HAL
open science

Plasticité moléculaire de deux écotypes de pin maritime soumis à un stress osmotique

Philippe Chaumeil

To cite this version:

Philippe Chaumeil. Plasticité moléculaire de deux écotypes de pin maritime soumis à un stress osmotique. Biologie végétale. Université Henri Poincaré - Nancy 1, 2006. Français. NNT : 2006NAN10016 . tel-01748169

HAL Id: tel-01748169

https://hal.univ-lorraine.fr/tel-01748169

Submitted on 29 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE LORRAINE

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

FACULTE des SCIENCES et TECHNIQUES
U.F.R. Sciences et Techniques Biologiques
Ecole Doctorale : Ingénieries des Ressources, Procédés, Produits et Environnement
Département de Formation Doctorale : Biologie Végétale et Forestière

Thèse

Présentée pour l'obtention du titre de

Docteur de l'Université Henri Poincaré, Nancy I

en Biologie Végétale et Forestière

par **Philippe CHAUMEIL**

Plasticité moléculaire de deux écotypes de pin maritime soumis à un stress osmotique

Soutenance publique le 13 Avril 2006

Membres du jury :

Président :	P. DIZENGREMEL	Professeur des Universités, U.H.P., Nancy I
Rapporteurs :	B. GARBAY	Professeur des Universités, ESTBB, Bordeaux
	M. ZIVY	Chargé de Recherche, CNRS, Gif sur Yvette
Examinateurs :	O. BRENDEL	Chargé de Recherche, INRA Nancy
	P. LABEL	Directeur de Recherche, INRA Orléans
Directeur de thèse :	C. PLOMION	Directeur de recherche, INRA Bordeaux

INRA-UMR BIOGECO – Equipe de Génétique – 33612 CESTAS Cedex – France

Remerciements

~

A la fois introduction et conclusion de la thèse, ils permettent avant tout de mettre l'accent sur l'aspect humain de tout projet de recherche qui est l'aboutissement de nombreuses contributions en particulier dans le cadre de cette thèse. Je tiens à remercier ici toutes les personnes qui ont participé à son élaboration.

- Antoine Kremer qui m'a accueilli dans son laboratoire où j'ai eu la chance d'effectuer mon DEA et de poursuivre en thèse. Tous ceux qui sont un jour passés à «Pierroton » ont remarqué son mode de fonctionnement ouvert aux échanges entre personnes et entre équipes qui est si enrichissant.
- Christophe Plomion qui a dirigé cette thèse, pour la confiance qu'il m'a accordée et la liberté dans le choix d'orientation et de réalisation de mes axes de recherche. Je tiens à souligner la façon dont Christophe sait valoriser, motiver et développer les compétences des membres de son équipe. Merci pour ta réactivité, la disponibilité dans les moments délicats et surtout d'avoir rendu possible de mener à son terme cette thèse. Merci également de savoir rester proche des étudiants et pour les bons moments passés à Salles.
- Concernant ce mémoire, je tiens à remercier Bertrand Garbay et Michel Zivy d'avoir accepté d'en être les rapporteurs.
- Oliver Brendel, Philippe Label et Pierre Dizengremel d'avoir accepté de faire partie de ce jury.
- Sébastien Duplessis, Virginie Garcia, Annick Moing, Christophe Rothan pour leur participation active aux comités de thèse.

~

Pour leur contribution directe à la réalisation de ce travail, je voudrais remercier :

- Patrick Léger pour toutes les discussions constructives, les améliorations dans le fonctionnement du labo et sans qui nous n'aurions pas pu générer les ESTs de racines.
- Aurélien Barré, Nacer Mohellibi et l'équipe du CBiB qui ont réalisé le pipeline d'annotation et ont répondus à mes nombreux coups de téléphones.
- Christophe Boury aka « Le Jeune » pour les heures passées à découvrir les joies de l'extraction, la qPCR et surtout l'annotation! Je conserverai également un bon souvenir des soirées « barbec » et de tes talents artistiques de montage photo qui font encore partie des murs ou plutôt des portes !
- Marie Foulongne avec qui j'ai eu plaisir à travailler et à discuter … des extractions d'ARN, des puces à ADN et du reste !
- Virginie Garcia et Johan Petit pour avoir partager leur bureau, leur bonne humeur, leurs conseils et leur expertise sur la réalisation et le traitement des données des puces.
- Manon Moreau qui a réalisé avec Céline Lalanne l'analyse du protéome (alors, ça polymérise jamais dans les tuyaux ? ;-)) et qui est devenue accro à la qPCR.
- Silvia Fluch (société PICME) qui a réalisé les puces.
- Jean-Marc Frigério qui a réalisé les banques de racines et d'aiguilles m'a convertit au Perl ("=~('(?{'.('+[@.)}'^'])@]]').'"'.('+(@,/^}'^'[@)@@+_').',$/})')) mais pas encore au Pingouin…
- L'équipe américaine de Ron Sederoff qui m'a accueilli chaleureusement.

Je n'oublie pas non plus tous ceux qui ont su répondre présents pour m'aider tout au long de la thèse et ceux qui sont devenus plus que des collègues de travail.

En particulier Céline qui doit toujours se demander si je ne vais pas tomber dans les pommes ;-) ; Manon pour sa gentillesse et la bonne humeur qu'elle sait apporter autour d'elle ; Agnès et David Chagné pour les fous rires et l'ambiance inimitable qu'ils mettaient dans les labos ; MH qui m'a ouvert les portes de son club ldp VIP et m'a soutenu dans les moments difficiles ; HCl puis Franck Salin avec qui j'ai partagé mon goût pour la bidouille informatique ce qui nous a valu le surnom de « Megabytes » … ; Fred qui a le mérite de supporter mes visites inopinées dans son bureau, la petite tape dans le dos et qui m'a permis de lâcher la pression ; Greg qui m'a initié à l'ARN et m'a fait découvrir le surfcasting et Jeannot pour ses blagues et la pêche sur le bassin.

Je remercie également Jeremy avec qui j'ai partagé le bureau et qui a supporté ma vision radicalement différente d'un bureau de travail. Je m'incline devant ta capacité à conserver un bureau parfaitement rangé et classé. Merci aussi pour ta bonne humeur et ton sens de l'humour. Et n'oublies pas, selon le vent, la voile :-D.

Mes remerciements iront également à Jérôme et à Jorge pour le soutient moral entre étudiants en fin de thèse, Manuela pour les discussions tardives, Carlos et sa petite famille pour les soirées Pisco. Je remercie aussi les « secrétaires » Annie, Corinne et Florence qui tentent de nous décoder les mystères administratifs. Merci à Annie et Valérie pour leur maîtrise de l'orthographe et de la grammaire. Merci à tous les permanents, non permanents et visiteurs qui ont contribué à ce que je ne vois pas passer ces 5 années à pierroton. Il m'est impossible de les citer tous, j'espère que les bretons et non bretons (hein JMG !) se reconnaîtront, de même que les argentins, australiens, brésiliens, chiliens, espagnols, italiens, japonais, mexicains (de souche ou de cœur) et portugais.

Une dédicace spéciale à Dje et Florence de Nancy car ce sont de vrais Amis que je me suis fait là bas et avec qui j'ai partagé des moments forts.

Un clin d'œil de PKM à la team oncle Albert, je remercie Cyborg pour avoir réussi à rapprocher deux électrons libres, Pikatchu pour ses connaissances de la plateforme de Toulouse ;-) et LoL qui est allé se perdre au pays du burger.

Je terminerai en remerciant mes parents et ma famille qui m'ont permis d'arriver jusqu'ici et surtout Gwen, Sterenn a Garantez, qui a supporté au quotidien les hauts comme les bas depuis maintenant 4 ans et sans qui je ne conçois plus mon avenir.

Abréviations

~

Liste des principales abréviations utilisées :

ABA : Acide abscissique
ACC : 1-aminocyclo-propane-1-carboxylic acid
ADN / DNA: Acide Desoxyribonucléique
ADNc : ADN complémentaire
AGP : Arabinogalactan proteins
ARN / RNA : Acide Ribonucléique
ARNm : ARN messagers
ARNt : ARN totaux
ASR : ABA-, Stress-, and Ripening-induced protein
Blast : Basic Local Alignment Search Tool
CCoAOMT : Caffeoyl Coenzyme A 3-O-methyltransferase
CN: Consensus
CL : Cluster
CT : Contig
COMT : Catechol-O-methyltransferase
EF : Ecotype Français
EM : Ecotype Marocain
EST : Expressed Sequences Tags
FT : Facteur de Transcription
GA : Gibbérelline(s)
Hsp / HSP : Heat-shock proteins
LEA : Late embryogenesis Abundant protein
miRNA : microARN
PA : Polyamine(s)
PAL : Phenylalanine ammonia-lyase
PCR : Polymersae Chain Reaction
PEG : Polyéthylène glycol
QTL : Qunatitative Trait Loci
qPCR : PCR quantitative en temps réel
RFO : Raffinose family oligosaccharide(s)
SC : Expérience « Stress Court »
SL : Expérience « Stress Long »
SNP : Single Nucleotide Polymorphism
SOD : Superoxyde Dismutase
TC : Tentative Consensus
XTH / XET : Xyloglucan endotransglucosylase/hydrolases
4CL : 4-coumarate-CoA ligase

Sommaire

A INTRODUCTION ---------- ***2***

A.1 Un climat en cours de modification ---------- **2**

A.2 Hydrologie et géologie de la région landaise ---------- **3**

A.3 Des marais à la sécheresse, un équilibre fragile ---------- **4**

A.4 Des incertitudes sur la durabilité du massif forestier aquitain ---------- **5**

