de BSA, 200 µM de chaque dNTP, 0,05% de détergent W-1 (Invitrogen), 12,5 µl de tampon 10X stérile [166 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 670 mM Tris HCl pH=8, 20 mM MgCl<sub>2</sub>, 0,7 % mercaptoéthanol] et 1U de Taq DNA polymérase (Invitrogen). Les conditions d'amplification sont les suivantes : une dénaturation initiale de 4 min à 94°C suivie de 35 cycles comprenant une dénaturation (30 secondes (sec) à 94°C), une hybridation (1 min à une température spécifique pour chaque couple d'amorces) et une élongation (1 min à 72 °C), et enfin une élongation finale de 5 min à 72 °C. Les températures d'hybridation spécifiques à chaque couple d'amorces fournies par la compagnie RAIZ ont d'abord été testées, et selon les besoins, des tests de gradient de température ont été réalisés. D'autres tests ont été réalisés sur la composition du mélange réactionnel des PCR : la quantité d'amorces (de 0,1 µM à 0,4 µM), la concentration en MgCl<sub>2</sub> (de 0 mM à 3,5 mM), la nature, la concentration du cation et leur combinaison (KCl à 25 mM ou 50 mM, et NH<sub>4</sub> à 16,6 mM ou 8mM), la provenance du 10X (soit produit au laboratoire, soit fourni avec la Taq DNA polymerase (Invitrogen)) ont été les paramètres variants de ces tests.

Pour chaque produit PCR, la spécificité d'amplification est vérifiée après une électrophorèse sur gel d'agarose à 2% dans du TBE (Tris Borate EDTA) 0,5X et une coloration au BET (bromure d'éthidium).

## 5. Révélation du polymorphisme chez les géniteurs

### Par SSCP

La technique SSCP, développée par Orita *et al.* (1989), est basée sur une différence de conformation secondaire des simples brins d'ADN en conditions non dénaturantes. Ces conformations secondaires sont dues à un appariement des bases entre elles. Ainsi la substitution d'une base au sein du fragment amplifié peut donner un repliement du brin d'ADN différent. En conditions non dénaturantes, ces dissimilitudes de séquences entraînent des différences de migration détectables sur un gel de polyacrylamide.

Pour les analyses SSCP, le protocole utilisé est celui décrit par Bodénès *et al.* (1996). Brièvement, 7 µl de produit PCR sont ajoutés à 7 µl d'une solution contenant 95% de formamide, 10 mM de NaOH, 0,05% de xylene cyanol, et 0,05% de bleu de bromophénol. Les échantillons sont dénaturés par un passage à 94°C pendant 4 min puis immédiatement refroidis dans la glace, empêchant le ré-appariement des deux brins complémentaires, et chargés sur un gel d'acrylamide (acrylamide/bisacrylamide : 37,5/1 ; 0,6X TBE ; 0,75 mm x 16 cm x 18 cm ; peigne 28 dents) non dénaturant pour une électrophorèse à température et voltage constant dans du TBE 0,6X. Pour chaque séquence, les conditions de migration ont été optimisées sur les 2 parents et 10 descendants en faisant varier la composition du gel (pourcentage d'acrylamide, ajout de 5% de glycérol), la température du tampon de cuve, la durée de migration et le voltage.

La coloration des gels s'effectue selon la méthode de Creste *et al.* (2001). Les gels sont fixés dans un bain d'acide acétique 1% et éthanol 10% pendant 10 min, rincés puis prétraités 3 min dans de l'acide nitrique 1,5% et rincés à nouveau. L'étape d'imprégnation a lieu dans une solution d'AgNO<sub>3</sub> 0,2% pendant 20 min. Les gels sont ensuite rincés deux fois puis placés dans deux bains de révélation successifs contenant du Na<sub>2</sub>CO<sub>3</sub> à 30 g/l et du formaldéhyde à 0,02%. La coloration des gels est stoppée dans un bain d'acide acétique à 7,5%. Après un dernier rinçage, les gels sont photographiés et séchés sous cellophane 2 h à 80°C.

## Par PCR-RFLP

Afin de révéler du polymorphisme sur des fragments de taille plus grande, une autre technique a été employée : la PCR-RFLP, Cleaved Amplified Polymorphic Sequence (CAPS). Les fragments amplifiés sont digérés par des enzymes de restriction spécifiques, les produits issus de la digestion sont ensuite visualisés après une électrophorèse sur un gel d'agarose et une coloration au BET. Concrètement, à 10 µl de chaque produit PCR est ajouté 15 µl d'une solution d'un tampon 1X spécifique (fourni avec l'enzyme) contenant 5U de l'enzyme de restriction choisie. La digestion se déroule à température et durée variables en fonction de l'enzyme (protocole fourni à l'achat) puis est arrêtée par un ajout de 4 µl de bleu de charge (2,5 g/l bleu de bromophénol, 50 mM EDTA Na<sub>2</sub>, 2H<sub>2</sub>O, TBE 1X, sodium dodecyl sulfate 0,2%). Les échantillons sont ensuite chargés sur un gel d'acrylamide 8% (acrylamide/bisacrylamide : 29/1 ; 1X TBE ; 1mm x 16cm x 18cm) pour une électrophorèse à température ambiante, voltage constant (200 V) pendant 3h30 dans du TBE 1X et révéler dans un bain de BET.

## Séquençage des produits PCR

La dernière technique employée consiste à séquencer les EST chez les deux géniteurs. Le séquençage de produits PCR bruts donne accès aux séquences des deux allèles l'eucalyptus étant diploïde. Il permet de vérifier la nature du fragment amplifié et donne une approximation de la variation existante entre les 2 allèles d'un même génotype (détection d'évènements mutationnels).

Les séquences concernées par le séquençage sont toutes les EST ayant présenté un profil d'amplification spécifique dans cette étude (selon le protocole d'amplification précédemment décrit), et d'une étude précédente (Carrouché, 2003) ainsi que les deux gènes candidats. Un volume de 30 µl des produits PCR obtenus dans les conditions précédemment décrites a été envoyé à Génome Express (Grenoble, France) qui a réalisé le séquençage. Les chromatogrammes sont ensuite visualisés par le logiciel BIOEDIT.

Pour chaque EST, les séquences parentales sont analysées séparément, chacune d'elles avec celle d'*E. gunnii* (origine de la définition des primers) et enfin un alignement des trois séquences est réalisé avec la fonction ClustalW. La séquence obtenue est un consensus des deux allèles ; le chromatogramme nous permet de visualiser leur présence simultanée ou non, et de déterminer leurs différences. Compte tenu que nous comparons du génome exprimé (*E. gunnii*) et du génome complet (*E. urophylla* et *E. grandis*), nous nous attendons à détecter des régions introniques pour les deux géniteurs de notre étude.

## 6. Cartographie génétique

Les cartes construites à partir de ce croisement (Verhaegen et Plomion, 1996 ; Verhaegen *et al.*, 1997 ; Gion *et al.*, 2000 ; Carrouché 2003) reposent principalement sur des marqueurs dominants de type RAPD. Les marqueurs SSCP étant des marqueurs codominants, il est donc nécessaire de les coder comme des marqueurs dominants. En conséquence, pour les deux allèles d'un gène qui ségrégent dans la descendance, la présence de l'une est notée 1 tandis que la présence de l'autre est notée 0, le choix étant arbitraire mais fixé pour une même séquence.

Si un seul des gènes est hétérozygote pour une séquence donnée, la ségrégation des allèles dans la descendance doit correspondre à une ségrégation mendélienne de type 1 : 1. cette EST ne sera positionnée que chez le gène polymorphe. Lorsque les deux gènes montrent du polymorphisme, la ségrégation mendélienne attendue est de type 1 : 1 : 1 : 1. L'EST sera alors cartographiée indépendamment chez les deux gènes et permettra de distinguer des groupes de synténie entre les deux espèces.

Afin de vérifier que les ségrégations observées soient conformes à des ségrégations mendéliennes, un test du  $\chi^2$  ( $\alpha = 5\%$ ) a été réalisé.

Deux marqueurs sont dits génétiquement liés lorsqu'ils coségrégent. Les marqueurs sont donc soumis deux à deux à un test d'indépendance de ségrégation ou test de liaison, pour lequel l'hypothèse nulle correspond à une absence de liaison ( $H_0 : \theta = 0.5$ ) où  $\theta$  représente le taux de recombinaison dont la valeur varie entre 0 et 0,5. Le test de liaison est réalisé à partir du calcul du rapport de vraisemblance (LOD score) défini par :

$$\text{LOD} = \log_{10} [L(\theta) / L(1/2)]$$

$L(\theta)$  représente la vraisemblance d'un taux de recombinaison compris entre 0 et 0,5, et  $L(1/2)$ , la vraisemblance d'une absence de liaison entre les 2 marqueurs. Pour positionner les gènes et les EST sur les groupes de liaison, un LOD score seuil de 4 correspondant à une probabilité de  $10^{-4}$  de se tromper et un  $\theta$  maximum de 0,4 ont été choisis.

Au moyen du logiciel MAPMAKER 2.0 (Lander *et al.*, 1987) nous avons défini des groupes de liaison avec les valeurs de  $\theta$  et de LOD score définies précédemment, puis teste l'ordre des marqueurs au sein d'un groupe par la fonction d'ordonnement. La fonction *Lod table* permet de visualiser la matrice des LOD score et des  $\theta$  rendant compte de la force de liaison entre les marqueurs pris deux à deux. Parallèlement, la fonction *ripple* du logiciel est utilisée, elle permet de définir l'ordre le plus probable des marqueurs pris trois à trois. Finalement, le logiciel convertit les pourcentages de recombinaison en distances génétiques par la fonction *Kosambi* (Kosambi, 1944) et forme une carte génétique à partir de l'ordre optimum des marqueurs choisi.



Figure 3 : Profils électrophorétiques sur gel d'agarose à 2%, pour un gradient de température réalisé sur les 3 groupes à tester, à savoir le groupe A, B et C. L'EST 08107 est une phase d'optimisation à été nécessaire. Les températures d'hybridation de 49°C à 57°C avec une température optimale d'hybridation de 53°C.



À l'aide du logiciel MAPMAKER 2.0 (Lander et al., 1987) nous avons défini des groupes de marqueurs en fonction des valeurs de  $\theta$  et de LOD scores définies précédemment, puis nous avons ordonné les marqueurs en fonction de leur position sur le chromosome. La fonction LOD score permet de calculer la probabilité de la présence d'un gène dans une région donnée. Plus la valeur de LOD score est élevée, plus la probabilité est grande. Par conséquent, la fonction LOD score permet de sélectionner les marqueurs les plus probables des marqueurs qui sont à tester. Ensuite, elle permet de sélectionner les marqueurs les plus probables des marqueurs qui sont à tester. Finalement, le logiciel MAPMAKER 2.0 permet de calculer les pourcentages de recombinaison en distance génétique par la fonction Kosambi (Kosambi, 1983) et de faire une carte génétique à partir de l'ordre optimal des marqueurs choisis.

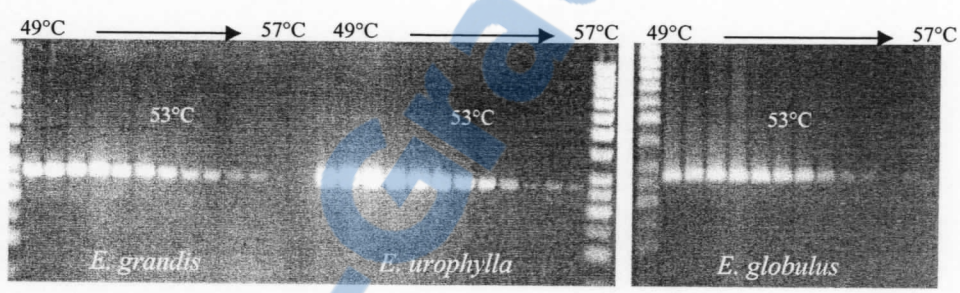


Figure 2 : Profils électrophorétiques sur gel d'agarose à 2%, pour un gradient de température réalisé sur les 3 génotypes *E. grandis*, *E. urophylla* et *E. globulus*, de l'EST 08H07 où une phase d'optimisation a été nécessaire. Les températures s'échelonnent de 49°C à 57°C avec une température optimale d'hybridation de 53°C.

Rapport

### 3. Résultats

#### 1. Amplification des EST et des gènes

Lors de l'optimisation des températures d'hybridation, tous les couples d'amorces ont été testés sur trois espèces différentes : *E. urophylla* (individu 14.144) et *E. grandis* (individu 9.21), les deux géniteurs, et *E. globulus* (individu CN 32 L6 de RAIZ) servant de témoin et d'élément de comparaison. Cette opération est nécessaire dans la mesure où les conditions de laboratoire (thermocycleur, composition du mélange réactionnel) sont différentes tout comme les espèces et les génotypes. La figure 2 montre les profils électrophorétiques de l'EST 08H07 sur les 3 génotypes où une optimisation a été nécessaire ; en effet la température d'hybridation optimale proposée était de 58°C (tableau 2) mais un gradient de température allant de 49°C à 57°C nous a montré qu'à 53°C l'amplification était de meilleure qualité. Cette température permet l'amplification d'un fragment de 250-300pb pour les 3 espèces correspondant à la taille obtenue (294pb) sur la séquence d'*E. gunnii*.

Sur les 33 EST testées, 29 couples d'amorces ont donné un produit d'amplification spécifique (une seule bande sur gel d'agarose) sur les 3 espèces testées (tableau 2). Quelque soit les conditions, l'EST 08A09 montre un profil multi bandes sur *E. grandis* et a donc été écartée de l'étude. Les 2 derniers couples d'amorces, 08G09 et 08H01, n'ont donné aucune amplification quelle que soit l'espèce. Une mutation a pu s'insérer sur le site de fixation des amorces ; en effet les EST correspondent, pour beaucoup, à des régions 3' UTR, régions les moins conservées des gènes (Paux *et al.*, sous presse). Parmi ces 30 produits d'amplification à bande unique pour les 2 géniteurs, 23 ont une taille observée identique à la taille attendue par rapport à la séquence d'*E. gunnii*, 5 autres ont une taille supérieure à la taille attendue laissant supposé l'amplification d'un intron et une EST, 09D01, a une taille inférieure à la taille attendue laissant supposer un problème de définition d'amorces.

Enfin pour les 2 gènes testés, un produit d'amplification unique a été obtenu sur les 3 génotypes testés. Le premier gène, codé par GF#27, a donné un produit d'amplification conforme à la taille attendue, par contre le second, codé par GF#86, a donné un produit d'amplification d'environ 400pb (taille attendue 352pb).