A.5 Objectifs ---------- **6**

A.6 Plan de la thèse ---------- **8**

B SYNTHESE BIBLIOGRAPHIQUE ---------- ***12***

B.1 Le Pin maritime ---------- **12**

B.1.1 Botanique et Taxonomie ---------- 12

B.1.2 L'appareil racinaire ---------- 14

B.1.3 Répartition géographique ---------- 15

B.1.4 Habitat ---------- 16

B.2 Les plantes et la sécheresse ---------- **17**

B.2.1 L'eau et la plante ---------- 17

B.2.1.1 L'eau du sol ---------- 17

B.2.1.2 L'eau dans la plante ---------- 18

B.2.1.3 L'eau et les racines ---------- 19

B.2.2 Le stress hydrique ---------- 22

B.2.2.1 Les stratégies de réponses ---------- 22

B.2.2.2 Les systèmes modèles ---------- 23

B.2.2.2.1 Systèmes tolérants ---------- 23

B.2.2.2.2 Modèles génétiques ---------- 25

B.2.2.2.3 Plantes de grande culture ---------- 26

B.2.2.3 Réponses morphologiques et physiologiques ---------- 26

B.2.2.3.1 Régulation de la croissance ---------- 26

B.2.2.3.2 Croissance racinaire ---------- 28

B.2.2.3.3 Les hormones ---------- 30

B.2.2.3.3.1 Acide abscissique ---------- 31

B.2.2.3.3.2 Ethylène ---------- 34

B.2.2.3.3.3 Polyamines (PA) ---------- 36

B.2.2.3.3.4 Les auxines ---------- 38

B.2.2.3.3.5 Les Gibbérellines ---------- 38

B.2.2.3.3.6 Cytokinines ---------- 39

B.2.2.4 Taille du génome et adaptation au stress hydrique ---------- 39

B.2.2.5 Gènes et voies métaboliques régulées au cours du stress ---------- 40

B.2.2.5.1 Méthodes de détection des gènes régulés ---------- 41
B.2.2.5.2 Détoxication ---------- 42
B.2.2.5.3 Fonctions chaperonne ---------- 43
B.2.2.5.3.1 Les Hsp (Heat-shock proteins) ---------- 43
B.2.2.5.3.2 Les LEA (Late Embryogenesis Abundant Proteins) ---------- 45
B.2.2.5.4 Ajustement osmotique ---------- 46
B.2.2.5.4.1 Proline ---------- 47
B.2.2.5.4.2 Glycine bétaïne ---------- 48
B.2.2.5.4.3 Sucres et dérivés ---------- 48
B.2.2.5.5 Transport de l'eau et des solutés ---------- 52
B.2.2.5.6 Modification structurale de la cellule ---------- 54
B.2.2.5.6.1 Modification de la paroi ---------- 55
B.2.2.5.6.2 Dégradation de la paroi ---------- 56
B.2.2.5.6.3 Biosynthèse de la paroi ---------- 56
B.2.2.5.6.4 Cytosquelette ---------- 57
B.2.2.5.7 Régulation de la transcription ---------- 57
B.2.2.5.7.1 proteines (b-ZIP) ---------- 58
B.2.2.5.7.2 Protéines Myb-like et Myc-like bHLH ---------- 58
B.2.2.5.7.3 Protéines AP2(APETALA2)/EREBP ---------- 59
B.2.2.5.7.4 Protéines WRKY ---------- 60
B.2.2.5.7.5 Autres Protéines ---------- 61
B.2.2.5.7.6 Histones et Méthylations ---------- 61
B.2.2.5.7.7 Autres mécanismes de régulation ---------- 62
B.2.2.5.8 Perception et transduction du signal ---------- 63
B.2.2.5.8.1 Perception du signal ---------- 63
B.2.2.5.8.2 Calcium et second messagers ---------- 64
B.2.2.5.8.3 Cascade des MAPKs ---------- 65
B.2.2.5.8.4 Protéines régulées par le Calcium ---------- 65

B.3 Le stress hydrique chez les arbres ---------- 65
B.3.1 Une problématique spécifique ---------- 65
B.3.1.1 Particularités physiologiques ---------- 65
B.3.1.2 Ressources moléculaires ---------- 66
B.3.2 Le stress hydrique chez le pin maritime ---------- 68
B.3.2.1 Les approches physiologiques ---------- 68
B.3.2.2 Les Approches génétiques et moléculaires ---------- 71
B.3.2.2.1 Cartographie et détection de QTL ---------- 71
B.3.2.2.2 Identification des protéines impliquées dans la réponse au stress hydrique ---------- 72
B.3.2.2.3 Identification des transcrits impliqués dans la réponse au stress hydrique ---------- 73

B.4 Stratégies et techniques pour la détection des gènes impliqués dans l'adaptation du pin maritime à la sécheresse. ----------74

C MATERIEL ET METHODES ----------78

C.1 Matériel végétal et conditions de culture----------78

C.1.1 Germination---------- 78

C.1.2 Culture Hydroponique---------- 79

C.1.3 Contrôle des pathogènes ---------- 79

C.1.4 Eclairage ---------- 80

C.1.4.1 Photosynthèse ---------- 80

C.1.4.2 Action Phytochromique---------- 81

C.1.4.3 Sources ---------- 81

C.1.5 Application du stress ---------- 82

C.1.5.1 Stress court ---------- 82

C.1.5.2 Stress long ---------- 82

C.1.6 Récolte du matériel végétal---------- 83

C.2 Mesure du niveau de stress et paramètres physiologiques ----------83

C.2.1 Mesure du potentiel de base ---------- 83

C.2.2 Dosage des terpènes ---------- 84

C.3 Analyse du transcriptome et caractérisation des EST ----------84

C.3.1 Extraction des ARN ---------- 84

C.3.2 Banques d'ADNc ---------- 85

C.3.2.1 Banque de xylème ---------- 85

C.3.2.2 Banques de racines et aiguilles---------- 86

C.3.3 Obtention des séquences ---------- 86

C.4 Analyse d'expression----------87

C.4.1 Synthèse des ADNc---------- 87

C.4.2 Microréseaux à ADNc---------- 88

C.4.2.1 Réalisation de l'Unigène---------- 88

C.4.2.2 Dépôt des sondes sur lame de verre ---------- 90

C.4.2.3 Plan d'expérience ---------- 91

C.4.2.4 Synthèse des cibles ---------- 91

C.4.2.5 Préparation, hybridation et lecture des lames ---------- 92

C.4.2.6 Traitement des données ---------- 93

C.4.2.7 Identification des gènes différentiellement exprimés ---------- 94

C.4.2.8 Agrégation des données (Clustering)---------- 95

C.4.3 PCR quantitative en temps réel---------- 97

C.4.3.1 Experimentation---------- 97

C.4.3.2 Choix des amorces---------- 98

C.4.3.3 Choix des gènes de référence ---------- 99
C.4.3.4 Traitement des données de fluorescence ---------- 99

C.5 Analyses Protéiques ---------- 100
C.5.1 Extraction des protéines et Electrophorèse bidimensionnelle ---------- 100
C.5.2 Analyse des gels ---------- 101
C.5.3 Analyse quantitative ---------- 101
C.5.4 Choix des spots à séquencer ---------- 101
C.5.5 Séquençage et identification des protéines ---------- 102

C.6 Analyses Bioinformatiques ---------- 103
C.6.1 Analyse et Assemblage des séquences ---------- 103
C.6.2 Annotation des Consensus et Singletons ---------- 105
C.6.3 Niveau d'expression *in silico* des EST (Northern electronique) ---------- 106
C.6.4 Annotation des spots protéiques ---------- 106

D RESULTATS ET DISCUSSION ---------- 109

D.1 Matériel végétal et paramètres physiologiques ---------- 109
D.1.1 Culture Hydroponique ---------- 109
D.1.2 Potentiel de base ---------- 109
D.1.3 Dosage des terpènes ---------- 110
D.1.4 Discussion ---------- 111

D.2 Analyse et annotation des séquences ---------- 115
D.2.1 Séquençage ---------- 115
D.2.2 Assemblage des séquences ---------- 117
D.2.3 Annotation des séquences ---------- 120
D.2.3.1 Annotation des consensus et des singletons ---------- 120
D.2.3.2 Assignation des catégories fonctionnelles ---------- 124
D.2.4 Northern électronique ---------- 125
D.2.5 Unigène ---------- 129
D.2.6 Discussion ---------- 131

D.3 Analyse du transcriptome ---------- 134
D.3.1 Extraction des ARN totaux ---------- 134
D.3.2 Puces à ADNc ---------- 134
D.3.2.1 Caractéristiques des lames produites et mise en œuvre ---------- 135
D.3.2.1.1 Qualité des dépôts d'ADNc ---------- 135
D.3.2.1.2 Synthèse des cibles et hybridation ---------- 136
D.3.2.1.3 Acquisition des données ---------- 136
D.3.2.2 Traitement des données brutes de fluorescence ---------- 137
D.3.2.2.1 Préparation et étude descriptive des données ---------- 137
D.3.2.2.2 Correction des données ---------- 139

D.3.2.2.3 Ajustement des données ---------- 140
D.3.2.3 Analyse de l'expression différentielle ---------- 142
D.3.2.3.1 Choix des paramètres statistiques ---------- 142
D.3.2.3.2 Quels sont les gènes différentiellement exprimés ? ---------- 143
D.3.2.3.3 Quelles sont les fonctions des gènes détectés ? ---------- 144
D.3.2.3.3.1 Communication / Transduction du signal ---------- 145
D.3.2.3.3.2 Croissance cellulaire, division ---------- 145
D.3.2.3.3.3 Devenir des protéines ---------- 146
D.3.2.3.3.4 Energie ---------- 147
D.3.2.3.3.5 Métabolisme ---------- 147
D.3.2.3.3.6 Organisation cellulaire ---------- 149
D.3.2.3.3.7 Réponse au Stress ---------- 150
D.3.2.3.3.8 Synthèse de protéines ---------- 152
D.3.2.3.3.9 Trafic intracellulaire ---------- 153
D.3.2.3.3.10 Transcription ---------- 153
D.3.2.3.3.11 Transport ---------- 154
D.3.2.3.3.12 Non-classé ---------- 155
D.3.2.3.4 Bilan des principales fonctions régulées ---------- 155
D.3.2.3.5 Les gènes détectés sont-ils les même dans les deux expériences ? ---------- 157
D.3.2.4 Regroupement et classification des données d'expression ---------- 158
D.3.2.4.1 Organisation et profil d'expression des groupes générés ---------- 158
D.3.2.4.2 Plasticité moléculaire et gènes d'intérêt ---------- 161
D.3.2.4.3 Profils d'expression et fonctions associées ---------- 163
D.3.2.4.3.1 Gènes dont l'expression diffère entre écotypes ---------- 163
D.3.2.4.3.2 Gènes dont l'expression dépend de la condition ---------- 165
D.3.2.4.3.3 Gènes répondant différemment entre écotypes lors du stress ---------- 166
D.3.2.4.3.4 Gènes dont le profil d'expression est atypique ---------- 169
D.3.3 PCR quantitative en temps réel ---------- 170
D.3.3.1 Choix des gènes ---------- 170
D.3.3.2 Niveau d'expression des gènes ---------- 172
D.3.3.2.1 Gènes vérifiés conjointement en stress court et stress long ---------- 172
D.3.3.2.2 Gènes vérifiés en stress long ---------- 176
D.3.3.2.3 Gènes vérifiés en stress court ---------- 177
D.3.3.2.4 Gènes de la biosynthèse de l'éthylène ---------- 178
D.3.4 Discussion ---------- 179

D.4 Analyse du protéome ---------- 183
D.4.1 Détection des protéines ---------- 183
D.4.2 Identification des spots protéiques différentiellement exprimés ---------- 183
D.4.3 Annotation des spots protéiques ---------- 184
D.4.4 Expression du protéome et du transcriptome ---------- 185

D.4.4.1 « Métabolisme des acides aminés » ---------- 185
D.4.4.2 « Biosynthèse des métabolites secondaires »---------- 186
D.4.4.3 « Métabolisme » ---------- 187
D.4.4.4 « Cytosquelette »---------- 188
D.4.4.5 « Réponse au stress »---------- 188
D.4.4.6 Autres catégories---------- 189
D.4.5 De la protéine au gène : cas de l'ACC désaminase ---------- 189
D.4.6 Discussion---------- 191

E CONCLUSIONS ET PERSPECTIVES---------- 194

E.1 Développement et utilisation d'une approche à haut débit pour l'étude de la réponse au stress hydrique.---------- 194

E.2 Perspectives de cette étude ---------- 197

E.2.1 Exploitation des ressources mises en place ---------- 197
E.2.2 Potentiel adaptatif des gènes candidats---------- 198
E.2.3 De l'organe à la cellule---------- 199
E.2.4 Une autre voie de régulation : les micros ARN ---------- 199
E.2.5 Contrôle épigénétique ---------- 199

BIBLIOGRAPHIE----------201

ANNEXES----------217

Introduction

A INTRODUCTION

A.1 Un climat en cours de modification

Depuis quelques années, des études ont été menées pour vérifier qu'une évolution rapide du climat était en cours au niveau mondial. Que ce soit par l'enregistrement et l'analyse de relevés de précipitations, de température, de concentration atmosphérique en différentes molécules ou par des relevés et modélisations météorologiques, l'ensemble des facteurs étudiés confirme une évolution rapide et globale du climat. En 1998, l'organisation mondiale de météorologie (WMO) et le programme environnement des Nations Unies ont constitué une structure intergouvernementale pour étudier les changements climatiques (IPCC : http://www.ipcc.ch/).

Si au cours du dernier millénaire les températures globales de la surface terrestre sont restées stables (Fig. A-1), on assiste à une augmentation nette au cours des 100 dernières années. La même tendance est observée pour la concentration atmosphérique en CO_2, N_2O, CH_4 et SO_4^{2-} produits par l'activité humaine (Fig. A-2). En France, l'augmentation constatée des températures moyennes pour la région Aquitaine est de 1.1°C et celle des maximales de 0.9°C (données Météo France Fig. A-3 ; A-4). Les années 2001 à 2005 ont été marquées par des déficits de précipitations, des sécheresses estivales et des températures au-dessus de la normale (Fig. A-5). L'année 2003 a été particulièrement chaude en Europe, surtout dans le Sud-ouest. Des études récentes (Schar and Jendritzky 2004; Schar *et al.* 2004) essayant de déterminer si un tel phénomène est de l'ordre de l'exceptionnel tendent à montrer que plus de la moitié du risque d'un tel évènement est dû à un changement anthropique du climat.

Les modèles de prévisions climatiques pour le siècle à venir proposés par le GIEC (Groupe International d'Experts sur le Climat : http://www.ipcc.ch/) prévoient une augmentation de température de l'ordre de 2-3 °C à 3-4 °C pour la France avec une augmentation nettement plus marquée pour la saison chaude. En ce qui concerne les précipitations, aucun changement n'est prévu pour la saison hivernale, mais la saison estivale devrait connaître des précipitations nettement plus faibles ou incertaines selon les scénarios (Fig. A-6). Si les modélisations à l'échelle mondiale sont incertaines, c'est d'autant plus le cas à une échelle régionale. Nous pouvons toutefois nous attendre à une augmentation des températures avec des saisons plus contrastées, ainsi qu'une augmentation de la variabilité du climat et de certains phénomènes extrêmes (GIEC 2001).

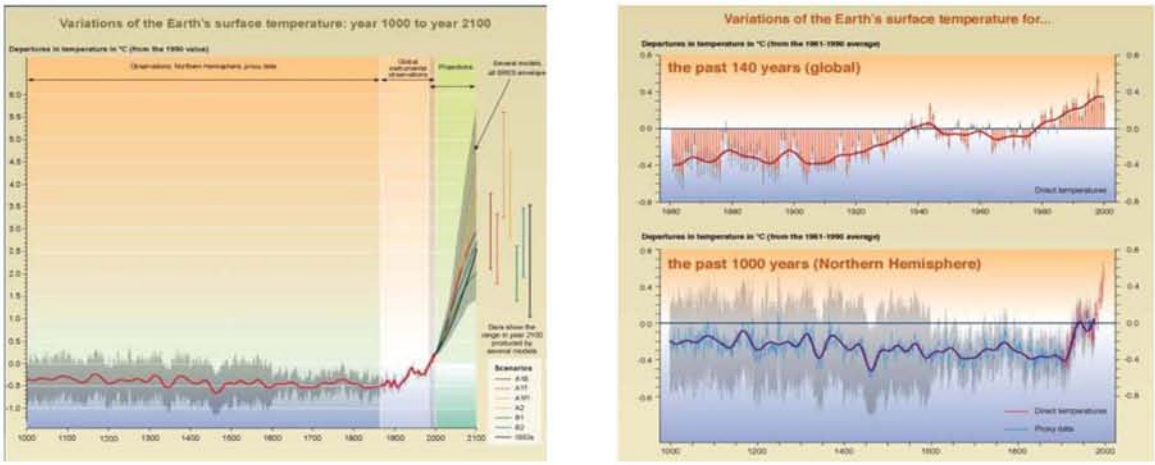


Figure A-1 : Evolution de la température de la surface terrestre, source IPCC.

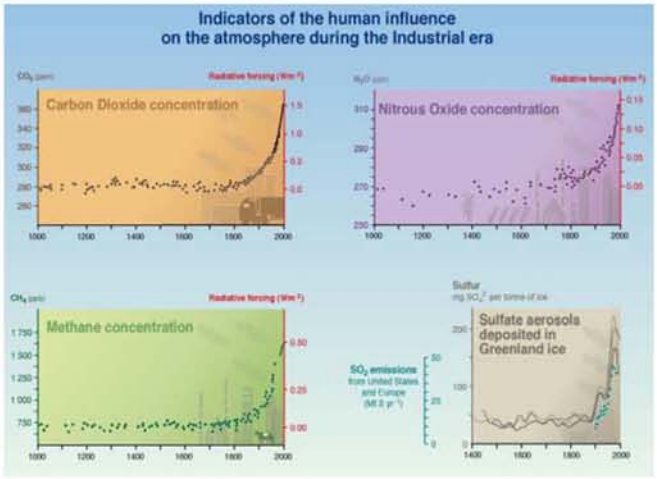


Figure A-2 : Concentration en CO_2, N_2O, CH_4 et SO_4^{2-}, source IPCC.

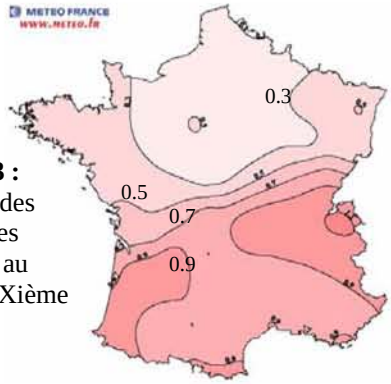


Figure A-3 : Evolution des températures maximales au cours du XXième siècle.