Tableau 2 : Conditions d'amplification des EST, et des gènes Tailles des fragments observées pour les deux laboratoires. Un \* signale qu'un intron a été amplifié. Un \*\* indique que la taille observée est significativement plus petite que la taille attendue. Echec désigne les EST pour lesquelles aucune amplification n'a été obtenue.

| nom séquence. | conditions d'amplification de RAIZ |                   |                   | conditions d'amplifications du CIRAD |                           |
|---------------|------------------------------------|-------------------|-------------------|--------------------------------------|---------------------------|
|               | température d'hybridation          | tailles attendues | tailles observées | tailles observées                    | température d'hybridation |
| 08A09         | 45                                 | 217               | 200<L<250         | 200<L<250                            | 50                        |
| 08B02         | 49                                 | 224               | 200<L<250         | 200<L<250                            | 49                        |
| 08B06         | 58                                 | 326               | 300<L<350         | 300<L<350                            | 58                        |
| 08C02         | 58                                 | 217               | 200<L<250         | 250<L<300                            | 60                        |
| 08C05         | 58                                 | 347               | 850<L<1000*       | 900<L<1000*                          | 58                        |
| 08C08         | 58                                 | 312               | 300<L<350         | 350<L<400                            | 58                        |
| 08C10         | 58                                 | 259               | 250<L<300         | 250<L<300                            | 58                        |
| 08E04         | 58                                 | 236               | 200<L<250         | 200<L<250                            | 58                        |
| 08E06         | 58                                 | 248               | 200<L<250         | 200<L<250                            | 58                        |
| 08E11         | 58                                 | 216               | 200<L<250         | 300<L<350*                           | 58                        |
| 08E12         | échec                              | 217               | échec             | 200<L<250                            | 47                        |
| 08F01         | 58                                 | 128               | 100<L<150         | 100<L<150                            | 58                        |
| 08F03         | 58                                 | 358               | 350<L<400         | 350<L<400                            | 62                        |
| 08F09         | 58                                 | 255               | 250<L<300         | 250<L<300                            | 58                        |
| 08G01         | 58                                 | 250               | 250<L<300         | 250<L<300                            | 62                        |
| 08G06         | 58                                 | 210               | 200<L<250         | 200<L<250                            | 58                        |
| 08G08         | échec                              | 121               | échec             | 100<L<150                            | 58                        |
| 08G09         | échec                              | 279               | échec             | échec                                | échec                     |
| 08H01         | 55                                 | 127               | 100<L<150         | échec                                | échec                     |
| 08H04         | 58                                 | 293               | 250<L<300         | 250<L<300                            | 58                        |
| 08H07         | 58                                 | 294               | 250<L<300         | 250<L<300                            | 53                        |
| 08H11         | 49                                 | 238               | 200<L<250         | 350<L<400*                           | 58                        |
| 09A11         | 58                                 | 104               | 1600<L<1700*      | 1200<L<1500*                         | 47                        |
| 09C05         | 58                                 | 169               | 150<L<200         | 150<L<200                            | 58                        |
| 09C10         | 58                                 | 265               | 250<L<300         | 250<L<300                            | 58                        |
| 09D01         | 62                                 | 316               | 300<L<350         | 100<L<150**                          | 58                        |
| 09D03         | 58                                 | 273               | 250<L<300         | 250<L<300                            | 58                        |
| 08D04         | 49                                 | 112               | 100<L<150         | 100<L<150                            | 49                        |
| 09F08         | 58                                 | 332               | 300<L<350         | 300<L<350                            | 58                        |
| 09H01         | 54                                 | 323               | 300<L<350         | 400<L<450*                           | 56                        |
| 09H02         | 58                                 | 210               | 200<L<250         | 200<L<250                            | 54                        |
| 09H05         | 58                                 | 228               | 200<L<250         | 250<L<300                            | 53                        |
| 09H06         | 58                                 | 278               | 250<L<300         | 250<L<300                            | 58                        |
| GF#27         | 58                                 | 434               | 400<L<450         | 400<L<450                            | 58                        |
| GF#86         | 58                                 | 352               | 350<L<400         | 400<L<450                            | 58                        |

D'une manière générale, les résultats obtenus dans cette étude confirment les résultats obtenus sur *E. globulus*. On peut cependant noter que pour 2 des 3 EST (08E12 et 08G08) n'ayant pu être amplifiées par RAIZ donnent ici des résultats positifs. L'EST 08H01 n'a pas été amplifiée ici. De plus, 5 EST (08E11, 08H11, 09A11, 09D01 et 09H01) donnent un produit d'amplification de taille différente par rapport à celle attendue.

L'amplification de fragments d'ADN est également sensible à d'autres variables. Afin d'optimiser au mieux les amplifications, différents tests ont été réalisés pour réduire les profils multi bandes d'une part et d'autre part améliorer l'efficacité de l'amplification pour soit des EST ayant un signal d'amplification faible mais unique sur agarose, soit des EST impossible à amplifier. Les différentes conditions (cf. matériels et méthodes) ont été testées sur plusieurs EST sans résultat probant (résultats non présentés).

## 2. Détection du polymorphisme

### Par SSCP

Différentes conditions SSCP ont été testées pour optimiser la révélation du polymorphisme au sein des séquences étudiées. Cette phase d'optimisation est nécessaire car la SSCP est sensible à différents paramètres qu'il est possible de faire varier : la température de migration, la présence ou l'absence de glycérol dans le gel, le temps de migration, le voltage et la concentration en acrylamide. Selon Bodénès *et al.* (1996) et Plomion *et al.* (1999) ces facteurs sont les principaux influant sur la qualité de la résolution des différentes bandes. Un autre facteur relevé par Gion *et al.* (2000) a montré une influence notable due au chargement des échantillons en présence ou en l'absence de tampon dans les puits. A titre d'exemple la figure 3 montre les profils SSCP de l'EST 08F09 chargés en présence (D) ou en absence (C) de tampon de migration. Le chargement en présence du tampon fait apparaître un nombre de bandes plus élevé qu'en son absence, chose qui serait liée à plusieurs conformations simples brins semi stables (Gion *et al.*, 2000).

Tableau 3 : Conditions SSCP utilisées pour chaque séquence après optimisation et polymorphisme observé pour chaque géniteur. P indique que la séquence est polymorphe pour le géniteur considéré, M qu'elle est monomorphe.

| EST /<br>gènes<br>id. | température<br>de<br>migration | concentration<br>en acrylamide | glycérol | voltage<br>(V) | durée de<br>migration | <i>E.</i><br><i>grandis</i><br>9.21 | <i>E.</i><br><i>urophylla</i><br>14.144 |
|-----------------------|--------------------------------|--------------------------------|----------|----------------|-----------------------|-------------------------------------|---|
| 08B02                 | 15°C                           | 12%                            | 5%       | 150            | 20h00                 | P                                   | P                                       |
| 08C02                 | 15°C                           | 10%                            | -        | 250            | 7h00                  | M                                   | P                                       |
| 08C08                 | 15°C                           | 12%                            | -        | 200            | 24h00                 | P                                   | P                                       |
| 08E04                 | 15°C                           | 10%                            | -        | 250            | 7h00                  | P                                   | P                                       |
| 08E12                 | 15°C                           | 12%                            | 5%       | 150            | 20h00                 | P                                   | P                                       |
| 08F03                 | 15°C                           | 10%                            | -        | 150            | 20h00                 | P                                   | P                                       |
| 08F09                 | 15°C                           | 10%                            | -        | 250            | 7h00                  | P                                   | M                                       |
| 08G01                 | 15°C                           | 12%                            | 5%       | 200            | 21h30                 | M                                   | P                                       |
| 08H04                 | 15°C                           | 10%                            | -        | 250            | 16h00                 | P                                   | M                                       |
| 08H07                 | 15°C                           | 12%                            | 5%       | 150            | 21h30                 | P                                   | P                                       |
| 08H11                 | 15°C                           | 12%                            | -        | 200            | 23h00                 | M                                   | P                                       |
| 09C05                 | 15°C                           | 10%                            | -        | 200            | 5h00                  | P                                   | M                                       |
| 09D01                 | 15°C                           | 10%                            | -        | 200            | 5h00                  | P                                   | M                                       |
| 09D03                 | 15°C                           | 12%                            | -        | 200            | 15h00                 | M                                   | P                                       |
| 09F08                 | 15°C                           | 12%                            | -        | 150            | 17h00                 | M                                   | P                                       |
| 09H02                 | 15°C                           | 10%                            | -        | 250            | 7h00                  | M                                   | P                                       |
| 09H06                 | 15°C                           | 10%                            | -        | 250            | 8h00                  | P                                   | P                                       |
| GF#27                 | 15°C                           | 8%                             | -        | 250            | 20h00                 | M                                   | P                                       |
| GF#86                 | 15°C                           | 12%                            | -        | 200            | 24h00                 | P                                   | P                                       |

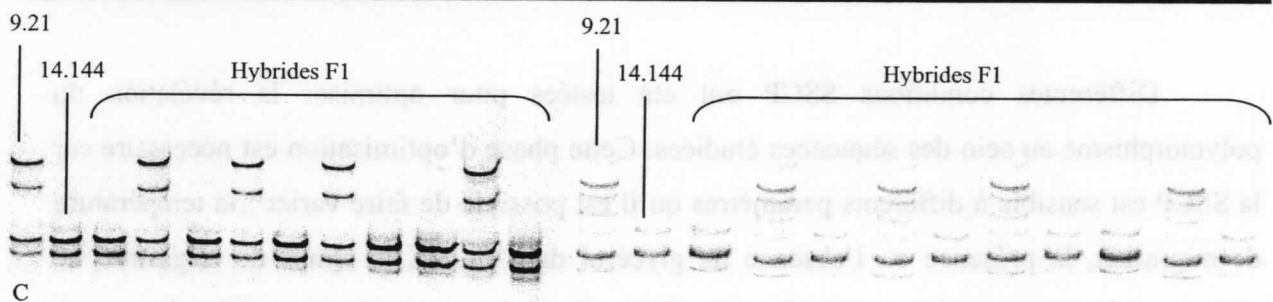


Figure 3 : Comparaison des profils SSCP de 08F09 après chargement des produits d'amplification sur gel en absence (C) ou en présence (D) de tampon de migration. Les deux premières pistes correspondent aux profils des géniteurs, les autres aux hybrides. Les conditions de migration sont identiques dans les deux cas : gel à 10% d'acrylamide, migration 7h00 à 250V et 15°C.

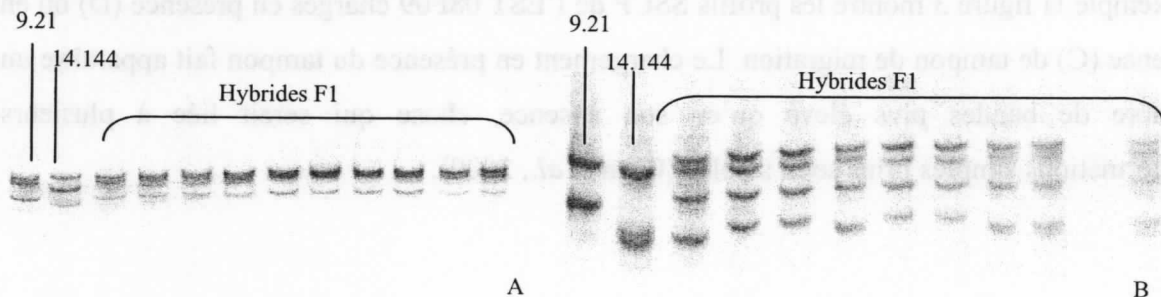


Figure 4 : Comparaison des profils SSCP de 08G01 avant (A) et après (B) optimisation des conditions SSCP. Les deux premières pistes correspondent aux profils parentaux, les autres à la descendance. Conditions pour A : gel à 10% d'acrylamide, migration 7h00 à 250V et 15°C ; conditions pour B : gel à 12% d'acrylamide + 5% de glycérol, migration 21h30 à 200V et 15°C. La distance de migration est identique pour les deux conditions

Dans une étude précédente, Carrouché, (2003) propose une température de migration unique de 15°C que nous avons conservé. Par contre, selon la taille des fragments amplifiés, la durée de migration et le voltage ont été adaptés. L'utilisation de glycérol à 5 % incorporé au gel d'acrylamide a permis, dans certains cas, une meilleure distinction des profils. Glavac et Dean (1993) ont démontré que la concentration en acrylamide influençait la résolution des gels. Les électrophorèses ont donc été réalisées avec des gels de trois concentrations différentes : 8, 10 et 12% et nous avons noté que l'augmentation de la concentration permet une meilleure séparation des bandes pour un même profil. Par conséquent, des conditions de migration spécifiques ont été définies pour chacune des EST. La figure 4 montre deux profils SSCP pour l'EST 08G01 avant (A) et après (B) optimisation des conditions SSCP. Les premières conditions testées (température de migration de 15°C, 10% d'acrylamide, 7h00 à 250V) ne permettent pas de déceler du polymorphisme pour cette EST alors que les secondes (température de migration de 15°C, 12% d'acrylamide, 5% de glycérol, 21h30 à 200V) révèlent un polymorphisme maternel.

Les conditions SSCP mises au point ont permis de révéler du polymorphisme chez 17 EST ainsi que chez les deux gènes (tableau 3). Sur ces 17 EST, 7 sont polymorphes pour les deux géniteurs, 6 montrent un polymorphisme maternel et 4 un polymorphisme paternel. Le gène candidat GF#86 est polymorphe chez les deux parents tandis que GF#27 est polymorphe uniquement chez le géniteur femelle. Le polymorphisme semble plus important chez *E. urophylla* que chez *E. grandis* dans ce croisement. Les douze EST restante n'ont pas montré de polymorphisme quelles que soient les conditions SSCP testées.

La SSCP est une technique simple, sensible (une mutation ponctuelle est visible) et peu coûteuse pour révéler du polymorphisme au sein de fragments PCR de tailles différentes. Néanmoins, la phase d'optimisation des conditions de migration est longue et nécessite de nombreux tests. Mais les résultats de cette étude et ceux de Carrouché (2003) permettent de diriger plus efficacement la mise au point des conditions. Toutefois Sheffield *et al.* (1993) ont démontré que la taille du fragment limitait la sensibilité de la SSCP aux mutations ponctuelles ; c'est la raison pour laquelle une autre technique a été envisagée pour détecter du polymorphisme.

Tableau 4 : Enzymes de restriction utilisées pour digérer l'EST 09A11, leur site de restriction et le polymorphisme observé pour les deux géniteurs.

| Enzyme   | Site de restriction                               | Révélation du polymorphisme |                     |
|----------|---|-----------------------------|---------------------|
|          |   | <i>E. grandis</i>           | <i>E. urophylla</i> |
| Eco0109  | 5'... PuG ↓GNCCPy ...3'<br>3'... PuGGNC↑CPy ...5' | oui                         | oui                 |
| Hind III | 5'... A↓AGCTT ...3'<br>3'... AAGCT↑T ...5'        | non                         | oui                 |
| Hinf I   | 5'... G↓ANTC ...3'<br>3'... GANT↑C ...5'          | non                         | oui                 |
| Hha I    | 5'... G↓CGC ...3'<br>3'... GCG↑C ...5'            | non                         | non                 |

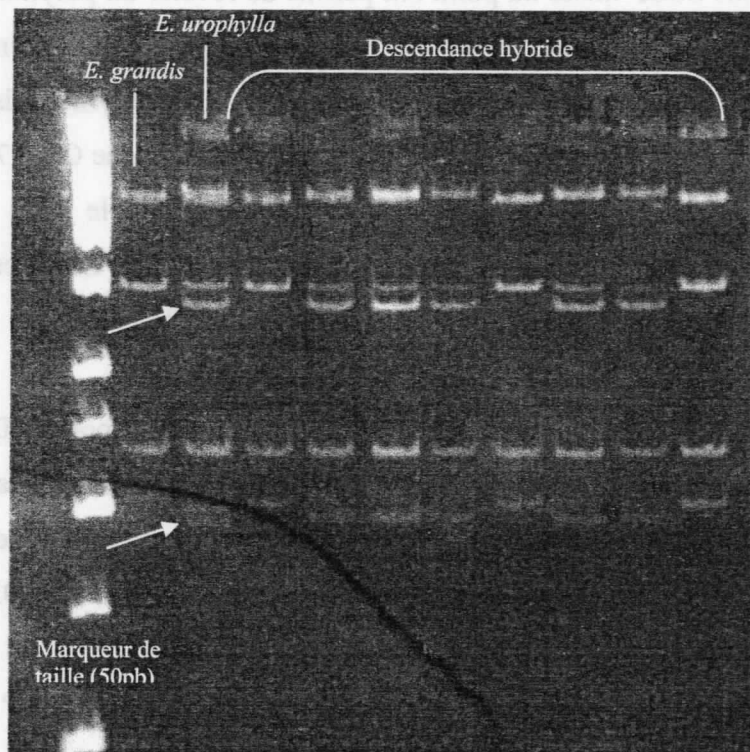


Figure 5 : Profils CAPS de l'EST 09A11 digérée par Eco0109. Les flèches indiquent les bandes ségrégant dans la descendance. Les deux premières pistes correspondent aux géniteurs, les autres aux hybrides.

## Par utilisation d'enzymes de restriction

Cette technique a été abordée uniquement sur l'EST 09A11, la taille du fragment étant de 1200/1500 pb. Si l'on suppose la répartition des 4 bases équiprobables, une enzyme ayant un site de restriction de 4 pb (Mbo I, Msp I, Alu I, Taq I....) coupera l'ADN toutes les 256 bases en moyenne ( $4^4$ ). Le fragment cible peut donc produire environ 5 fragments de taille diverse ; la spécificité du site de restriction peut entraîner un polymorphisme de longueur de fragments dans la descendance si au moins un des deux géniteurs est hétérozygote pour le croisement.

Différentes enzymes ont été choisies en fonction de leur site de restriction pour digérer les produits d'amplification parentaux et ceux de la descendance (tableau 4). Trois enzymes révèlent du polymorphisme chez le géniteur femelle uniquement (Hind III, Hinf I et Eco0109), la figure 5 nous montre les profils digérés par Eco0109 sur les deux géniteurs et 8 descendants. On peut clairement voir plusieurs bandes chez *E. urophylla* ségrégant dans la descendance (notées par les flèches), les bandes d'*E. grandis* ne montrent aucune bande ségrégant clairement. D'autres tests sont à réaliser pour affirmer si le géniteur mâle est polymorphe ou monomorphe. Une enzyme de restriction (Hha I) n'a pas montré de polymorphisme dans ce fragment PCR.

Il s'agissait ici de chercher une voie alternative à la SSCP en conséquence la descendance complète du croisement n'a pas été amplifiée pour réaliser la cartographie de cette EST. Il serait également intéressant d'utiliser cette technique sur les EST n'ayant montré aucun polymorphisme avec la SSCP.

## Séquençage des produits d'amplification

Dans un premier temps, un tri des séquences a été réalisé en fonction de la lisibilité des chromatogrammes. Celles dont les chromatogrammes étaient de mauvaise qualité (pic chaotique et irrégulier, taille de séquence incohérente, faible signal mélangé au bruit de fond...) ont été éliminées. Le séquençage a permis de produire 67 séquences sur les 96 envoyées au départ, soit un taux de 70% de séquences exploitables (tableau 5). Parmi les EST écartées de cette analyse, du polymorphisme a été mis en évidence chez les géniteurs pour 64% d'entre elles, ce qui laisse supposer une variabilité trop importante entre les deux allèles.



Tableau 5 : Caractéristiques des variations observées sur les séquences.

Les données suivantes sont présentées : pourcentage d'homologies entre espèces, insertions-délétions (INDEL) détectées au sein d'un génotype, nombre de SNP détectés à l'état hétérozygote pour chaque géniteur, nombre de mutations (SNP) détectées à l'état homozygote et nombre d'insertions-délétions détectées à l'état homozygote. A : absence, P : présence, - : détection impossible.

| séquence id. | taille du fragment | pourcentage d'homologies (%)    |                                |                              | INDEL intra-spécifique |        | Nombre de SNP détectés |        | Nombre de mutations détectées entre génotypes | Nombre d'INDEL interspécifiques détectées entre génotypes |
|--------------|--------------------|---------------------------------|--------------------------------|------------------------------|------------------------|--------|------------------------|--------|---|---|
|              |                    | <i>E. urophylla/ E. grandis</i> | <i>E. urophylla/ E. gunnii</i> | <i>E. grandis/ E. gunnii</i> | 9.21                   | 14.144 | 9.21                   | 14.144 |   |   |
| 06A01        | 350/400            | 99                              | 91                             | 91                           | A                      | A      | 1                      | 1      | 1   | 0   |
| 06A07        | 200/250            | 62                              | 69                             | 56                           | P                      | P      | -                      | -      | 0   | 1   |
| 06A08        | 350/400            | 99                              | 90                             | 88                           | A                      | A      | 0                      | 0      | 4   | 0   |
| 06A11        | 300/350            | 63                              | 62                             | 45                           | P                      | A      | -                      | 1      | 1   | 0   |
| 06A12        | 150                | 89                              | 72                             | 73                           | A                      | A      | 1                      | 2      | 0   | 0   |
| 06B07        | 100/150            | 97                              | 77                             | 75                           | A                      | A      | 0                      | 0      | 0   | 1   |
| 06C08        | 100/150            | 100                             | 65                             | 65                           | P                      | P      | -                      | -      | 0   | 0   |
| 06D01        | 300                | 68                              | 62                             | 88                           | A                      | P      | 0                      | 0      | 0   | 0   |
| 06D02        | 200                | 96                              | 75                             | 75                           | A                      | A      | 0                      | 0      | 0   | 0   |
| 06D03        | 150/200            | 82                              | 66                             | 75                           | A                      | A      | 0                      | 1      | 1   | 0   |
| 06D07        | 550*               | 96                              | 25                             | 26                           | A                      | A      | 2                      | 4      | 0   | 1   |
| 06D12        | 700/800*           | 80                              | 17                             | 18                           | A                      | A      | 1                      | 1      | 1   | 2   |
| 06F06        | 200/250            | 98                              | 85                             | 83                           | A                      | A      | 0                      | 0      | 0   | 0   |
| 06H12        | 150                | 87                              | 71                             | 76                           | A                      | A      | 0                      | 0      | 0   | 1   |
| 07A01        | 300/400            | 66                              | 88                             | 62                           | P                      | A      | -                      | 0      | -   | 0   |
| 08A09        | 200/250            | -                               | 78                             | -                            | -                      | A      | -                      | 2      | -   | -   |
| 08B02        | 200/250            | 52                              | 57                             | 65                           | P                      | A      | -                      | 2      | -   | 0   |
| 08B06        | 300/350            | 94                              | 66                             | 64                           | A                      | A      | 2                      | 0      | 0   | 1   |
| 08C02        | 250/300*           | 34                              | 45                             | 32                           | A                      | A      | -                      | -      | -   | 1   |
| 08C10        | 250/300            | 49                              | 54                             | 43                           | A                      | A      | -                      | -      | -   | 0   |
| 08D04        | 100/150            | 89                              | 52                             | 56                           | A                      | A      | 1                      | 1      | 3   | 1   |
| 08E04        | 200/250            | 97                              | 82                             | 84                           | A                      | P      | 0                      | 1      | 1   | 0   |
| 08E11        | 300/350*           | 88                              | 53                             | 55                           | A                      | A      | 0                      | 1      | 4   | 3   |
| 08F01        | 100/150            | 38                              | 71                             | 47                           | A                      | A      | 1                      | 0      | 0   | 3   |
| 08F09        | 250/300            | 96                              | 85                             | 86                           | A                      | A      | 1                      | 1      | 1   | 0   |
| 08G01        | 250/300            | 94                              | 87                             | 87                           | A                      | A      | 0                      | 2      | 1   | 2   |
| 08G08        | 100/150            | 88                              | 75                             | 80                           | A                      | A      | 0                      | 0      | 2   | 0   |
| 08H07        | 250/300            | 90                              | 66                             | 69                           | A                      | P      | 0                      | -      | 0   | 4   |
| 09C10        | 250/300            | 91                              | 76                             | 76                           | A                      | A      | 0                      | 0      | 2   | 0   |
| 09D01        | 100/150            | 95                              | 39                             | 40                           | A                      | A      | 1                      | 0      | -   | 1   |
| 09D03        | 250/300            | 80                              | 66                             | 69                           | A                      | P      | 1                      | 1      | 3   | 0   |
| 09H02        | 200/250            | 94                              | 79                             | 80                           | A                      | A      | 0                      | 0      | 3   | 2   |
| GF#27        | 400/450            | 92                              | 88                             | 83                           | A                      | A      | 1                      | 1      | 2   | 0   |
| GF#86        | 400/450*           | 93                              | 70                             | 74                           | A                      | A      | 2                      | 3      | 2   | 0   |

Pour les séquences dont le chromatogramme est illisible, il est possible d'envisager un séquençage à partir de l'amorce antisens, de purifier l'ADN avant le sous traitement du séquençage ou encore de cloner les deux brins d'ADN donnant alors accès aux séquences respectives des deux allèles.

D'une manière générale, nous observons des homologies plus fortes entre les séquences des deux géniteurs par rapport à celles observées avec *E. gunnii*. Pour la plupart des EST, une zone de 20 à 30 pb en début de séquence d'*E. gunnii* est inexistante chez *E. grandis* et d'*E. urophylla*. Cette séquence ayant été prise en compte dans l'alignement des fragments et le calcul du pourcentage d'homologies, le pourcentage d'homologies s'en trouve affecté. A titre d'exemple, si l'on ôte les 22 premières bases de la séquence d'*E. gunnii* et que l'on refait un alignement avec chaque géniteur, les pourcentages d'homologies augmentent de 11% en moyenne.

L'alignement des trois séquences nous permet de confirmer la présence d'introns dans les séquences dont la taille observée est supérieure à la taille attendue (06D07, 06D12 et 08E11).

L'EST 09D01 dont la taille observée était inférieure à la taille attendue lors de l'amplification présente 39% d'homologies entre *E. urophylla* et *E. gunnii* et 40% entre *E. grandis* et *E. gunnii*. Cette homologie représente un domaine conservé d'une centaine de paires de bases en fin de séquence. Nous pouvons donc supposer que nous avons amplifié un domaine protéique d'une trentaine d'acides aminés identique à celui codé par une partie de l'EST.

Pour toutes les EST où le chromatogramme est de bonne qualité, le séquençage permet de confirmer la nature des produits d'amplification chez *E. urophylla* et *E. grandis* avec celle d'*E. gunnii*.

Différents types de variations nucléotidiques ont pu être mis en évidence (tableau 5) : un motif microsatellite, 5 INDEL (insertions délétions) chez le géniteur femelle, *E. urophylla* 4 INDEL chez le géniteur mâle, *E. grandis* (figure 1a, annexe 2), et 22 INDEL observées lors de la comparaison des trois génotypes (figure 1b, annexe 2). Parallèlement, 25 et 15 SNP ont été détectés respectivement chez *E. urophylla* et *E. grandis* (figure 1d, annexe 2) ainsi que 32 mutations chez au moins un des trois génotypes (figure 1c, annexe 2).

Tableau 6 : Interprétation génétique des profils SSCP pour les 19 séquences polymorphes et test de conformité à une ségrégation mendélienne par un test du  $\chi^2$  à 5%. (a) valeur de  $\chi^2$  observée pour chaque séquence. Pour P=95%, la valeur seuil à 1 degré de liberté (ddl) est de 3,84, 7,81 pour 3 ddl et 14,1 pour 7 ddl. \* indique que la valeur est significativement différente au seuil de 5% mais est acceptée au seuil de 1%. \*\* indique que la valeur est significativement différente aux seuils de 5 et 1%. Un - indique un allèle nul.

| EST   | Génotypes parentaux<br><i>E. urophylla</i> / <i>E. grandis</i> | Génotypes des hybrides                                       | Effectifs observés pour chaque génotype hybride | $\chi^2$ (a) | ddl |
|-------|--|--|---|--------------|-----|
| 08B02 | (12 ;34)/(56 ;78)  | (12 ;56)/(12 ;78)/(34 ;56)/(34 ;78)                          | 40/47/44/43                                     | 0,57         | 3   |
| 08C02 | (12 ;12)/(34 ;56)  | (12 ;34)/(12 ;56)  | 95/79   | 1,47         | 1   |
| 08C08 | (12 ;34)/(56 ;78)  | (12 ;56)/(12 ;78)/(34 ;56)/(34 ;78)                          | 27/47/34/67                                     | 21,18**      | 3   |
| 08E04 | (12 ;--)/(56 ;78)  | (12 ;56)/(12 ;78)/(-- ;56)/(-- ;78)                          | 41/48/28/57                                     | 10,3*        | 3   |
| 08F03 | (12 ;34)/(56 ;78)  | (12 ;56)/(12 ;78)/(34 ;56)/(34 ;78)                          | 38/55/47/38                                     | 4,52         | 3   |
| 08F09 | (12 ;34)/(34 ;34)  | (12 ;34)/(34 ;34)  | 87/86   | $5.10^{-3}$  | 1   |
| 08G01 | (12 ;12)/(34 ;56)  | (12 ;34)/(12 ;56)  | 96/80   | 1,45         | 1   |
| 08H04 | (12 ;34)/(34 ;34)  | (12 ;34)/(34 ;34)  | 70/104  | 6,64*        | 1   |
| 08H07 | (12 ;--)/(56 ;78)  | (12 ;56)/(12 ;78)/(-- ;56)/(-- ;78)                          | 41/44/57/29                                     | 9,28*        | 3   |
| 08H11 | (12 ;12)/(34 ;56)  | (12 ;34)/(12 ;56)  | 87/88   | $5.10^{-3}$  | 1   |
| 09C05 | (12 ;34)/(34 ;34)  | (12 ;34)/(34 ;34)  | 92/82   | 0,57         | 1   |
| 09D01 | (12 ;34)/(34 ;34)  | (12 ;34)/(34 ;34)  | 78/94   | 1,5          | 1   |
| 09D03 | (12 ;12)/(12 ;34)  | (12 ;12)/(12 ;34)  | 85/85   | $10^{-14}$   | 1   |
| 09F08 | (12 ;12)/(12 ;34)  | (12 ;12)/(12 ;34)  | 97/77   | 2,3          | 1   |
| 09H02 | (12 ;12)/(34 ;--)  | (12 ;34)/(12 ;--)  | 98/76   | 2,8          | 1   |
| 09H06 | (12 ;34)/(56 ;78)  | (12 ;56)/(12 ;78)/(34 ;56)/(34 ;78)                          | 46/40/39/53                                     | 2,8          | 3   |
| GF27  | (12 ;12)/(12 ;34)  | (12 ;12)/(12 ;34)  | 82/88   | 0,21         | 1   |
| GF86  | (12 ;34)/(56 ;78)  | (12 ;56)/(12 ;78)/(34 ;56)/(34 ;78)                          | 47/42/43/43                                     | 0,34         | 3   |
| 08E12 | (a ; b ; c ; d) /<br>(e ; f ; g ; -)                           | (aceg)/(acfg)/(bceg)/(bcfg)/(adeg) /<br>(adfg)/(bdeg)/(bdfg) | 27/27/20/16/18/28/23/14                         | 9,65         | 7   |

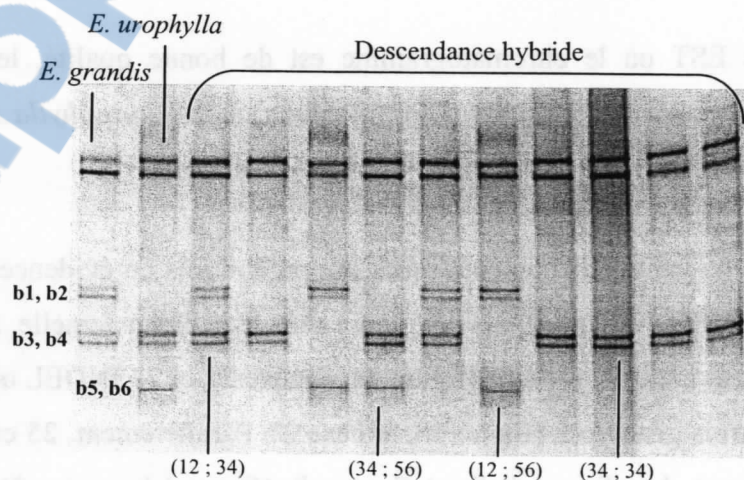


Figure 6 : Révélation du polymorphisme pour GF#86. Le géniteur mâle *E. grandis* est (12 ; 34) et le géniteur femelle *E. urophylla* est (34 ; 56). Quatre classes de descendants visibles : (12 ; 34), (34 ; 56), (12 ; 56) et (34 ; 34).