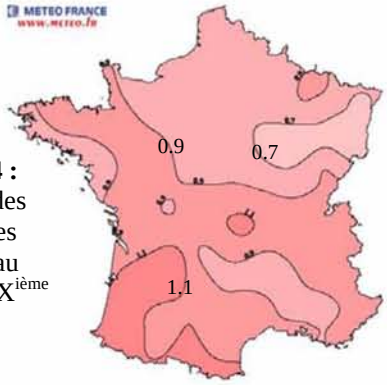
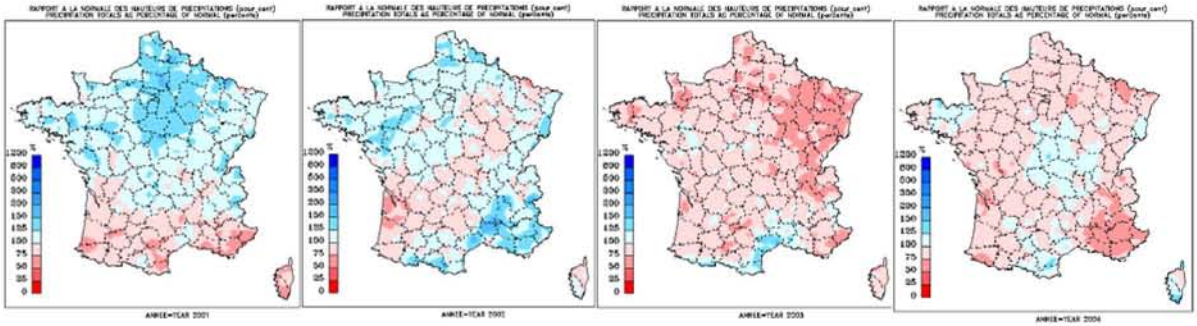


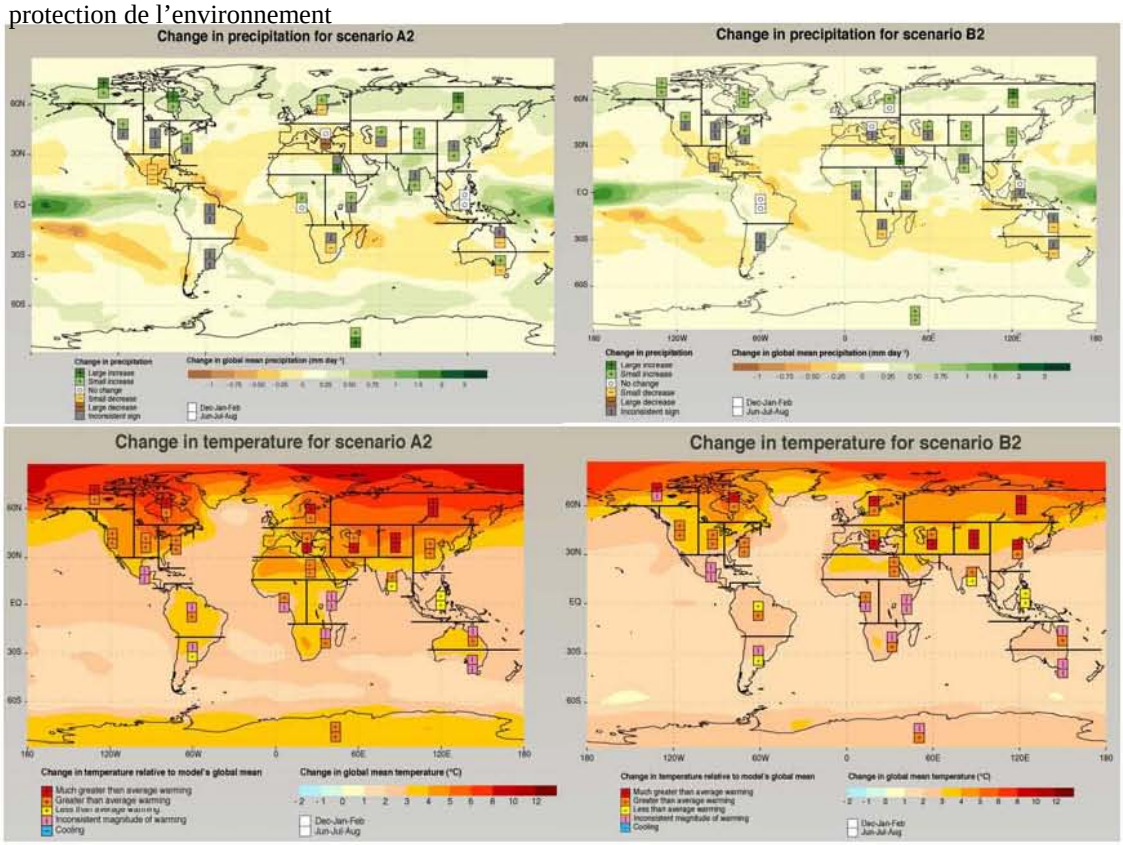
Figure A-4 : Evolution des températures moyennes au cours du XXième siècle.



2001 **2002** **2003** **2004**

Figure A-5 : Evolution des précipitations en France au cours des années 2001 à 2004 (données Météo-France).

Figure A-6 : Prévisions de l'évolution des précipitations et températures au XXIième siècle :
Scenario A2 = monde hétérogène, démographie converge progressivement, accroissement continu de la population, développement économique plus lent que dans les autres scénarios et régionalisé.
Scenario B2 = monde à fonctionnement régionalisé, démographie à accroissement continu de la population plus faible que A2, développement économique régional hétérogène, évolution technologique plus faible que A2, protection de l'environnement



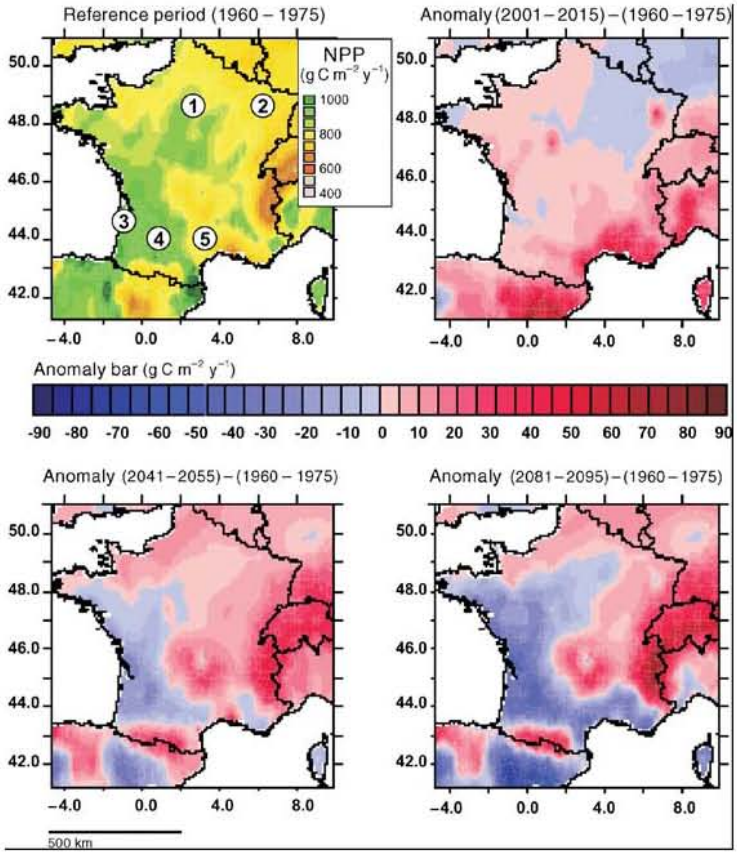


Figure A-7 : Productivité Primaire Annuelle Nette (NPP) des feuilles et aiguilles du système tempéré en France (en haut à gauche) et les prévisions d'évolution de la NPP en 2008, 2048 et 2088 (respectivement haut-droit, bas-gauche et bas-droit). D'après Lousteau *et al.* (2005).

L'évolution des conditions climatiques aura des répercussions directes sur le développement et la croissance des végétaux et en particulier sur la productivité des forêts cultivées (Fig. A-7) (Loustau *et al.* 2005). La question est d'autant plus sensible en sylviculture que les semences plantées aujourd'hui donneront des arbres qui - contrairement aux cultures annuelles - devront faire face aux évolutions climatiques des 50 à 100 prochaines années.

Face à des changements climatiques lents, les écosystèmes forestiers peuvent évoluer grâce à leur plasticité et à leur diversité génétique. Cependant, leur adaptation à des perturbations brutales n'est pas du tout aussi certaine. Si les teneurs supérieures en CO_2 et le réchauffement peuvent stimuler la croissance dans certaines régions, l'instabilité du microclimat et la sécheresse peuvent en annuler ces effets pouvant, à l'extrême, amener à planter des espèces plus méridionales pour maintenir un couvert forestier (Broadmeadow *et al.* 2005).

A.2 Hydrologie et géologie de la région landaise

La région Aquitaine regroupe à elle seule près d'un million d'hectares de pins maritimes sur les 1.357 millions plantés en France. L'aire d'exploitation française du pin maritime est donc centrée sur cette région, plus particulièrement le département des Landes, qui présente des spécificités pédologiques particulières.
La capacité de croissance des végétaux dépend directement des caractéristiques du sol et du sous-sol dans lequel les racines se développent. Le sol landais actuel se singularise par une vaste zone sableuse triangulaire de près d'un million d'hectares.