La ségrégation de l'EST 08E12 dans la descendance hybride suggère l'existence de deux gènes amplifiés dont un est homozygote pour le géniteur *E. urophylla*. Cette ségrégation est conforme à une ségrégation mendélienne de deux gènes au seuil de 5%. Les amorces ont été définies pour être spécifiques de l'EST visée, de plus le profil électrophorétique sur agarose ne montre qu'une seule bande amplifiée, nous pouvons donc émettre l'hypothèse que nous avons amplifié deux gènes d'une même famille. Pour cette EST, le code des combinaisons alléliques a été modifié pour plus de compréhension ainsi les deux bandes représentant un allèle sont regroupées en une lettre. Par exemple, les bandes B1 et B2 caractérisant la présence d'un allèle paternel sont regroupées sous la lettre « a ».

Plomion *et al.* (1999) et Bodénès *et al.* (1996) relèvent que pour les eucalyptus, individu diploïde, un maximum de 4 bandes doit être observé en SSCP chez un individu hétérozygote. Dans cette étude, plusieurs EST ont montré un profil plus complexe avec un nombre de bandes plus élevé rendant difficile l'interprétation des gels (profil multi bandes). Ce résultat suggère l'existence de plusieurs conformations stables des simples brins capables de générer des niveaux de bandes différents. Ainsi pour les 19 séquences polymorphes chez au moins un des deux géniteurs, un nombre variable de bandes est observé pour chaque allèle : 2 bandes, 3 bandes, 4 bandes ou encore un profil multi bandes. Mais l'observation des descendants permet de retrouver les bandes qui ségrègent ensemble et donc de les considérer comme un seul et même allèle.

Les séquences polymorphes ont été cartographiées sur les cartes génétiques d'*E. grandis* et *E. urophylla* déjà existantes pour ce croisement (Gion *et al.* 2000). Elles ont été établies à partir de marqueurs dominants, les RAPD, et de marqueurs codominants, les microsatellites, les gènes de fonction connue et les EST. Des liaisons significatives entre les « marqueurs jalons » et les 19 séquences ont été établies. L'existence de 15 groupes de liaison chez *E. urophylla* ( $2n=22$  chez le genre eucalyptus), contre 11 groupes de liaison chez *E. grandis* (figure 7) semble indiquer que le niveau de saturation des cartes n'est pas atteint.

Les longueurs des 11 groupes de liaison majeurs oscillent entre 73 et 150 cM pour *E. grandis*, et entre 77 et 163 cM pour *E. urophylla* avec une longueur moyenne d'environ 113 cM pour les deux génotypes.

Tableau 7 : Récapitulatif des marqueurs codominants positionnés sur les cartes génétiques des deux géniteurs. (EST : expressed sequence tag, SSR : single sequence repeat, CG : candidate gene).

| Groupe de liaison | Type de marqueur | Nbre de marqueurs | Nbre cartographiés chez <i>E. urophylla</i> | Nbre cartographiés chez <i>E. grandis</i> | Nbre commun |
|-------------------|------------------|-------------------|---|---|-------------|
| 1                 | EST              | 5                 | 4   | 3   | 2           |
| 1                 | SSR              | 1                 | 1   | 1   | 1           |
| 2                 | EST              | 3                 | 2   | 2   | 1           |
| 2                 | SSR              | 1                 | 1   | 0   | 0           |
| 2                 | CG               | 2                 | 2   | 2   | 2           |
| 3                 | EST              | 4                 | 3   | 2   | 1           |
| 3                 | SSR              | 1                 | 1   | 1   | 1           |
| 4                 | EST              | 5                 | 3   | 4   | 2           |
| 4                 | SSR              | 1                 | 1   | 1   | 1           |
| 4                 | CG               | 2                 | 2   | 0   | 0           |
| 5                 | EST              | 2                 | 2   | 2   | 2           |
| 6                 | EST              | 3                 | 1   | 3   | 1           |
| 6                 | SSR              | 1                 | 1   | 0   | 0           |
| 6                 | CG               | 2                 | 1   | 1   | 1           |
| 7                 | EST              | 2                 | 1   | 1   | 0           |
| 7                 | SSR              | 3                 | 3   | 1   | 1           |
| 7                 | CG               | 3                 | 3   | 0   | 0           |
| 8                 | EST              | 1                 | 1   | 0   | 0           |
| 8                 | SSR              | 2                 | 2   | 1   | 1           |
| 9                 | EST              | 2                 | 2   | 1   | 1           |
| 9                 | SSR              | 1                 | 1   | 0   | 0           |
| 9                 | CG               | 2                 | 2   | 1   | 1           |
| 10                | EST              | 2                 | 1   | 2   | 1           |
| 10                | CG               | 2                 | 2   | 2   | 2           |
| 10                | SSR              | 1                 | 1   | 0   | 0           |
| 11                | EST              | 2                 | 1   | 2   | 1           |
| 11                | CG               | 1                 | 1   | 0   | 0           |

### **Localisation des séquences polymorphes uniquement chez *E. urophylla***

Pour ce géniteur, les 7 séquences polymorphes ont pu être positionnées sur les groupes de liaison existants. L'EST 09F08 a été positionnée sur le groupe de liaison 1a (GL1a), 08G01 et 08C02 ont été cartographiées sur le GL3, 08H11 sur le GL4b, 09D03 sur le GL7, 09H02 sur le GL8 et enfin le gène candidat GF#27 sur le GL9.

### **Localisation des EST polymorphes uniquement chez *E. grandis***

Les 5 marqueurs polymorphes chez ce géniteur ont pu être cartographiés sur les groupes de liaison existants. Ainsi, l'EST 09D01 et 08E12 gène1 sont positionnées sur le GL1, 08F09 sur le GL 4, 08H04 sur le GL7 et enfin 09C05 sur le GL10.

### **Localisation des séquences polymorphes chez les deux géniteurs**

Dans cette étude, 8 séquences se sont révélées polymorphes chez les deux géniteurs. Elles ont toutes pu être cartographiées sur des groupes de liaison existants. Le gène fonctionnel GF#86 est cartographié sur le GL2. L'EST 09H06 se positionne sur le GL3 pour les deux géniteurs, 08E04 et 08H07 sur le GL4 (GL4a pour *E. urophylla*), 08C08 et 08E12 gène2 sur le GL5, 08B02 sur le GL6, et enfin 08F03 sur le GL9.

Avec les résultats de notre étude, 57 marqueurs codominants sont placés sur les cartes génétiques de ce croisement élite (tableau 7): 31 EST, 12 microsatellites (SSR) et 14 gènes candidats. Les cartes génétiques d'*E. urophylla* et d'*E. grandis* possèdent respectivement 46 et 33 marqueurs codominants dont 23 sont en commun. Ces derniers nous permettent de valider la quasi-totalité des homologies de groupes de liaison entre les deux espèces. Aucun gros réarrangement chromosomique n'est à noter entre des deux espèces, en effet tous les marqueurs codominants communs se positionnent sur les groupes de liaison homologues des deux espèces ou ne sont pas liés à un groupe existant. Les distances respectives entre deux marqueurs communs ne sont pas conservées. A titre d'exemple, la distance entre les deux gènes de fonction CAD et CCoAOMT localisés sur le GL10 de chaque parent est respectivement de 1,9 et 7,1 cM chez *E. urophylla* et *E. grandis*. De même, la distance entre les EST 08H07 et 08E04 positionnées sur le GL4 est respectivement de 30,3 et 26,6 cM chez *E. grandis* et *E. urophylla*. On peut également noter que cette étude a permis de positionner des EST en bout de groupe de liaison pour quatre d'entre eux. Ceci nous apporte de nouvelles informations sur la définition des groupes et la détection de QTL.

## 4. Discussion

### 1. SSCP

Beaucoup d'auteurs s'accordent à dire que la SSCP est une méthode sensible, rapide, et peu coûteuse (Sunnuck *et al.*, 2000, Plomion *et al.*, 1999, Bodenes *et al.*, 1996, Fukuoka *et al.*, 1994, Sheffield *et al.*, 1993, Orita *et al.*, 1989). Elle permet de détecter des mutations ponctuelles sur des fragments de taille allant de 95 à plus de 600 pb (Sheffield *et al.*, 1993). Dans notre étude, la SSCP a permis de révéler du polymorphisme dans 62,5% des fragments amplifiés. Si l'on compare ces résultats en fonction de la taille des fragments en différenciant trois classes : <250pb, 250/300pb et >300pb nous constatons que respectivement 9 fragments sur 24, 9/22 et 10/22 révèlent du polymorphisme. La détection de polymorphisme par la SSCP n'est ici pas affectée par la taille des fragments, des résultats similaires ont été décrits par Fukuoka *et al.* (1994) ainsi que par Plomion *et al.* (1999). Ces derniers relèvent également que la présence d'introns (8 séquences sur 16 séquences de pin maritime) n'améliore pas la détection de polymorphisme. Dans notre étude sur les 5 EST pour lesquelles nous détectons un intron, deux montrent un polymorphisme maternel, nous ne pouvons pas relier présence d'intron et niveau de polymorphisme. Dans ce sens, Champurney (2003) a montré que les variations nucléotidiques au sein du gène CCR pour 40 génotypes d'*E. urophylla* étaient principalement situées dans les zones introniques. L'absence de détection de polymorphisme par SSCP ne signifie pas l'absence de variation nucléotidique, en effet Sheffield *et al.* (1993) ont démontré que la position de la mutation influait sur la sensibilité de la SSCP. Dans notre étude, pour les 12 EST n'ayant pas montré de polymorphisme, il est possible d'envisager d'autres approches pour détecter du polymorphisme : CAPS, denaturing gradient gel electrophoresis ou DGGE, temperature gradient gel electrophoresis ou TGGE, Ecotilling... Suivant les cas, il est possible d'exploiter ses régions flanquantes par une nouvelle définition d'amorces. Il est probable également qu'il n'existe pas de variations de séquence entre les deux géniteurs du croisement.

Tableau 8 : Détection des SNP et des INDEL (insertions délétions) par la technique SSCP

| Séquence<br>id. | Insertions délétions<br>détectées |                             | Nombre de SNP<br>détectés     |                             | Détection du polymorphisme<br>par SSCP |                   |
|-----------------|-----------------------------------|-----------------------------|-------------------------------|-----------------------------|--|-------------------|
|                 | <i>E.</i><br><i>urophylla</i>     | <i>E.</i><br><i>grandis</i> | <i>E.</i><br><i>urophylla</i> | <i>E.</i><br><i>grandis</i> | <i>E. urophylla</i>                    | <i>E. grandis</i> |
|                 | 06A01                             | non                         | non                           | 1                           | 1                                      | -                 |
| 06A08           | non                               | non                         | 0                             | 0                           | -                                      | -                 |
| 06B07           | non                               | non                         | 0                             | 0                           | -                                      | -                 |
| 06C08           | oui                               | oui                         | -                             | -                           | -                                      | -                 |
| 06D02           | non                               | non                         | 0                             | 0                           | -                                      | -                 |
| 06F06           | non                               | non                         | 0                             | 0                           | -                                      | -                 |
| 06H12           | non                               | non                         | 0                             | 0                           | -                                      | -                 |
| 08B06           | non                               | non                         | 0                             | 2                           | -                                      | -                 |
| 08C10           | non                               | non                         | 0                             | 0                           | -                                      | -                 |
| 08D04           | non                               | non                         | 1                             | 1                           | -                                      | -                 |
| 08E11           | non                               | non                         | 1                             | 0                           | -                                      | -                 |
| 08F01           | non                               | non                         | 0                             | 1                           | -                                      | -                 |
| 08G08           | non                               | non                         | 0                             | 0                           | -                                      | -                 |
| 09C10           | non                               | non                         | 0                             | 0                           | -                                      | -                 |
| 06D07           | non                               | non                         | 2                             | 1                           | -                                      | +                 |
| 07A01           | non                               | oui                         | 0                             | -                           | -                                      | +                 |
| 08F09           | non                               | non                         | 1                             | 1                           | -                                      | +                 |
| 09D01           | non                               | non                         | 0                             | 1                           | -                                      | +                 |
| 06A11           | non                               | oui                         | 1                             | -                           | +                                      | -                 |
| 06A12           | non                               | non                         | 2                             | 1                           | +                                      | -                 |
| 06D03           | non                               | non                         | 1                             | 0                           | +                                      | -                 |
| 06D12           | non                               | non                         | 1                             | 1                           | +                                      | -                 |
| 08C02           | non                               | non                         | 2                             | 0                           | +                                      | -                 |
| 08G01           | non                               | non                         | 2                             | 0                           | +                                      | -                 |
| 09D03           | oui                               | non                         | 1                             | 1                           | +                                      | -                 |
| 09H02           | non                               | non                         | 0                             | 0                           | +                                      | -                 |
| GF#27           | non                               | non                         | 1                             | 1                           | +                                      | -                 |
| 06A07           | oui                               | oui                         | -                             | -                           | +                                      | +                 |
| 06D01           | oui                               | non                         | 0                             | 0                           | +                                      | +                 |
| 08B02           | non                               | oui                         | 2                             | -                           | +                                      | +                 |
| 08E04           | oui                               | non                         | 1                             | 0                           | +                                      | +                 |
| 08H07           | oui                               | non                         | -                             | 0                           | +                                      | +                 |
| GF#86           | non                               | non                         | 3                             | 2                           | +                                      | +                 |



## 2. Les causes moléculaires de la détection de polymorphisme en SSCP

Les résultats de séquençage des produits PCR, où les deux allèles sont analysés simultanément, confirment les ségrégations observées pour chaque marqueur SSCP (tableau 8). Deux types de séquences ont été observés :

- ✓ Des séquences d'EST présentant des indéterminations pour certains nucléotides. Ceci peut indiquer que l'individu est hétérozygote pour le fragment amplifié.
- ✓ Des séquences ne présentant aucune indétermination dans la séquence nucléotidique du fragment amplifié. Dans ce cas le produit PCR correspond à une seule séquence, indiquant soit l'homozygotie soit l'hétérozygotie avec un allèle non amplifié (allèle nul) où une mutation s'est insérée dans le site de fixation d'une des amorces.

Les EST ayant montré un profil SSCP comportant un allèle nul (08E04, 08H07, 09H02 et 06D01) montrent après séquençage un chromatogramme sans variation pour le gène en question confirmant l'absence d'amplification pour un des deux allèles de ce gène.

Dans cette étude, 7/9 EST présentant des INDEL et 53% des SNP supposés ont pu être détectés par la SSCP. Sheffield *et al.* (1993) ont montré que la détection de mutations ponctuelles est optimale pour des fragments d'une taille inférieure à 250pb environ ; en effet, 70% des mutations ponctuelles sont détectées pour un fragment de 212 pb. Or 8 de nos fragments (sur un total de 12) pour lesquels aucun SNP n'a été détecté en SSCP, ont une taille supérieure à 250 pb. Ces résultats confirment la sensibilité de la SSCP et de notre approche sur des fragments de taille inférieure à 250 pb. De plus, les fragments en question ne possèdent qu'un seul SNP dans leur séquence, ce qui limite davantage les chances de révéler du polymorphisme au sein des fragments. Un autre facteur important, avancé par Sheffield *et al.*, (1993), dans la détection du polymorphisme par SSCP est la position de la mutation sur le fragment amplifié. Les SNP non détectés se positionnent en général en début ou en fin de séquence (11 cas sur 13), la totalité des mutations ponctuelles observées se trouvant dispersées tout au long des séquences contrairement aux résultats de Sheffield *et al.* (1993) qui ne distinguaient aucune zone particulière susceptible de ne pas être détectée en SSCP.

L'efficacité de la SSCP est dépendante de la taille des fragments et de la position de la mutation, et peut varier significativement d'un fragment à l'autre. Même s'il existe une corrélation négative entre la taille du fragment et la sensibilité de la SSCP, l'optimisation des conditions de migration s'est révélée payante et les fragments de taille plus grande peuvent également être analysés. Par exemple, le fragment 06D12 (700/800pb) présentant un SNP pour chacun des géniteurs, a révélé un polymorphisme maternel.

Les autres techniques couramment employées incluent la denaturing gradient gel electrophoresis (DGGE), la temperature gradient gel electrophoresis (TGGE), les analyses des heteroduplex et leurs dérivées ou encore la cleaved amplified polymorphism sequence (CAPS). L'intérêt de ces techniques varie en fonction de leurs avantages et de leurs inconvénients en termes de sensibilité, optimisation, coûts et équipements nécessaires. La SSCP est la technique, selon Sunnucks *et al.* (2000), qui offre le meilleur compromis.

Néanmoins, nous avons testé la CAPS sur une EST présentant un intron dont la taille du fragment amplifié, 1200/1500pb, pouvait nous faire suggérer un nombre de mutation suffisamment important pour inclure un site de coupure spécifique à chacun des deux géniteurs. Cette technique nous a permis de mettre en évidence du polymorphisme chez le géniteur femelle tout comme en SSCP. Cette technique simple d'emploi semble pourtant ne pas être aussi sensible que la SSCP, en effet Sunnucks *et al.* (2000) rapportent que la CAPS a permis de détecter 1 site polymorphe sur 20, alors que la SSCP en a détecté 90%. Sato et Nishio (2003) ont mis au point une variante combinant ces deux techniques, la « PCR-RF-SSCP » pour les fragments de taille importante (dans l'étude 2Kb). Après amplification des produits, le fragment est digéré par une enzyme de restriction, puis le produit issu de la digestion est passé en SSCP. Plus de 70% des mutations ponctuelles du gène amplifié ont été détecté par cette technique. Il serait donc intéressant de la tester sur nos fragments de grandes tailles. On peut également envisager l'étude de gènes complets de grande taille en limitant le nombre de contigs.

### 3. Les autres apports du séquençage

La mise en évidence de SNP au sein des fragments amplifiés permet l'application de méthodes de génotypage à haut débit sur la descendance complète disponible ; cette étude a porté sur la première descendance plantée qui représentait 201 individus, grâce à un second croisement réalisé à partir des mêmes géniteurs, une autre famille de 600 descendants est disponible. De nombreuses innovations techniques voient le jour dans le génotypage à haut débit, leurs principes sont variables mais la technique qui semble la plus utilisée aujourd'hui, dite du « Taqman assay », est basée sur la différence d'intensité de fluorescence existant entre deux groupes d'individus en fonction de la composition respective des deux allèles au SNP. Cette technique semble être une des suites logiques de notre étude afin de positionner le maximum d'EST en un temps réduit. Il est également possible de rechercher une enzyme de restriction capable de couper au site SNP, générant ainsi des marqueurs CAPS ségrégant dans la descendance (Quint *et al.* 2002). Il serait également intéressant de déterminer si ces mutations ponctuelles apparaissent comme des mutations silencieuses ou non synonymes notamment pour les séquences colocalisant avec des caractères d'intérêts (QTL). Ces résultats pourraient nous permettre de déterminer l'effet respectif de chaque allèle sur les caractères d'intérêts ciblés en sélection et d'appliquer ces résultats pour la sélection. A titre d'exemple, Champurney (2003) a montré qu'*E. urophylla* portait une mutation non synonyme chez le gène CCR, gène colocalisant avec un QTL de teneur en lignines.

Pour les trois EST où une amplification intronique a été décelée au séquençage, le type et le nombre de sites variables se répartissent différemment entre les introns et les exons. 72% des SNP se positionnent sur les zones introniques, ainsi que la totalité des INDEL observées. Ces dernières sont de taille significativement plus grande, de 6 à 10 bases en général, contre une ou deux bases en moyenne pour les autres EST. Les introns correspondent à des régions non codantes du gène, ce qui implique que les mutations peuvent s'y accumuler sans résultat sélectif sur les plantes (Plomion *et al.*, 1999). Nous nous attendions à ce que le polymorphisme soit plus élevé dans les zones introniques que dans les exons. Nos résultats ne vont pas dans ce sens car aucune INDEL intra spécifique n'a pu être détectée dans ces EST et l'intron semble même être bien conservé d'une espèce à l'autre, nous observons 95% d'homologies entre les espèces pour l'intron détecté dans l'EST 06D07.

Le séquençage nous a permis de confirmer la présence d'un microsatellite sur l'EST 06A07. Les microsatellites étant des marqueurs hautement polymorphes, ce type de marqueur pourrait faciliter la cartographie des EST. De plus, Khlestkina *et al.* (2004) ont démontré que les marqueurs microsatellites issus de séquences EST étaient mieux conservés que les microsatellites issus de séquence du génome (STS) au sein des céréales. Ce type de microsatellite pourrait donc être utilisé pour l'étude de l'évolution des génomes du genre *Eucalyptus*.

#### 4. La transférabilité des EST

Les amorces utilisées pour amplifier les EST chez *E. globulus*, *E. grandis* et *E. urophylla* ont été mises au point sur des séquences d'ADNc développées chez *E. gunnii*. Elles ont permis d'obtenir 90% de séquences exploitables dans la recherche de polymorphisme pour les deux géniteurs, *E. urophylla* et *E. grandis* et *E. globulus*. Après séquençage, une moyenne de 68% d'homologies de séquence est retrouvée entre chacun des géniteurs et *E. gunnii*, alors que la moyenne entre les deux géniteurs est de 83%. Puisque au sein du sous genre *Symphyomyrtus*, *E. gunnii* appartient à la sous section *Maidenaria*, *E. grandis* et *E. urophylla* sont regroupés dans la sous section *Transversaria*, nos résultats nous suggèrent donc que les séquences sont mieux conservées au sein d'espèces d'une même sous section. D'autres études corroborent ces résultats, en effet Gion *et al.* (2000) ont utilisé avec succès des séquences de *E. globulus* et de *E. gunnii* pour définir des amorces spécifiques de 6 gènes candidats sur *E. grandis* et *E. urophylla*. Champurney (2003) a également étudié la transférabilité du gène CCR au sein de quatre espèces différentes *E. urophylla*, *E. gunnii*, *E. saligna* et *E. globulus*. Les marqueurs microsatellites semblent également être bien conservés au sein du genre *Eucalyptus*. Byrne *et al.* (1996) ont montré que 90% des microsatellites mis au point sur *E. nitens* étaient transférables, au travers de 6 espèces, dans le sous genre *Symphyomyrtus* et avec une moindre efficacité aux sous genre *Monocalyptus* et *Corymbia*. De même, Marques *et al.* (2002) ont démontré la transférabilité des microsatellites mis au point par Brondani *et al.* (1998) sur un croisement *E. urophylla* x *E. grandis* à un croisement entre *E. tereticornis* et *E. globulus*. Une autre étude plus récente (Cabannes, 2004) a révélé le possible transfert des amorces mises au point sur les EST de *E. gunnii* à 87 génotypes de 29 espèces différentes, réparties comme suit : 4 espèces appartenant au sous genre *Corymbia*, 5 espèces au sous genre *Monocalyptus*, 1 espèce aux *Idiogènes* et 19 aux *Symphyomyrtus*. Ces différents résultats révèlent une très bonne conservation du génome

entre des différentes espèces d'eucalyptus existantes. Un set de 200 000 EST est en cours d'étude dans le cadre du programme Génolyptus (association d'organismes d'Amérique du Sud essentiellement) sur *E. grandis* et *E. urophylla*, la conservation des ressources génomiques au sein du genre permet d'envisager une exploitation de ce set d'EST au genre entier pour un grand nombre d'espèces.

## 5. Cartographie

Les cartes font une distance totale de 1238 et 1372 cM respectivement pour *E. grandis* et *E. urophylla*. Verhaegen et Plomion (1996) avaient estimé la longueur des cartes génétiques de ce croisement à 1332 cM et à 1483 cM. La longueur observée, pour *E. urophylla*, supérieure à celle définie précédemment s'expliquerait par l'existence d'un nombre supérieur de groupes de liaison ainsi qu'à des distances élevées entre quelques marqueurs. A titre d'exemple, 6 distances entre marqueurs adjacents au sein des groupes de liaison sont supérieures à 27 cM *Kosambi*. Ces résultats reflètent la non saturation des cartes génétiques des deux géniteurs du croisement déjà évoquée par Carrouché (2003) et par Gion (2001) qui estimait la couverture des cartes génétiques à 80% pour *E. grandis* et à 94% pour *E. urophylla*. Le positionnement des différentes EST et gènes candidats sur l'ensemble des groupes de liaison n'a révélé aucun cluster de gènes susceptibles d'avoir un effet sur la qualité du bois. Cette indépendance entre locus constitue une situation plus favorable à la distinction des gènes impliqués dans le contrôle génétique de la qualité du bois. Ce résultat augmente également la probabilité de regrouper un maximum d'allèles favorables dans un même génotype.

L'EST distordue au seuil de 5% a été cartographiée sur le groupe de liaison 5 pour les deux géniteurs. La ségrégation observée pour cette EST est en adéquation avec une ségrégation postzygotique (Ky *et al.*, 2000). Les études antérieures de cartographie sur *Eucalyptus* relatent toutes ce problème de distorsion de ségrégation avec un taux oscillant entre 15 et 30% des marqueurs (Marques *et al.*, 1998, Myburg *et al.*, 2003, Thamarus *et al.*, 2002). Myburg *et al.* (2003 et 2004) imputent ces observations à des incompatibilités géniques lors de la fécondation même si les eucalyptus ont un génome bien conservé entre espèces. Bien que les croisements interspécifiques soient possibles, les taux de réussite sont très faibles souvent inférieur à 1% (96% des hybrides F1 non viables après fécondation pour le croisement réalisé dans l'étude de Myburg *et al.*, 2003). Ces problèmes d'incompatibilité

Tableau 9 : Comparaison de la position de différents marqueurs codominants sur les groupes de liaisons des cartes génétiques de plusieurs espèces d'eucalyptus. Abréviations utilisées SSR : single sequence repeat, CG : candidate gene, EST : expressed sequence tag, ne : non étudié, - : étudié mais non polymorphe dans le croisement considéré.

| Types de marqueurs | Nom du marqueur | Brondani et al. (1998)           |                                | Thamarus et al. (2002)           |                                | Marques et al. (2002)           |                                 |                                     |
|--------------------|-----------------|----------------------------------|--------------------------------|----------------------------------|--------------------------------|---------------------------------|---------------------------------|-------------------------------------|
|                    |                 | Position sur <i>E. urophylla</i> | Position sur <i>E. grandis</i> | Position sur <i>E. urophylla</i> | Position sur <i>E. grandis</i> | Position sur <i>E. globulus</i> | Position sur <i>E. globulus</i> | Position sur <i>E. tereticornis</i> |
| SSR                | embra 7         | 2                                | -                              | -                                | 9                              | ne                              | 5                               | 3                                   |
|                    | embra 13        | 10                               | -                              | 9                                | 9                              | ne                              | -                               | 4                                   |
|                    | embra 6         | ne                               | ne                             | 1                                | 1                              | 7                               | 8                               | 10                                  |
|                    | embra 11        | 7                                | -                              | 1                                | 1                              | ne                              | 8                               | 10                                  |
|                    | embra 12        | 7                                | 7                              | 1                                | 1                              | ne                              | 8                               | 10                                  |
|                    | embra 5         | 8                                | -                              | 5                                | 5                              | 9                               | 10                              | 6                                   |
|                    | embra 9         | 8                                | 8                              | 5                                | 5                              | ne                              | non lié                         | 6                                   |
|                    | embra 8         | 1                                | 1                              | 6                                | -                              | ne                              | 1                               | 1                                   |
|                    | embra 10        | 6                                | -                              | 10                               | 10                             | ne                              | 4                               | 14                                  |
|                    | embra 2         | 4                                | 4                              | -                                | 11                             | 3                               | ne                              | ne                                  |
| CG                 | ELF1            | 4                                | -                              | ne                               | ne                             | 3                               | ne                              | ne                                  |
|                    | 4CL             | 11                               | -                              | ne                               | ne                             | 3                               | ne                              | ne                                  |
|                    | COMT1/2         | 7                                | 7                              | ne                               | ne                             | 7                               | ne                              | ne                                  |
|                    | CCoAOMT         | 10                               | 10                             | ne                               | ne                             | 2                               | ne                              | ne                                  |
|                    | CCR             | 6                                | 6                              | ne                               | ne                             | 10                              | ne                              | ne                                  |
| PAL                | 6               | -                                | ne                             | ne                               | 4                              | ne                              | ne                              |                                     |
| EST/CG             | GF27/e360a      | 9                                | -                              | ne                               | ne                             | 2                               | ne                              | ne                                  |
|                    | GF27/e360b      | 9                                | -                              | ne                               | ne                             | 10                              | ne                              | ne                                  |

d'origine génétique se répercutent sur l'équiprobabilité d'obtenir l'une ou l'autre allèle du géniteur mâle dans la descendance. Cependant Myburg *et al.* (2004) observent une distorsion de ségrégation équivalente pour les deux géniteurs. Taylor et Ingvarsson (2003) relatent l'existence de « gamete eliminators », l'association entre un pollen et un ovule incompatible entraîne la mort des deux. Ce mécanisme pourrait être une des causes de la distorsion observée. Les distorsions de ségrégation ayant un impact direct sur la sélection, il serait intéressant de cartographier le maximum de marqueurs distordus afin de repérer les zones candidates à la distorsion et de visualiser une éventuelle colocalisation avec des QTL pour une meilleure précision de ces derniers.