Cette vaste plaine s'est formée durant le Jurassique à partir d'un vaste golfe qui s'est comblé progressivement par des dépôts d'épaisses séries sédimentaires sur plusieurs milliers de mètres. A la suite de mouvements tectoniques au tertiaire, certaines parties du triangle landais se sont enfoncées ou soulevées et le comblement s'est poursuivi. Au quaternaire, le comblement s'est poursuivi par d'abondants apports fluviatiles où se mêlent de façon variable des sables, argiles, graviers et galets. La présence des niveaux argileux dans la partie supérieure de la formation est particulièrement importante. Sur l'axe landais, ils constituent souvent le soubassement direct du sable des Landes et ont localement un grand rôle sur le régime de l'eau dans le sol. C'est pendant le quaternaire que les Landes vont devenir la plus grande plaine sableuse de France. Avec les changements climatiques et la fin des grandes glaciations, on observe une élévation du niveau de la mer qui rapproche progressivement le

rivage de sa position actuelle et remanie constamment les sédiments de la plateforme continentale. Les réseaux hydrographiques actuels se mettent en place et le vent alimente en sable l'intérieur des terres (Montané 1994).

Il se forme alors un constituant essentiel du sol landais : l'alios. Le lessivage des sables forme un horizon supérieur gris clair (podzol) de quelques décimètres d'épaisseur. Au-dessous se forme l'alios, qui est un grès tendre où les grains de sable sont cimentés par les acides humiques et l'oxyde de fer donnant une teinte brune à rougeâtre. Sa constitution interdit tout passage des racines vers l'eau de la nappe. L'épaisseur et la compacité de l'alios vont déterminer la vitesse de pénétration des eaux pluviales vers la nappe phréatique qui, associée aux très faibles pentes, va créer un écotype de lande sèche ou de lande humide.

On distingue ainsi différents types de landes correspondant à des écosystèmes différents fonctions de la profondeur de la nappe d'eau et de la fertilité du sol :

- En lande humide, la Molinie et l'Ajonc nain côtoient la Bruyère à quatre angles et la Bruyère ciliée. L'alios ainsi que les nappes sont très proches de la surface à 30-50cm.
- La lande mésophile fait place à la Fougère aigle, à l'Ajonc, au Genêt et à la Brande.
- Sur le sol sec de la lande sèche se trouvent la Callune, la Bruyère cendrée, l'Hélianthème. Les nappes sont plus profondes et l'alios plus éloigné de la surface.

A.3 Des marais à la sécheresse, un équilibre fragile

Jusqu'au milieu du XIXe siècle, l'eau apportée par les pluies hivernales ne pouvant pas traverser rapidement l'alios, et le système hydrique de cette vaste étendue sans pente ne pouvant évacuer l'eau, les landes de Gascogne se transforment en zone marécageuse. Les parties racinaires sont donc noyées et seules les espèces adaptées à l'ennoyage peuvent subsister. L'eau, croupie par décomposition des végétaux, s'évacue progressivement à l'approche de la saison estivale puis l'eau disponible aux racines dans l'horizon supérieur est rapidement épuisée. De telles conditions ne permettant pas la culture ou la sylviculture, l'élevage et l'apiculture sont pratiquement les seules activités. Il existe cependant quelques bois de pin maritime fournissant le bois, le charbon et la gemme. A cela s'ajoute la progression vers l'intérieur des terres des dunes littorales non stabilisées.

A la fin du XVIIIe siècle, Brémontier réussit à fixer les dunes par des plantations de végétation dunaire protégée du vent par des branchages. Puis Chambrelent mit en place un plan d'assainissement des Landes de Gascogne basé sur la réalisation d'un réseau de canaux collecteurs à l'origine de la loi du 19 juin 1857 obligeant les communes à assainir et à

ensemencer leurs landes. Les terres drainées ont dès lors pu être plantées et cultivées. Progressivement les forêts de pin maritime ont remplacé les anciens marais.

Aujourd'hui nous assistons à un déficit hydrique marqué entraînant des dépérissements massifs des peuplements de Pin maritime. D'après les informations relevées par les professionnels de la sylviculture et l'équipe d'écophysiologie de l'INRA de Pierroton, près de 30000 hectares sont touchés dans la région de Solferino et Escource au cœur des Landes (communication personnelle Alexandre Bosc, UMR EPHYSE). Considérant l'étendue des pertes, la cause hydrique peut être mise au premier plan, suivie par d'autres stress comme des attaques de pathogènes profitant d'arbres affaiblis. Il s'agit d'une zone de landes humides avec un alios situé à 40-50cm qui stoppe la pénétration des racines. L'horizon accessible par le système racinaire est donc très limité mais la nappe des landes humides est proche de la surface. En temps normal, l'eau liquide atteint les racines en dehors de la saison estivale avec l'accumulation des précipitations. Un important système hydraulique avec de gros fossés permet de drainer l'eau en hiver pour éviter l'ennoyage. Cependant, les prélèvements conséquents dans les nappes pour l'irrigation des terres agricoles, combinés à 3 années de déficit de précipitations et de sécheresse, ont provoqué une dépression du niveau de la nappe qui, à son niveau supérieur, ne doit plus dépasser l'alios et donc atteindre les racines.

A.4 Des incertitudes sur la durabilité du massif forestier aquitain

La forêt aquitaine a un double rôle à jouer : économique et environnemental. Avec 2.5 milliards d'euros de chiffre d'affaires, elle génère 21000 emplois directs auxquels s'ajoutent 7000 emplois en menuiserie, meubles et commerce (www.mediaforest.net). De sa bonne gestion dépend la préservation de la santé de l'économie et de l'écosystème d'une région entière. Les choix politiques et économiques récents soulignent la nécessité de préserver ce moteur économique essentiel en soutenant la création du pôle de compétitivité « Industries et Pin maritime du Futur » qui allie la recherche publique et privée.

En 2003, le pin maritime représentait 26% (2.9Mm3) du bois de trituration et 36% (5.4Mm3) du bois d'œuvre résineux au plan national.

Les utilisations du pin maritime sont multiples. Son bois peut être utilisé sous sa forme native en construction (bois de charpente, bardage, lambris, planches, poteaux, etc.), en ameublement, en emballage (palettes, caisses et coffrage) ou transformé en pâte à papier et panneaux de particules. D'autres usages sont plus spécifiques comme l'écorce pour le jardin ou la résine et autres extraits (picnogenol, parfums).

Face à la demande croissante en bois, cette forêt cultivée, dévolue à une fonction de production, aura un rôle de plus en plus important à jouer. L'amélioration des pratiques sylvicoles associées à un programme d'amélioration génétique initié dans les années 1960 à l'INRA (Baradat and Pastuszka 1992) visant à augmenter la production et la qualité du bois, a permis d'augmenter la productivité de façon significative. En effet, le gain attendu sur le volume de bois est de l'ordre de 30% pour les variétés actuellement plantées en Aquitaine (Alazard and Raffin 2002).

Cependant, dans un milieu changeant, on peut se poser la question de savoir si les gains espérés seront réalisés. Une estimation indique actuellement que si l'on considère les précipitations, l'évapotranspiration et la production de bois à l'hectare, 850 litres d'eau sont nécessaires pour produire 1kg de bois. Les variétés à croissance rapide auront probablement des exigences accrues en eau dans un avenir qui s'annonce peu favorable. Il est donc essentiel de prendre en compte d'autres critères que la croissance et la qualité du bois pour permettre un développement durable de ce massif forestier. Ainsi, il apparaît essentiel de prendre en compte l'adaptabilité des variétés futures à la modification de l'écosystème. Cela peut passer par l'augmentation de la diversité génétique des peuplements voire le mélange d'espèces et/ou l'amélioration génétique de l'écotype landais (sur lequel est essentiellement basé le programme d'amélioration) pour la résistance à la sécheresse. Des études récentes ont en effet montré qu'il existait au sein des populations aquitaines une forte variabilité pour l'efficience d'utilisation de l'eau (Brendel *et al.* 2002), un des caractères liés à l'adaptation à un déficit hydrique.

A.5 Objectifs

Mon travail de thèse s'inscrit dans le cadre de la poursuite des travaux initiés par l'INRA sur le déterminisme génétique (Costa 1999; Brendel *et al.* 2002), moléculaire (Costa 1999) et physiologique (Nguyen and Lamant 1988; 1989b; 1989a; Costa 1999; Nguyen-Queyrens *et al.* 2002) de la réponse du pin maritime à un déficit d'alimentation en eau. Ces études ont pour but d'améliorer nos connaissances sur le sujet et de permettre d'identifier des critères de sélection pour le choix de génotypes plus performants et mieux adaptés à leur environnement. Si les études physiologiques ont permis de mettre en évidence une variabilité de comportement entre écotypes de pin maritime, les QTL d'efficience d'utilisation de l'eau et d'autres caractères physiologiques soulignent l'existence d'un contrôle génétique de la

réponse au déficit hydrique. Il reste néanmoins à identifier les gènes impliqués dans l'adaptation à la sécheresse chez cette espèce.

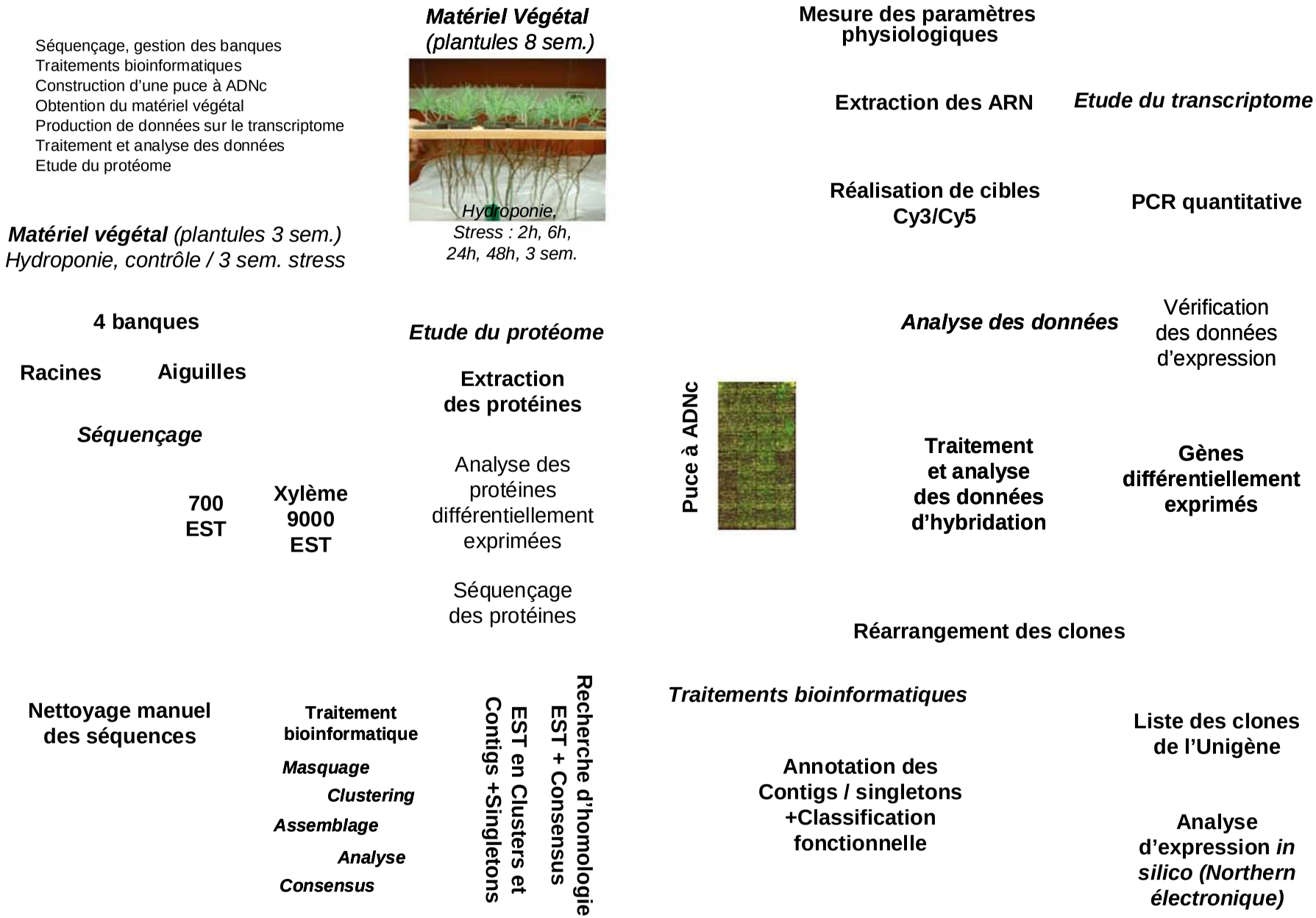
Mon travail se focalise sur ce dernier aspect. Il concerne l'identification de gènes régulés en condition de stress hydrique édaphique et qui présentent une variation d'expression chez des écotypes contrastés quant à leur réponse adaptative à la contrainte hydrique. Les recherches se concentreront sur les gènes exprimés dans les racines ; organe clé impliqué dans le captage de l'eau et qui a jusqu'ici fait l'objet de peu d'études chez les plantes en général.

Parallèlement à cet objectif scientifique, mon travail comprenait également des défis technologiques. Il me fallait en effet développer une approche « génomique » en mettant en place de nouveaux outils et méthodes d'étude de l'expression du génome : le séquençage massif d'étiquettes de séquences exprimées (EST) pour obtenir une base de ressources génomiques, le développement d'outils bioinformatiques pour les analyser et les rendre accessibles à la communauté scientifique, la création et l'utilisation des premiers réseaux d'ADNc (microarrays) chez le pin maritime.

La figure A-8 synthétise la démarche adoptée au cours de ce travail. Un code couleur permet de visualiser les différentes étapes suivies. La stratégie s'articule autour de 7 grands axes : 1) Production de séquences, 2) Traitements bioinformatiques, 3) Construction d'une puce à ADNc, 4) Obtention du matériel végétal, 5) Production de données sur le transcriptome, 6) Analyse de données d'expression au niveau transcriptomique 7) Etude du protéome. Ces sept étapes sont ici brièvement présentées :

1. Le projet a débuté par la production de séquences issues d'ARN extraits de racines (indiqué dans la partie gauche de la figure A-8). Ce programme de séquençage d'EST portait sur deux banques de racines obtenues sur de jeunes plants (3 semaines) de pins maritimes élevés en hydroponie : l'une correspondant à des plants témoins et l'autre à des plants soumis à 3 semaines de stress osmotique (PEG, -0.45MPa). Je disposais des banques et j'ai pris en charge les aspects séquençages avec l'aide d'un technicien du laboratoire.
2. Ces séquences ont été injectées dans une chaîne d'analyse bioinformatique pour l'assemblage et la recherche d'homologie contre les banques publiques. J'ai pris en charge la partie des traitements bioinformatiques concernant l'annotation, la

Figure A-8 :
Les différentes étapes réalisées pendant la thèse.



classification fonctionnelle, l'analyse d'expression *in silico* et l'obtention d'un Unigène pour la construction d'une puce à ADNc.

3. Après réarrangement et duplication des banques, que j'ai transmises à notre partenaire autrichien PICME (http://www.picme.at), une puce à ADNc a été construite.
4. Dans le même temps j'ai mis en place le dispositif pour obtenir le matériel végétal utilisé pour les études d'expression. Il s'agissait de plants âgés de 8 semaines, cultivés en hydroponie et soumis à un stress osmotique par ajout de PEG.
5. Ce matériel a fait l'objet de mesures physiologiques et d'une étude du transcriptome. Après extraction des ARN, j'ai synthétisé les cibles et hybridé les lames produites.
6. Les données générées ont alors été traitées et analysées afin d'identifier des gènes différentiellement exprimés lors du stress hydriques chez deux écotypes, français et marocain.
7. Le protéome du même matériel végétal a également été analysé.

A.6 Plan de la thèse

Ma thèse est structurée en quatre parties principales :

La **partie 1** est une synthèse bibliographique. J'ai choisi de focaliser cette étude sur les mécanismes de réponses moléculaires et physiologiques des plantes au stress hydrique. Je traiterai aussi, brièvement, des stratégies possibles pour la découverte des gènes majeurs pour l'adaptation à la sécheresse chez le pin maritime. Cette partie bibliographique est divisée en quatre chapitres.

- Le premier chapitre présentera les caractéristiques botaniques de l'espèce étudiée, les écotypes et son aire de répartition géographique.
- Le second chapitre traitera des réponses des plantes aux stress. Après avoir décrit les relations de l'eau et de la plante, nous aborderons les mécanismes de régulation physiologique que la plante met en œuvre au cours du stress. Puis, nous passerons en revue les voies métaboliques et les gènes connus dont l'expression est régulée au cours du stress.
- Le troisième chapitre sera consacré au stress hydrique chez les arbres. Nous découvrirons les travaux de recherches menés jusqu'à présent sur cette thématique chez le pin maritime. Les approches physiologiques seront présentées, puis nous passerons au déterminisme génétique avec les études de détection de QTL et de

génétique quantitative, puis aux mécanismes moléculaires avec l'étude du protéome et du transcriptome.

- La dernière partie abordera les aspects techniques et les stratégies pouvant être mises en œuvre pour la détection des gènes impliqués dans l'adaptation du pin maritime à la sécheresse.

La **partie 2** détaille les principales méthodologies utilisées. Je présenterai successivement le matériel végétal et le dispositif expérimental, les techniques de biologie moléculaire utilisées pour obtenir les données d'expression transcriptomique et protéomique, puis les méthodes bioinformatiques et statistiques mises en œuvre pour traiter les données.

La **partie 3** se structure également en 4 chapitres. Chaque chapitre présente les résultats produits et leur interprétation.

- Le chapitre I concerne l'obtention du matériel végétal et les paramètres physiologiques mesurés au sein du dispositif expérimental que j'ai mis en œuvre.
- Le chapitre II concerne l'analyse et le traitement des séquences. Nous nous intéresserons tout d'abord au prétraitement et à l'assemblage des séquences. Ces étapes ont été réalisées en faisant appel à un nouveau système d'interface web associé à un pipeline bioinformatique auquel j'ai participé pendant 2 ans en tant que beta testeur. Ce nouvel outil fait l'objet d'une note dont je suis co-auteur (article soumis, Annexe Px-1). Les séquences produites dans ce projet pin maritime ont également été le support pour la recherche *in silico* de SNP (Single Nucleotide Polymorphism) par une chaîne d'analyse automatique présentée dans une publication de Plant Molecular Biology dont je suis également co-auteur, annexe Px-2 (Le Dantec *et al.* 2004). Nous présenterons ensuite le travail d'annotation fonctionnelle réalisé sur les séquences puis leur utilisation pour l'approche par « Northern électronique » dont l'objectif était de détecter les variations d'expression des transcrits par comptage du nombre de séquences présentes dans les bases de données d'EST. Cet aspect a fait l'objet du stage de DESS de Christophe Boury que j'ai co-encadré. La fin de ce chapitre donnera les informations sur l'Unigène que j'ai généré et sera suivie d'une discussion des résultats observés.
- Le chapitre III est dédié à l'analyse du transcriptome. Il traitera dans un premier temps de l'ensemble des données d'expression que j'ai obtenues en utilisant la technologie

des puces à ADNc. Nous passerons en revue les résultats des analyses statistiques et des méthodes de classification. Dans un second temps, les niveaux d'expression obtenus par PCR quantitative sur un ensemble de gènes sélectionnés seront présentés. Les résultats de ces deux approches complémentaires seront interprétés et discutés. Une étudiante de DES que j'ai co-encadrée (Manon Moreau) a également participé à l'acquisition de ces résultats.