Plusieurs cartes génétiques ont été construites pour les principales espèces de reboisement (*E. urophylla* et *E. grandis* : Grattapaglia and Sederoff, 1994, Verhaegen and Plomion, 1996 ; *E. nitens* : Byrne *et al.*, 1994 ; *E. globulus* : Marques *et al.*, 1998, Myburg *et al.*, 2003, Thamarus *et al.*, 2002). Elles ont toutes été principalement construites à partir de marqueurs dominants (RAPD, AFLP) dans lesquelles peu à peu sont positionnés des marqueurs codominants (SSR, EST, STS...). Ces derniers permettent d'étudier la synténie entre différentes espèces. Dans le tableau 9, la position de chaque marqueur est indiquée par le numéro du groupe de liaison; il semblerait ainsi que le GL1 de Brondani *et al.* (1998) corresponde au GL7 de notre étude, au GL7 de Thamarus *et al.* (2002) et enfin, aux GL8 chez *E. globulus* et GL10 chez *E. tereticornis* de Marques *et al.* (2002). On dénote quelques différences pour certains marqueurs ainsi le gène CCR et le gène PAL qui sont tous deux placés sur le GL6 de notre étude se positionnent sur deux groupes différents chez *E. globulus* (Thamarus *et al.*, 2002) ; De même, le microsatellite Embra 02, le gène ELF1 et le gène 4CL placés sur le groupe 3 chez *E. globulus* sont cartographiés sur deux groupes distinct dans notre croisement (Embra 02 et ELF1 sur le GL4 et 4CL sur le GL11). Ces différences s'expliqueraient par des réarrangements chromosomiques entre espèces, *E. globulus* faisant partie d'une autre sous-section (*Maidenaria*), il peut s'agir également de familles multi géniques, chaque étude ayant étudié un gène différent de la famille. En ce sens, Thamarus *et al.* (2002) ont démontré l'existence de deux gènes PAL. Par contre, deux microsatellites Embra 07 et Embra 13 placés sur deux groupes différents sur notre carte d'*E. urophylla*, et celle d'*E. tereticornis*, sont situés sur le même groupe chez *E. grandis* (Brondani *et al.*, 1998). Ceci semble indiquer qu'un réarrangement chromosomique s'est opéré dans la même sous-section. De même, au sein de notre croisement, certaines distances entre marqueurs ne sont pas conservées. Ces différences peuvent être dues à un effet de l'espèce, au sexe, à de petits

réarrangements chromosomiques au cours de l'évolution ou encore à une erreur d'échantillonnage (Gion, 2001). Marques *et al.* (2002) observent également des distances non conservées entre marqueurs d'une espèce à l'autre, mais aussi des permutations entre 2 marqueurs. Afin de vérifier ces hypothèses, nous pouvons envisager d'utiliser un nombre plus élevé de marqueurs communs sur chaque groupe de liaison, de réaliser le croisement inverse (*E. grandis* comme femelle et *E. urophylla* comme mâle) et de considérer un nombre d'individu plus important. Il serait également nécessaire de cartographier plus de marqueurs pour confirmer ces hypothèses, et permettre de se prononcer de façon plus certaine sur le degré d'homologie des génomes au sein du genre *Eucalyptus*. Ainsi, l'étude de la conservation des groupes de liaison et la comparaison de l'ordonnement des gènes à l'intérieur des mêmes groupes de liaison, permettrait d'étudier l'évolution du génome dans le genre *Eucalyptus*. L'utilisation d'un set de marqueurs codominants communs aux équipes de recherche pourrait conduire à la création d'une carte consensus du genre. Elle permettrait non seulement l'étude de l'évolution du génome mais aussi la possibilité de saturer des cartes génétiques spécifiques d'un croisement, l'opportunité de pouvoir construire une carte génétique pour les croisements de géniteurs à faible polymorphisme (croisements intra spécifiques surtout). Les marqueurs communs pourraient être aussi utilisés dans le maintien de la variabilité, la détection de QTL dans les différents fonds génétiques et la prédiction de localisation des gènes orthologues dans les autres espèces (Marques *et al.*, 2002).



## 5. Conclusion et perspectives

L'utilisation de la technique SSCP pour la cartographie d'EST et de gènes candidats fonctionnels est une technique nécessitant peu d'investissements pour révéler efficacement du polymorphisme de séquence. Nous avons également abordé la possibilité de générer des marqueurs CAPS à partir d'une séquence d'EST, cette technique représente une alternative à la SSCP dans le cas de fragments plus longs. Grâce au séquençage des produits PCR, nous avons pu confirmer la nature des fragments amplifiés chez *E. urophylla* et *E. grandis* et détecter des sites de mutations ponctuels non décelés en SSCP. Mais il s'agirait d'une étape préalable pour mettre en place d'autres stratégies de détection du polymorphisme plus ciblées sur les sites de mutation, le génotypage à haut débit ou encore la PCR-RF-SSCP sont, entre autres, des techniques envisageables.

Les résultats obtenus dans cette étude montrent que les EST développées chez *E. gunnii* sont des séquences bien conservées entre espèce du même sous genre, le haut degré d'homologie retrouvé entre les séquences des 3 espèces nous permettant de confirmer la nature des fragments amplifiés. L'étude de Cabannes (2004), plus exhaustive, confirme cette conservation dans les autres sous genres de l'*Eucalyptus*. Grâce à ces données, les approches de cartographie comparée déjà amorcées vont pouvoir se développer plus rapidement et conduire à la création d'une carte consensus du genre *Eucalyptus* d'une part, et d'autre part à l'étude de l'évolution des génomes des espèces de ce genre.

Afin d'envisager une utilisation en « sélection assistée par marqueurs », les EST nouvellement positionnées devront être intégrées dans les études de colocalisation marqueurs/QTL. Toutefois, une densification des cartes génétiques doit être réalisée au préalable notamment sur les sites de détection des QTL connus pour ce croisement. Dans le règne végétal, les études menées principalement sur les espèces de grandes cultures montrent que ces caractères sont sous le contrôle d'un petit nombre de gènes à effets forts, et donc potentiellement manipulables par le sélectionneur (Plomion, 1996). La colocalisation EST/QTL à effets forts nécessite une étude fonctionnelle *à posteriori* de l'EST dont la fonction n'est connue que par homologies de séquence. L'accès à la séquence entière du gène correspondant aux EST permettrait une étude plus exhaustive des sites de mutation et un repérage dans la population des différents allèles existants à ce locus pour finalement estimer leur valeur phénotypique propre.

Une autre possibilité de valorisation des cartes génétiques pourrait se développer. Dans une sélection à croisements non contrôlés, la recherche de paternité est réalisée sur les plantules obtenues pour déterminer la meilleure combinaison parentale. Cette recherche de paternité est réalisée à l'aide de marqueurs codominants (microsatellites dans l'étude de Grattapaglia *et al.*, 2004). Les valeurs phénotypiques des descendants sont déterminées pour les caractères d'intérêts agronomiques puis mises en relation avec la paternité observée. Dans l'étude de Grattapaglia *et al.* (2004), un géniteur mâle ressort plus fréquemment dans le lot d'arbres sélectionnés. Le même type d'approche pourrait être réalisé avec les EST. Dans ce cas un nombre plus important d'EST cartographiées pour les deux espèces serait indispensable afin de tester un maximum de géniteurs et de combinaisons interspécifiques. Le gain ne se fait plus sur la durée des rotations mais sur le nombre de croisements réalisés simultanément sans contrôle, et la capacité des géniteurs à générer un nombre de semences viables plus important, réel problème dans la sélection des hybrides interspécifiques.

L'étude et la cartographie d'EST n'est qu'une étape préliminaire à l'association gènes candidats / caractères d'intérêts. Dans ce contexte, la sélection assistée par marqueurs laisse présager des avancées significatives mais beaucoup de travaux restent à réaliser avant d'envisager une SAM effective.

## Bibliographie

- Afocel. Information *Eucalyptus*, Présentation générale de l'eucalyptus. Numéro 1, sept. 2003.
- Bodenes C., Laigret F., Kremer A. Inheritance and molecular variations of PCR-SSCP fragments in pedunculate oak (*Quercus robur* L.). *Theor. Appl. Genet.*, 93 : 348-354, 1996.
- Bossinger G., Leitch M.A. Isolation of cambium-specific genes from *Eucalyptus globulus* Labill. In : Savidge R., Barnett J., Napier R (eds) Cell and molecular biology of wood formation. BIOS Scientific, Oxford, 203-207, 2000
- Bouvet J.M. Evolution de la variabilité avec l'âge et corrélation juvénile adulte dans des populations d'*Eucalyptus*. Thèse de doctorat Institut National Agronomique Paris-Grignon. 236p, 1995.
- Brondani R.P.V., Brondani C., Grattapaglia D. Towards a genus-wide reference linkage map for Eucalyptus based exclusively on highly informative microsatellite markers. *Mol. Genet. Genomics*, 267: 338-347, 2002.
- Brondani R.P.V., Brondani C., i R., Grattapaglia D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. *Theor. Appl. Genet.*, 97: 816-827, 1998.
- Byrne M., Marquez-Garcia M.I., Uren T., Smith D.S. and Moran G.F. Conservation and genetic diversity of microsatellite loci in the genus *Eucalyptus*. *Aust. J. Bot.*, 44: 331-341, 1996.
- Byrne M., Murrel J.C., Owen J.V., Allen B., and Moran G.F. An integrated genetic linkage map for eucalypts using RFLP, RAPD and isozyme markers. *Theor. Appl. Genet.*, 91:869-875, 1994.
- Cabannes E. Transférabilité d'EST impliquées dans la xylogénèse au sein du genre *Eucalyptus*. Rapport de maîtrise, Université de Montpellier II, 2004.
- Carrouché A. Cartographie génétique de gènes candidats pour la qualité du bois. Rapport de DESS, Université de Toulouse III, 2003.
- Champurney N. Mise en évidence de polymorphisme au sein du gène CCR chez *Eucalyptus urophylla*. Rapport de DEA, Université de Montpellier II, 2003.
- Creste S., Tulmann Neto A. and Figueira A. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol. Biol. Report.* 19: 299-306, 2001.
- De Vienne D, Causse M., Maurice A., Santoni S. Les marqueurs moléculaires et leurs applications. In Biotechnologies végétales, cours du CNED, 87p, 1995.

Diatchenko L., Lau Y-F. C., Campbell A.P., Chenchik A., Moqadam F., Huang B., Lukyanov S., Gurskaya N., Sverdlov E.D., and Siebert P. Suppression subtractive hybridization : a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc. Natl. Acad. Sci. USA, 93: 6025-6030, 1996.

Fisher D.D., and Cyr R.J. Extending the microtubule/microfibril paradigm. Plant Physiol., 116 : 1043-1051, 1998.

Food and agricultural organisation. Situation des forêts du monde 2001. In Eds Organisation des Nations Unies pour l'alimentation et l'agriculture, Rome (Italie), 2001

Fukuoka S., Inoue T., Miyao A., Monna L., Zhong H.S., Sasaki T. And Minobe Y. Mapping of sequence-tagged sites in rice by single strand conformation polymorphism. DNA Research, 1: 271-277, 1994.

Gion J-M. Etude de l'architecture génétiques des caractères complexes chez l'eucalyptus : des marqueurs anonymes aux gènes candidats. Thèse de doctorat, Université de Rennes I, 2001

Gion J-M., Rech P., Grima-Pettenati J., Verhaegen D. And Plomion C. Mapping candidate genes in *Eucalyptus* with emphasis on lignification genes. Molecular breeding, 6: 441-449, 2000.

Glavac D. and Dean M. Optimization of the single strand conformation polymorphism (SSCP) technique for detection of point mutations. Human mutation, 2: 404-414, 1993.

Grattapaglia D. and Sederoff R. genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics, 137: 1121-1137, 1994.

Grattapaglia D., Ribeiro V.J. and Rezende G.D.S.P. Retrospective selection of elite parent trees using paternity testing with microsatellite markers : an alternative short term breeding tactic for *Eucalyptus*. Theor. Appl. Genet., 109: 192-199, 2004.

Khlestkina E.K., Myint Than M.H., Pestsova E.G., Röder M.S., Malyshev S.V., Korzun V., and Börner A. Mapping of 99 new microsatellite-derived loci in rye (*Secale cereale* L.) including 39 expressed sequence tags. Theor. Appl. Genet., sous presse, 2004.

Kosambi D.D. The estimation of map distance from recombination values. Ann. Eugen., 12: 172-175, 1944.

Kremer A. Diversité génétique et variabilité des caractères phénotypiques chez les arbres forestiers. Genet. Sel. Evol., 26: 105-123, 1994

Ky C-L., Barre P., Lorieux M., Trouslot P., Akaffou S., Louarn J., Charrier A., Hamon S., Noirot M. Interspecific genetic linkage map, segregation distortion and genetic conversion in coffee (*Coffea sp.*) Theor. Appl. Genet., 101 : 669-676, 2000.

Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., and Newburg L. MAPMAKER : an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics, 1: 174-181, 1987

Marques C.M., Araùja J.A., Ferreira J.G., Whetten R., O'Malley D.M., Liu B.-H., Sederoff R.R. AFLP genetic maps of *Eucalyptus globulus* and *E. tereticornis*. *Theor. Appl. Genet.*, 96: 727-737, 1998.

Marques C.M., Brondani R.P.V., Grattapaglia D. and Sederoff R. Conservation and synteny of SSR loci and QTLs for vegetative propagation in four *Eucalyptus* species. *Theor. Appl. Genet.*, 105: 474-478, 2002.