- Dans le chapitre IV nous synthétiserons les données obtenues par l'analyse du protéome (Stage de Master 2 de Manon Moreau) et les comparerons aux données du transcriptome.

Enfin la **partie 4** est une conclusion générale de l'ensemble des résultats obtenus dans le cadre de cette thèse et propose des perspectives au vu des informations recueillies.

Synthèse Bibliographique

B SYNTHESE BIBLIOGRAPHIQUE

Comme tout organisme vivant, la plante est essentiellement constituée d'eau. La plupart des végétaux sont soumis à des variations de la disponibilité en eau du milieu. Ils ont donc dû développer au cours de l'évolution des mécanismes de réponse adaptative pour faire face à des périodes de déficit en eau. De nombreux travaux ont permis de décrire les réponses des plantes en état de stress. Les effets sont multiples et vont de l'altération de l'expression des gènes et du métabolisme cellulaire, à la modification de la croissance avec des conséquences sur le rendement des cultures. Les réponses peuvent être directes, initiées par le stress et facilement identifiables, ou indirectes pour palier aux conséquences du stress. Au niveau moléculaire, cela implique la régulation d'un nombre conséquent de voies métaboliques. Après la présentation de l'espèce, nous ferons dans cette synthèse l'état des connaissances sur les mécanismes majeurs impliqués lors d'un déficit en eau. Nous décrirons également les voies métaboliques permettant la synthèse des métabolites impliqués dans la réponse au stress hydrique, la régulation de leur production étant déterminée par les enzymes des voies de biosynthèse et des facteurs de transcription.

B.1 Le Pin maritime

B.1.1 Botanique et Taxonomie

Les conifères appelés également résineux sont les représentants actuels des gymnospermes apparus à la fin de l'ère Primaire. Ces espèces ont colonisé tout l'hémisphère nord au cours de l'ère secondaire. Les traces fossiles du genre *Pinus* sont nombreuses et remontent au début du crétacé (Wang *et al.* 2000) soit il y a environ 140 millions d'années. Les changements climatiques avec les périodes glaciaires et interglaciaires du quaternaire ont eu pour effet la colonisation de territoires ou la création de zones refuges à l'origine de spéciations. La figure B-1a indique la chronologie de la différentiation des gymnospermes.

Les conifères forment l'ordre des *Coniférales* ou *Pinales*, divisé classiquement en six familles (www.species2000.org) : les *Araucariaceae* (araucarias, Agathis), les *Céphalotaxaceae* (Cephalotaxus), les *Cupressaceae* (cyprès, thuyas, genévrier), les *Pinaceae* (mélèzes, sapin, pins, cèdres épicéas), les *Podocarpaceae* (Podcarpus) et les *Taxodiaceae* (séquoias).

La famille des *Pinaceae* est divisée en sept genres : *Abies*, *Cedrus*, *Larix*, *Picea*, *Pinus*, *Pseudotsuga*, et *Tsuga*. Trois autres genres peuvent être distingués : *Cathay*, *Keteleeria* et *Pseudolarix* (http://tolweb.org).

Figure B-1a : Arbre phylogénétique et origine de la différentiation des gymnospermes

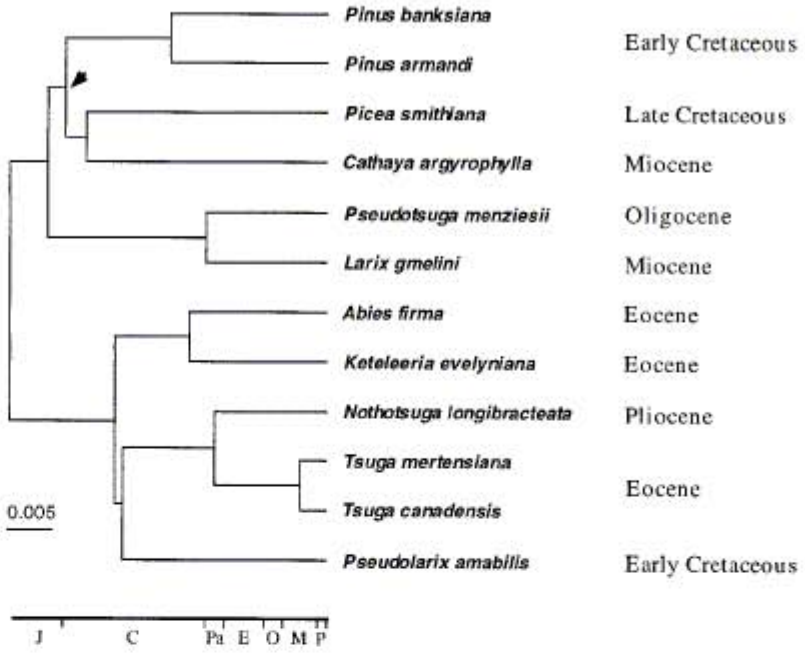


Figure B-1b : Arbre phylogénétique et classification botanique des espèces du genre Pinus

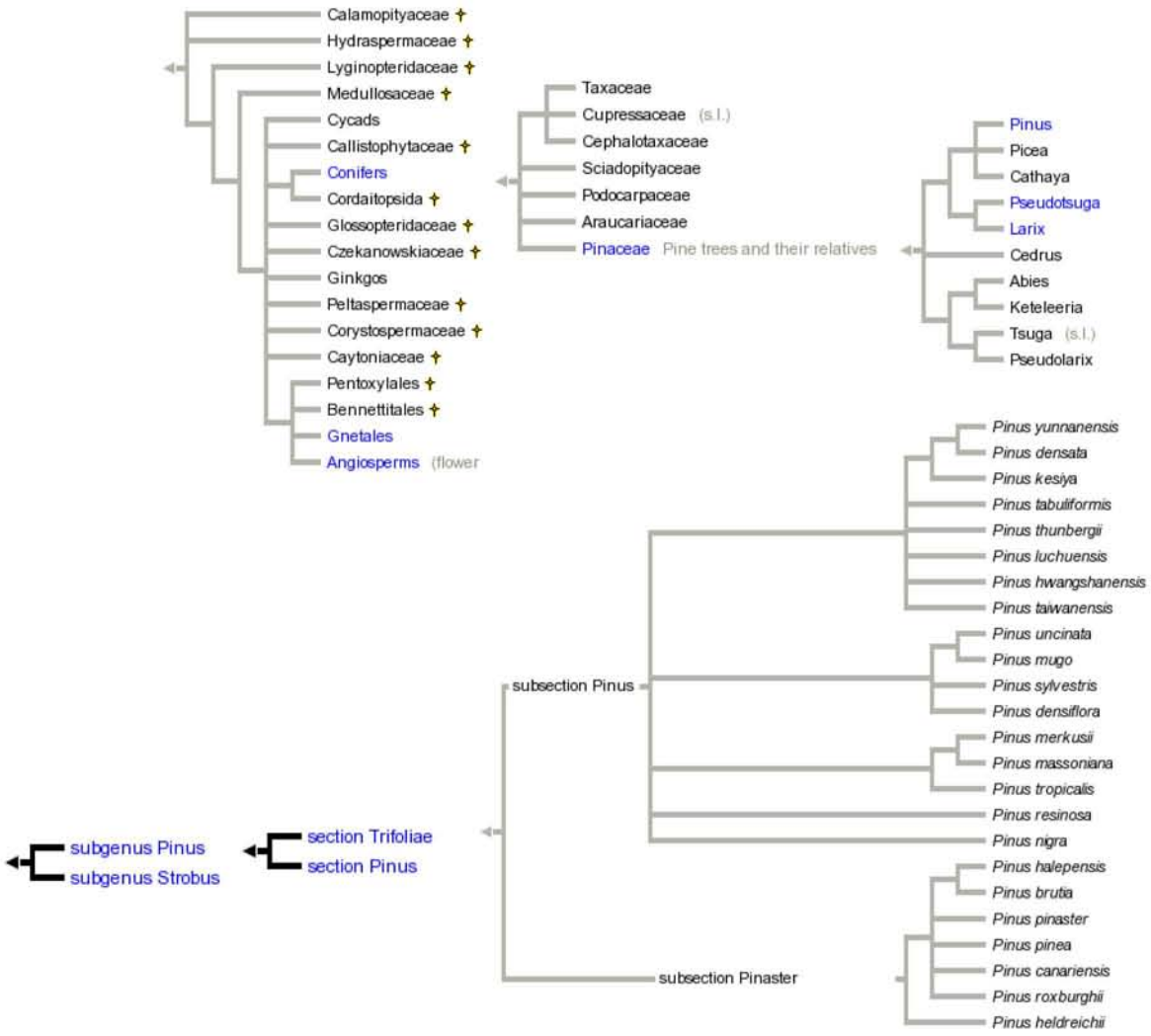


Figure B-2 : Pin maritime
(Pinus pinaster Ait.)