Myburg A.A., Griffin A.R., Sederoff R.R., and Whetten R.W. Comparative genetic linkage maps of *Eucalyptus grandis*, *Eucalyptus globulus* and their F1 hybrid based on a double pseudo-backcross mapping approach. *Theor. Appl. Genet.*, 107: 1028-1042, 2003

Myburg A.A., Vogl C., Griffin A.R., Sederoff R.R., and Whetten R.W. Genetics of postzygotic isolation in *Eucalyptus* : whole-genome analysis of barriers to introgression in a wide interspecific cross of *Eucalyptus grandis* and *E. globulus*. *Genetics*, 166 : 1405-1418, 2004.

Myers R.M., Lumelsky N., Lerman L.S., and Maniatis T. Detection of single base substitutions in total genomic DNA. *Nature*, 313: 495-497, 1985.

Orita M., Iwahana H., Kanazawa H., Hayashi K. And Sekiya T. Detection of polymorphisms of human DNA by gel electrophoresis as single strand conformation polymorphisms. *Proc. Natl. Acad. Sci. USA*, 86 : 2766-2770, 1989.

Orita M., Suzuki Y., Sekiya T. and Hayashi K. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics*, 5 : 874-879, 1989.

Osorio L.F., White T.L., and Huber D.A. Age-age and trait-trait correlations for *Eucalyptus grandis* Hill ex Maiden and their implications for optimal selection age and design of clonal trials. *Theor. Appl. Genet.*, 106: 735-743, 2002.

Paux E., Tamasloukht M'B., Ladouce N., Sivadon P., and Grima-Pettenati J. Identification of genes preferentially expressed during wood formation in *Eucalyptus*. *Plant. Mol. Biol. Sous-presse*.

Plomion C, Hurme P., Frigerio J-M., Ridolfi M., Pot D., Pionneau C., Avila C., Gallardo F., David H., Neutelings G., Campbell M., Canovas F.M., Savolainen O., Bodenes C. And Kremer A. Developing SSCP markers in two *Pinus* species. *Molecular breeding*, 5 : 21-31, 1999.

Plomion C., Durel C.E., Verhagen D. Utilisation des marqueurs moléculaires dans les programmes d'amélioration génétiques des arbres forestiers : exemple du pin maritime et de l'eucalyptus. *Ann. Sci. For.*, 53 : 819-848, 1996.

Plomion C., Leprovost G., and Stokes A. Wood formation in trees. *Plant Physiol.*, 127: 1513-1523, 2001.

Quint M., Mihaljevic R., Dussle C.M., Xu M.L., Melchinger A.E., and Lübberstedt T. Development of RGA-CAPS markers and genetic mapping of candidate genes for sugarcane mosaic virus resistance in maize. *Theor. Appl. Genet.*, 105: 355-363, 2002.

Riesner D.G., Steger U., Wiese M., Wulfert M., Heibey M., and Henco K. Temperature-gradient electrophoresis for the detection of polymorphic DNA and for quantitative polymerase chain reaction. *Electrophoresis*, 13, 1992.

Saghai-Marooif M.A., Soliman K.M., Jorgensen R.A. and Allard R.W. Ribosomal DNA spacer length polymorphism in barley : Mendelian inheritance, chromosomal location, and population dynamics. *Proc. Natl. Acad. Sci. USA*, 81: 8014-8018, 1984.

Sato Y., and Nishio T. Mutation detection in rice waxy mutants by PCR-RF-SSCP. *Theor. Appl. Genet.*, 107: 560-567, 2003.

Sheffield Val C., Beck J.S., Kwitek A., Sandstrom D. And Stone E.M. The sensitivity of single strand conformation polymorphism analysis for the detection of single base substitutions. *Genomics*, 16: 325-332, 1993.

Sunnucks P., Wilson A.C.C., Beheregaray L.B., Zenger K., French J. And Taylor A.C. SSCP is not so difficult: the application and utility of single strand conformation polymorphism in evolutionary biology and molecular ecology. *Molecular Ecology*, 9: 1699-1710, 2000.

Taylor D.R., and Ingvarsson P.K. Common features of segregation distortion in plants and animals. *Genetica*, 117: 27-35, 2003.

Thamarus K.A., Groom K., Murrell J., Byrne M. and Moran G.F. A genetic linkage map for *Eucalyptus globulus* with candidate loci for wood, fibre, and floral traits. *Theor. Appl. Genet.* 104: 379-387, 2002.

Tragoonrung S., Kanizin V., Hayes P.M., and Blake T.K. Sequence tagged-site-facilitated PCR for barley genome mapping. *Theor. Appl. Genet.*, 84, 1992.

Verhaegen D. and Plomion C. Genetic mapping in *Eucalyptus urophylla* and *Eucalyptus grandis* using RAPD markers. *Genome*, 39 : 1051-1061, 1996.

Verhaegen D., Plomion C., Gion J.M., Poitel M., Costa.P., Kremer A. Quantitative trait dissection analysis in *Eucalyptus* using RAPD markers: 1. Detection of QTL in interspecific hybrid progeny, stability of QTL expression across different ages. *Theor. Appl. Genet.*, 95 : 597-608, 1997.

Vigeneron P. and Bouvet J-M. Amélioration des plantes tropicales; les eucalyptus. Editions CIRAD, 1997

Rapport Gratuit.com

# Annexe 1

Primers and selective PCR amplification products used for orthologous gene revelation in *E. globulus*, *E. urophylla* and *E. grandis*

| Sequence information |  | Primer information |                             |             |      |                                   |                          | Selective amplification            |                      |                      | CIRAD selective amplification      |              |         |           |                      |
|----------------------|--|--------------------|-----------------------------|-------------|------|-----------------------------------|--------------------------|------------------------------------|----------------------|----------------------|------------------------------------|--------------|---------|-----------|----------------------|
| EST Id.              | BLAST homology (putative function)                               | Primer - direction | Primer sequences (5' to 3') | Length (bp) | % GC | Position on original DNA sequence | Melting temperature (°C) | Optimal annealing temperature (°C) | DNA fragment         |                      | Optimal annealing temperature (°C) | DNA fragment |         |           | Observed length (bp) |
|                      |  |                    |                             |             |      |                                   |                          |                                    | Expected length (bp) | Observed length (bp) |                                    | Globulus     | Grandis | Urophylla |                      |
| GF#27                | xyloglucan endo-1, 4-beta-D-glucanase                            | GF#27 Upper        | TTA TGT GGA TGA AGT GCC C   | 19          | 47.4 | 96-114                            | 56.9                     | 58                                 | 434                  | 400<L<450            | 60                                 | S            | S       | S         | 400<L<450            |
|                      |  | GF#27 Lower        | AAC TTG AGC CTC CCA GTG T   | 19          | 52.6 | 512-530                           | 57.3                     |                                    |                      |                      |                                    | =            | =       | =         |                      |
| GF#86                | alpha-tubulin  | GF#86 Upper        | CTT GCA CCC TGA GCA ACT C   | 19          | 57.9 | 48-66                             | 58.4                     | 58                                 | 352                  | 350<L<400            | 60                                 | S            | S       | S         | 400<L<450            |
|                      |  | GF#86 Lower        | AAA GCA CAG CAA CAT CGG T   | 19          | 47.4 | 382-400                           | 59.3                     |                                    |                      |                      |                                    | <            | =       | =         |                      |
| 08A09                | No Hits  | 08A09 Upper        | GTACGCGGGGATATGACAAG        | 20          | 45.0 | 1-20                              | 58.0                     | 45                                 | 217                  | 200<L<250            | 50                                 | S            | S       | S         | 200<L<250            |
|                      |  | 08A09 Lower        | GCCCGTTAATGAAAAGCAC         | 20          | 55.0 | -217                              | 54.9                     |                                    |                      |                      |                                    | =            | =       | >         |                      |
| 08B02                | hypothetical protein [Methanosarcina mazei Goe1]                 | 08B02 Upper        | ATTTACGCGGGTGACCAAGA        | 20          | 50.0 | 27-46                             | 57.6                     | 49                                 | 224                  | 200<L<250            | 49                                 | S            | S       | S         | 200<L<250            |
|                      |  | 08B02 Lower        | ATGCCGGCTCAAGGTGTATCA       | 20          | 45.0 | 231-250                           | 58.8                     |                                    |                      |                      |                                    | =            | =       | =         |                      |
| 08B06                | similar to DNA repair protein RAD50, putative                    | 08B06 Upper        | TTTACGAGCCTCTTCCTTGA        | 20          | 45.0 | 61-80                             | 55.1                     | 58                                 | 326                  | 300<L<350            | 60                                 | S            | S       | S         | 300<L<350            |
|                      |  | 08B06 Lower        | AAGCAGCAGCTATTGTTTCC        | 20          | 45.0 | 367-386                           | 55.3                     |                                    |                      |                      |                                    | =            | <       | =         |                      |
| 08C02                | outer capsid protein   | 08C02 Upper        | GAAACTGTGTTCTCCCAAG         | 20          | 45.0 | 193-212                           | 54.3                     | 58                                 | 217                  | 200<L<250            | 60                                 | multiband    | S       | S         | 250<L<300            |
|                      |  | 08C02 Lower        | CACAAGATGCCTCAACAATG        | 20          | 45.0 | 450-469                           | 51.8                     |                                    |                      |                      |                                    | <&>          | =       | =         |                      |
| 08C05                | Contains similarity to peptidyl-prolyl cis-trans isomerase       | 08C05 Upper        | AGAAGGGCGGGTATTAGGA         | 20          | 50.0 | 24-43                             | 57.8                     | 58                                 | 347                  | 850<L<1000           | 58                                 | S            | S       | S         | 900<L<1000           |
|                      |  | 08C05 Lower        | GACCCATCAAGATGTCTCCAGT      | 22          | 50.0 | 349-370                           | 58.9                     |                                    |                      |                      |                                    | <            | =       | =         |                      |
| 08C08                | coat protein [Cucumber mosaic virus]                             | 08C08 Upper        | ACTCACCCAACTTCAACCAA        | 20          | 45.0 | 105-124                           | 55.0                     | 58                                 | 312                  | 300<L<350            | 60                                 | multiband    | S       | S         | 350<L<400            |
|                      |  | 08C08 Lower        | CGTTTGCCACGATGATTTT         | 20          | 40.0 | 397-416                           | 56.4                     |                                    |                      |                      |                                    | <, =, >      | =       | <         |                      |
| 08C10                | Plasmodium falciparum trophozoite antigen r45-like protein       | 08C10 Upper        | TTGATGTGCCGTATTCCTTT        | 20          | 45.0 | 29-48                             | 55.7                     | 58                                 | 259                  | 250<L<300            | 60                                 | S            | S       | multiband | 250<L<300            |
|                      |  | 08C10 Lower        | TTGTTGCCCTTTACACTCAC        | 20          | 40.0 | 268-287                           | 53.4                     |                                    |                      |                      |                                    | =            | =       | <&>       |                      |
| 08E04                | cellulase (EC 3.2.1.4) T21L14.7 - Arabidopsis thaliana           | 08E04 Upper        | GCCTGCACCTTACAATACCGC       | 20          | 45.5 | 345-364                           | 58.4                     | 58                                 | 236                  | 200<L<250            | 60                                 | S            | S       | S         | 200<L<250            |
|                      |  | 08E04 Lower        | GATTACTACACGGGCTTCACAA      | 22          | 55.0 | 561-580                           | 57.3                     |                                    |                      |                      |                                    | <            | =       | =         |                      |
| 08E06                | P0529E05.5 [Oryza sativa (japonica cultivar-group)]              | 08E06 Upper        | CAATGGGCTGTTGAGGTTCT        | 20          | 55.0 | 65-84                             | 57.3                     | 58                                 | 248                  | 200<L<250            | 60                                 | S            | S       | S         | 200<L<250            |
|                      |  | 08E06 Lower        | GTGTTGGAGGGTGAGGGTTA        | 20          | 50.0 | 293-312                           | 57.0                     |                                    |                      |                      |                                    | =            | <       | =         |                      |
| 08E11                | hypothetical protein - Arabidopsis thaliana                      | 08E11 Upper        | GGGGCCGAAACTCGCTTCTC        | 20          | 55.0 | 137-146                           | 65.7                     | 58                                 | 216                  | 200<L<250            | 58                                 | S            | S       | S         | 300<L<350            |
|                      |  | 08E11 Lower        | ATGGCATCACGCTCTGGTCT        | 20          | 65.0 | 333-352                           | 60.1                     |                                    |                      |                      |                                    | =            | =       | =         |                      |
| 08E12                | hypothetical protein F11C1.10 - Arabidopsis thaliana             | 08E12 Upper        | TTCAATAATCCCTGGTCAAT        | 20          | 45.0 | 5-24                              | 52.0                     | failed                             | 217                  | failed               | 49                                 | S            | S       | S         | 200<L<250            |
|                      |  | 08E12 Lower        | CCAAACGCACCTTAATAGTC        | 20          | 35.0 | 202-221                           | 52.8                     |                                    |                      |                      |                                    | =            | =       | >         |                      |
| 08F01                | No Hits  | 08F01 Upper        | ATAGAACACGCGTCCACT          | 20          | 35.0 | 77-96                             | 52.3                     | 58                                 | 128                  | 100<L<150            | 58                                 | S            | S       | S         | 100<L<150            |
|                      |  | 08F01 Lower        | GCATTTTGATCGCTTTGATT        | 20          | 45.0 | 185-204                           | 55.6                     |                                    |                      |                      |                                    | =            | =       | =         |                      |
| 08F03                | ceo protein [Arabidopsis thaliana]                               | 08F03 Upper        | GAGGAGGAGCAGAAGCAGG         | 20          | 40.0 | 102-121                           | 59.0                     | 58                                 | 358                  | 350<L<400            | 64                                 | S            | S       | S         | 350<L<400            |
|                      |  | 08F03 Lower        | AAATCAGCATGAAAGCGGTA        | 20          | 60.0 | 450-459                           | 56.1                     |                                    |                      |                      |                                    | =            | =       | <         |                      |
| 08F08                | No Hits  | 08F08 Upper        | ACCGCCGATTTACTCTTACT        | 20          | 40.0 | 18-37                             | 53.8                     | 45                                 | 113                  | 100<L<150            | 49                                 | S            | S       | S         | 100<L<150            |
|                      |  | 08F08 Lower        | TCCGATAATCGACTATTCCA        | 20          | 45.0 | 111-130                           | 53.1                     |                                    |                      |                      |                                    | =            | =       | =         |                      |
| 08F09                | hypothetical protein; 44869-47459 [Arabidopsis thaliana]         | 08F09 Upper        | GGGCACCTTTCAGGCACCACC       | 20          | 60.0 | 210-229                           | 64.7                     | 58                                 | 255                  | 250<L<300            | 58                                 | S            | S       | S         | 250<L<300            |
|                      |  | 08F09 Lower        | GCATCGCATCGTATCGGGAG        | 20          | 65.0 | 445-464                           | 63.6                     |                                    |                      |                      |                                    | =            | =       | =         |                      |
| 08G01                | hypothetical protein [Plasmodium yoelii yoelii]                  | 08G01 Upper        | CGAAGCGAGGACTGAAGAAG        | 20          | 50.0 | 135-154                           | 57.6                     | 58                                 | 250                  | 250<L<300            | 62                                 | S            | S       | multiband | 250<L<300            |
|                      |  | 08G01 Lower        | CGACCGATGATTTGATAGCC        | 20          | 55.0 | 375-384                           | 58.0                     |                                    |                      |                      |                                    | =            | =       | =&>       |                      |
| 08G05                | gamma-aminobutyric acid (GABA) A receptor, beta 1 [Homo sapiens] | 08G05 Upper        | AAAGTTTCCTGTATCTTCTCA       | 22          | 40.9 | 35-56                             | 53.1                     | failed                             | 290                  | failed               | failed                             | -            | -       | -         | -                    |
|                      |  | 08G05 Lower        | CGTGTATTTTCAGTTTGCCTCTC     | 22          | 36.4 | 303-324                           | 54.7                     |                                    |                      |                      |                                    | -            | -       | -         |                      |
| 08G06                | putative harpin-induced protein [Arabidopsis thaliana]           | 08G06 Upper        | GGGAGTAGATGGTGACCTGGAA      | 22          | 55.0 | 7-28                              | 60.4                     | 58                                 | 210                  | 200<L<250            | 58                                 | S            | S       | S         | 200<L<250            |
|                      |  | 08G06 Lower        | GCGGGAAGCATGAAAGCAAG        | 20          | 54.5 | 197-216                           | 63.2                     |                                    |                      |                      |                                    | =            | =       | =         |                      |



| Sequence information |  | Primer information |                             |             |      |                                   |                          | Selective amplification            |                      |                      | CIRAD selective amplification      |              |           |           |                      |
|----------------------|--|--------------------|-----------------------------|-------------|------|-----------------------------------|--------------------------|------------------------------------|----------------------|----------------------|------------------------------------|--------------|-----------|-----------|----------------------|
| EST Id.              | BLAST homology (putative function)                             | Primer - direction | Primer sequences (5' to 3') | Length (bp) | % GC | Position on original DNA sequence | Melting temperature (°C) | Optimal annealing temperature (°C) | DNA fragment         |                      | Optimal annealing temperature (°C) | DNA fragment |           |           |                      |
|                      |  |                    |                             |             |      |                                   |                          |                                    | Expected length (bp) | Observed length (bp) |                                    | Globulus     | Grandis   | Urophylla | Observed length (bp) |
| 08G08                | No Hits  | 08G08 Upper        | TGAAGCAAATCGAGTTGTTT        | 20          | 35.0 | 15-34                             | 53.5                     | failed                             | 121                  | failed               | 58                                 | S            | S         | S         | 100<L<150            |
|                      |  | 08G08 Lower        | TTTCAGCCAAAACATGTCAA        | 20          | 40.0 | 116-135                           | 54.8                     |                                    |                      |                      |                                    | =            | =         | =         |                      |
| 08G09                | PUTATIVE TRANSPOSASE NUMBER 3 FOR INSERTION SEQUENCE ISRM26    | 08G09 Upper        | TTTTGATGATACGGTGGAA         | 20          | 55.0 | 22-41                             | 53.9                     | failed                             | 279                  | failed               | failed                             | -            | -         | -         | -                    |
|                      |  | 08G09 Lower        | GAGGGACTCTGTGGTTCCTTG       | 20          | 40.0 | 381-300                           | 56.6                     |                                    |                      |                      |                                    | -            | -         | -         |                      |
| 08H01                | conserved hypothetical protein [Streptococcus pyogenes M1 GAS] | 08H01 Upper        | CGGGGAAACCTACTTGTCTAA       | 20          | 45.0 | 17-36                             | 52.1                     | 55                                 | 127                  | 100<L<150            | failed                             | -            | -         | -         | -                    |
|                      |  | 08H01 Lower        | GCATAAGTGAACACCAGGAT        | 20          | 50.0 | 124-143                           | 57.4                     |                                    |                      |                      |                                    | -            | -         | -         |                      |
| 08H04                | unknown [Glycine max]  | 08H04 Upper        | TGAAGTTCAGCGGCAAGAAG        | 20          | 40.9 | 80-99                             | 58.5                     | 58                                 | 293                  | 250<L<300            | 58                                 | S            | S         | S         | 250<L<300            |
|                      |  | 08H04 Lower        | TTGGAGGAAACAGCTAACCAAT      | 22          | 50.0 | 351-372                           | 59.0                     |                                    |                      |                      |                                    | =            | =         | >         |                      |
| 08H07                | hypothetical protein [Azotobacter vinelandii]                  | 08H07 Upper        | AATTTATTTGTTGAGGGTG         | 20          | 45.0 | 107-126                           | 53.6                     | 58                                 | 294                  | 250<L<300            | 53                                 | S            | S         | S         | 250<L<300            |
|                      |  | 08H07 Lower        | CAACGCATTGTTGAGTTAGT        | 20          | 35.0 | 381-400                           | 52.3                     |                                    |                      |                      |                                    | =            | =         | =         |                      |
| 08H11                | unknown [Arabidopsis thaliana]                                 | 08H11 Upper        | AAGCATCGGTTGCCAAGGAA        | 20          | 65.0 | 154-173                           | 63.7                     | 49                                 | 238                  | 200<L<250            | 58                                 | S            | S         | S         | 350<L<400            |
|                      |  | 08H11 Lower        | GCTGCTGTGCCAGGTGAAG         | 20          | 50.0 | 372-391                           | 62.8                     |                                    |                      |                      |                                    | =            | =         | =         |                      |
| 09A11                | unknown protein [Arabidopsis thaliana]                         | 09A11 Upper        | ATGACAAATAAGTCTATGGAAA      | 22          | 45.0 | 13-34                             | 53.3                     | 58                                 | 104                  | 1600<L<1700          | 50                                 | S            | S         | S         | 1200<L<1500          |
|                      |  | 09A11 Lower        | TACCGCATCCTCTACAAGAA        | 20          | 27.3 | 97-116                            | 49.2                     |                                    |                      |                      |                                    | =            | >         | =         |                      |
| 09C05                | No Hits  | 09C05 Upper        | GCCGAGCTGATGCAATAACT        | 20          | 50.0 | 37-56                             | 56.4                     | 58                                 | 169                  | 150<L<200            | 58                                 | S            | S         | S         | 150<L<200            |
|                      |  | 09C05 Lower        | GACCTTCTGGGTCTGAAA          | 20          | 50.0 | 186-205                           | 58.0                     |                                    |                      |                      |                                    | =            | =         | =         |                      |
| 09C08                | No Hits  | 09C08 Upper        | GTACATTGCCAAGGGGATAA        | 20          | 50.0 | 1-20                              | 54.9                     | 58                                 | 162                  | 150<L<200            | 58                                 | S            | S         | S         | 150<L<200            |
|                      |  | 09C08 Lower        | CAGTCCTTCCGAATACAGA         | 20          | 45.0 | 143-162                           | 55.2                     |                                    |                      |                      |                                    | =            | =         | <         |                      |
| 09C10                | Unknown protein [Oryza sativa (japonica cultivar-group)]       | 09C10 Upper        | TGGACGATGAGGACCAGAAG        | 20          | 50.0 | 96-115                            | 54.1                     | 58                                 | 266                  | 250<L<300            | 58                                 | S            | S         | S         | 250<L<300            |
|                      |  | 09C10 Lower        | GGTGAACATAATCAGGGAG         | 20          | 55.0 | 341-360                           | 57.9                     |                                    |                      |                      |                                    | =            | =         | =         |                      |
| 09D01                | trans(ormer serine/arginine-rich ribonucleoprotein, putative   | 09D01 Upper        | ATCGAGGAGGCATACGACC         | 20          | 63.2 | 194-213                           | 60.5                     | 62                                 | 316                  | 300<L<350            | 60                                 | multiband    | S         | S         | 100<L<150            |
|                      |  | 09D01 Lower        | CGGCGGAGGCAGAAATAGAC        | 20          | 57.9 | 190-509                           | 57.1                     |                                    |                      |                      |                                    | 250<L<300    | =         | <         |                      |
| 09D03                | subtilisin-like proteinase                                     | 09D03 Upper        | GAAGGATGGGGTTCACAAGG        | 20          | 40.0 | 19-38                             | 56.6                     | 58                                 | 273                  | 250<L<300            | 58                                 | S            | S         | S         | 250<L<300            |
|                      |  | 09D03 Lower        | CAACATGGATCAACCGAAAA        | 20          | 55.0 | 272-291                           | 59.0                     |                                    |                      |                      |                                    | =            | =         | =         |                      |
| 08D04                | No Hits  | 09D04 Upper        | GTACGGTTTGCATGATTGTT        | 20          | 35.0 | 1-20                              | 53.9                     | 49                                 | 112                  | 100<L<150            | 49                                 | S            | S         | S         | 100<L<150            |
|                      |  | 09D04 Lower        | TTAATCGCAATCCACAACAT        | 20          | 40.0 | 93-112                            | 53.4                     |                                    |                      |                      |                                    | =            | =         | =         |                      |
| 09E01                | No Hits  | 09E01 Upper        | AGCTGGTTATTCTTCAGAA         | 19          | 47.4 | 32-50                             | 51.5                     | 49                                 | 123                  | 100<L<150            | 49                                 | S            | S         | S         | 100<L<150            |
|                      |  | 09E01 Lower        | TAAGTGAAGCAACCGAGTG         | 20          | 36.8 | 135-154                           | 46.9                     |                                    |                      |                      |                                    | =            | <         | =         |                      |
| 09F08                | VI   | 09F08 Upper        | AAAGCACAAGGAGAAGAAA         | 20          | 40.0 | 21-40                             | 50.2                     | 58                                 | 332                  | 300<L<350            | 58                                 | S            | S         | S         | 300<L<350            |
|                      |  | 09F08 Lower        | CTCAGAATTAGATGCCAAGA        | 20          | 40.0 | 333-352                           | 54.4                     |                                    |                      |                      |                                    | =            | =         | =         |                      |
| 09H01                | MAP protein kinase [Arabidopsis thaliana]                      | 09H01 Upper        | CGAAATGATTATTGGGAAAC        | 20          | 45.0 | 19-30                             | 51.4                     | 54                                 | 323                  | 300<L<350            | 58                                 | S            | S         | S         | 400<L<450            |
|                      |  | 09H01 Lower        | ATCACTTCTATGGACCGTGT        | 20          | 35.0 | 314-333                           | 52.6                     |                                    |                      |                      |                                    | =            | =         | =         |                      |
| 09H02                | hypothetical protein XP_243418 [Rattus norvegicus]             | 09H02 Upper        | ACTTTGCTTGTGGTATTTC         | 20          | 40.0 | 3-22                              | 53.1                     | 58                                 | 210                  | 200<L<250            | 58                                 | S            | S         | S         | 200<L<250            |
|                      |  | 09H02 Lower        | AATTCCTTCCGACTTGACTT        | 20          | 35.0 | 194-212                           | 49.7                     |                                    |                      |                      |                                    | =            | =         | =         |                      |
| 09H05                | putative translation initiation factor                         | 09H05 Upper        | GGACAGAGCTTCAGCAGATT        | 20          | 40.0 | 16-35                             | 52.3                     | 58                                 | 228                  | 200<L<250            | 53                                 | S            | S         | S         | 250<L<300            |
|                      |  | 09H05 Lower        | CCACTTACTCCAAGCAATT         | 20          | 50.0 | 224-243                           | 54.2                     |                                    |                      |                      |                                    | <            | >         | =         |                      |
| 09H06                | cell surface protein precursor [Lactobacillus plantarum WCFS1] | 09H06 Upper        | GGCATAGTCGATCACATTAG        | 20          | 50.0 | 13-52                             | 54.4                     | 58                                 | 278                  | 250<L<300            | 60                                 | multiband    | multiband | S         | 250<L<300            |
|                      |  | 09H06 Lower        | AGTCGCCTGAGATAAATGG         | 20          | 45.0 | 271-290                           | 50.4                     |                                    |                      |                      |                                    | =&>          | <&>       | =         |                      |

S : specific amplification  
- : no amplification  
comparison, for a same primer, of length fragment  
= : fragment with the same length for genotypes or middle fragment  
> : fragment with length longer than other genotypes  
< : fragment with length smaller than other genotypes

# Annexe 2



## Résumé

Dans le cadre de l'amélioration génétique des eucalyptus au Congo, le CIRAD Forêt a engagé des recherches visant à utiliser des marqueurs moléculaires en sélection. Des travaux de cartographie et de détection de QTL (Quantitative trait loci) pour la qualité du bois ont été entrepris sur un croisement interspécifique *E. urophylla* x *E. grandis* constitué par une famille de plein frères de 201 individus. Des cartes génétiques ont été construites à partir de marqueurs dominants (RAPD) et de marqueurs codominants (SSR et gènes candidats) pour ce croisement. Récemment, un projet mené en collaboration avec le CNRS de Toulouse et des laboratoires privés espagnol (ENCE) et portugais (RAIZ) a permis la création d'une banque d'EST d'*E. gunnii* basée sur leur expression différentielle entre deux tissus (xylème contre feuille) et dans différentes conditions de xylème en différenciation. L'objectif de cette collaboration est de localiser ces EST sur différentes cartes génétiques et de les colocaliser avec des QTL de qualité du bois. Après l'optimisation des conditions d'amplification, le polymorphisme d'un jeu de 33 nouvelles EST et deux gènes candidats a été révélé sur la descendance au moyen de plusieurs techniques : SSCP, CAPS et le séquençage de produits PCR. Au total, 19 séquences polymorphes ont été cartographiées grâce à la SSCP, une EST a révélé du polymorphisme en CAPS et 5 autres présentent des SNP sur leur séquence. Ce travail a permis de positionner sur les cartes génétiques d'*E. grandis* et d'*E. urophylla* de nouveaux marqueurs codominants dans l'objectif de déterminer des gènes candidats colocalisant avec des QTL et d'aborder les aspects de synténie au sein du genre *Eucalyptus*.

## Abstract

Within the framework of the genetic improvement of the eucalypti in Congo, CIRAD Forêt engaged research aiming at using molecular markers in selection. Genetic mapping and quantitative trait loci (QTL) analyses for wood quality were undertaken on a F1 interspecific progeny from *E. urophylla* x *E. grandis*. Genetic linkage maps have been constructed using dominant markers (RAPD) and codominant markers (SSR and candidate genes) for this crossing. Recently, a project carried out in collaboration with the CNRS of Toulouse and the private Spanish (ENCE) and Portuguese (RAIZ) laboratories, allowed the creation of an *E. gunnii* EST bank based on the isolation of xylem preferentially expressed genes (secondary xylem / leaves) and under various conditions of xylem in differentiation. The aim of this collaboration is to locate these EST on different genetic maps and to colocalise them with QTL for wood quality. After the optimization of the amplification conditions, the polymorphism of a set of 33 new EST and two candidate genes were revealed on the progeny by several techniques: SSCP, CAPS and sequencing of PCR products. A number of 19 polymorphic sequences were mapped thanks to the SSCP, one EST has shown polymorphism with the CAPS method and 5 others have SNP on their sequence. This work allowed the positioning of new co-dominant markers on the genetic maps of *E. grandis* and *E. urophylla* in order to determine candidate genes, as wood quality QTLs are localized on these DNA markers loci. These linkage maps provide a basis for investigating genome organisation and identifying synteny among the *Eucalyptus* genus.