Le genre *Pinus* est séparé en deux sous-genres : *Pinus* (2-3 aiguilles) et *Strobus* (5 aiguilles) qui regroupent un peu plus d'une soixantaine d'espèces. Ces sous genres sont divisés en sections, subdivisées elles-mêmes en sous-sections (Fig. B-1b).

Division	*Coniferophyta*
Classe	*Pinopsida / Coniferopsida*
Ordre	*Pinales / Coniférales*
Famille	*Pinaceae*
Genre	***Pinus***
Sous-Genre	*Pinus*
Section	*Pinaster / Pinus*
Sous-Section	*Australes / Pinaster*
Espèce	***Pinaster***

D'après http://www.itis.usda.gov/index.html

Les pins ont pour caractéristique commune la présence d'un appareil foliaire sous forme d'aiguilles pointues, longues ou courtes et réunies en groupes de 2, 3, 4 ou 5. Ce sont des plantes monoïques, dont l'inflorescence femelle ou cône mûrit en deux (rarement trois) ans après fécondation.

Pinus pinaster (Fig. B-2) est couramment appelé Pin maritime, Pin des Landes, Pin de Bordeaux, Pinastre, Pin de Corte pour la variété corse ou encore « Cluster Pine » en anglais. La floraison a lieu en avril/mai avec une production massive de pollen transporté par le vent. Les graines sont souvent ailées, ce qui facilite leur dissémination par le vent et l'extension de leur aire de distribution.

C'est une espèce allogame, sempervirente, à croissance initiale rapide, qui arrive à maturité à 40-50 ans en atteignant en général une taille de 20 à 35 m. Son espérance de vie peut cependant atteindre les 200 ans. Le tronc, flexueux à la base, est recouvert d'une écorce abondante (Taux d'écorce en volume = 25%) brun violet, rouge sombre en profondeur très épaisse et fortement crevassée. Le bois a une densité moyenne (Masse volumique brute : 880kg/m^3, masse volumique sans écorce = 700kg/m^3), rougeâtre foncé, et à aubier nettement plus clair ; les canaux résinifères sont nombreux, gros, et visibles à l'œil nu ; la sécrétion de résine est abondante. Son houppier est conique, étalé et peu compact. Les aiguilles sont rigides et présentent des stomates sur les deux faces. Elles sont regroupées en fascicules de deux aiguilles pouvant dépasser 15-20cm de long mais dans le très jeune âge, les sujets vigoureux en ont souvent trois, au moins sur la pousse principale. On distingue les euphylles

ou feuilles aciculaires qui apparaissent aussitôt après les cotylédons sur les tiges et rameaux du semis de l'année. Puis les aiguilles ou pseudophylles se développent à la place des euphylles qui sont très vite réduites à l'état d'écailles brunes (Maugé 1987).

B.1.2 L'appareil racinaire

Il a pour fonction l'ancrage de l'arbre et son alimentation en eau. C'est l'organe directement au contact de la ressource hydrique et de son développement va dépendre les capacités de réponses aux périodes de sécheresse.

Le système racinaire du Pin maritime est souvent qualifié de pivotant ou plongeant. En réalité, le système est double ; constitué d'un système plongeant très développé et d'un système traçant (Maugé 1987; Danjon *et al.* 1999).

Le système plongeant est le premier à se développer sur les jeunes plants à partir d'un axe unique qui se ramifie ensuite, donnant d'autres racines à développement vertical (Fig. B-3). Lorsque l'horizon atteint est riche en eau, le développement du système plongeant s'arrête au profit d'un pinceau racinaire constitué de nombreuses racines fines dont la taille peut atteindre 1m (dunes) ou sera réduite dans le cas d'une nappe. Cependant, cette organisation du système plongeant est modifiée selon la nature du sol et les possibilités de pénétration des racines en profondeur (Fig. B-5). La plasticité du système racinaire permet de modifier considérablement l'axe de développement préférentiellement vertical.

Si un obstacle infranchissable - comme l'alios d'une lande humide dont nous reparlerons dans les chapitres suivants- est rencontré, les racines verticales s'arrêtent et forment un moignon et un réseau racinaire enchevêtré circulaire d'environ 2 mètres de diamètre (Danjon *et al.* 2005)(Fig. B-6) offrant une résistance d'arrachement très limitée en cas de forts vents d'où les fréquents chablis rencontrés lors des tempêtes.

Si l'obstacle peut être contourné – comme l'alios très fissuré des landes sèches -, les racines verticales modifient leur progression en le longeant. Dès que le point de franchissement est atteint, les racines adoptent leur mode de développement initial (Fig. B-4). L'ancrage est alors très bon et les dégâts observés en cas de tempête sont plutôt des volis.

Le système traçant se développe juste sous la base du collet en émettant 5 à 7 racines d'ordre 1 à développement radial et parfaitement horizontal desquelles partent des racines d'ordre 2, 3, etc. conservant toujours la direction centrifuge des racines d'ordre 1. Elles se développent dans l'horizon, superficiel bien aéré et constitué de matières organiques en décomposition. Tout obstacle stoppe la croissance de la racine radiale qui est donc impossible

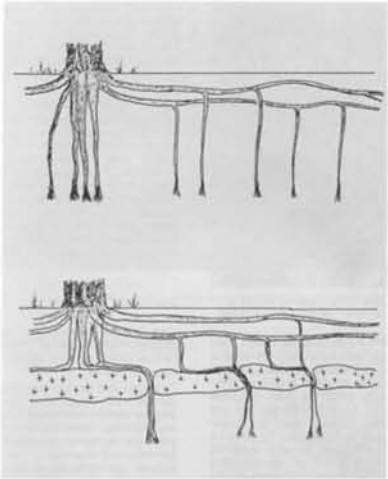


Figure B-3 : Développement du système racinaire plongeant et traçant dans un sol idéal chez le pin maritime

Figure B-4 : Développement du système racinaire plongeant et traçant en présence d'alios fissuré du type lande sèche chez le pin maritime (Maugé (1987))

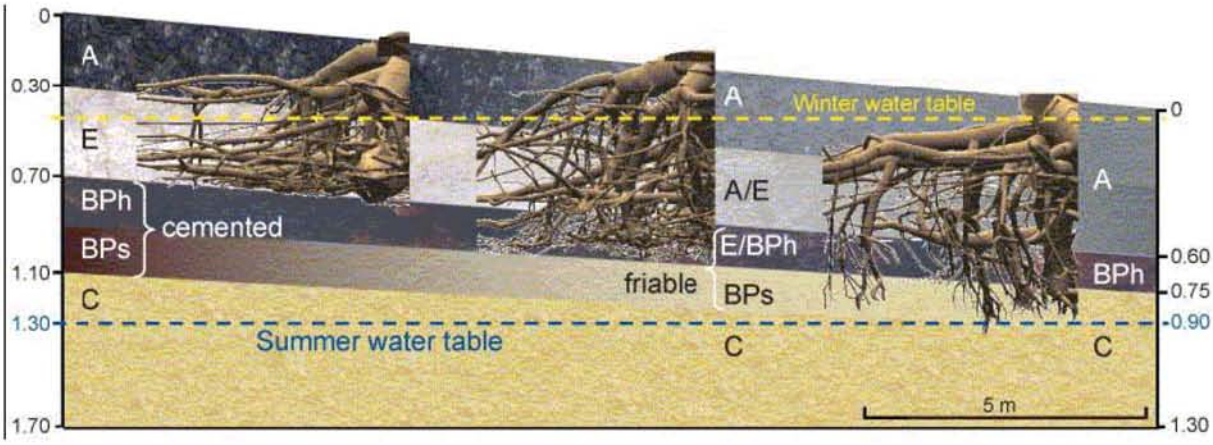


Figure B-5 : Nature de l'enracinement de pins maritimes (reconstitution 3D) en fonction de la nature du sol. (Danjon (2005)). A gauche : sol bien drainé, l'alios (BP) stoppe la progression des racines. A droite, sol mal drainé, l'alios friable permet un passage des racines.

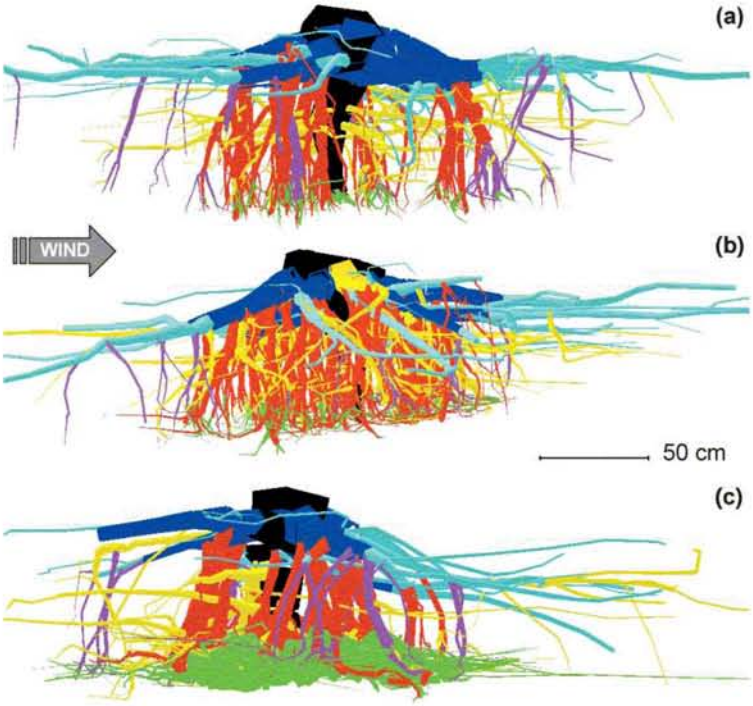


Figure B-6 : Reconstitution 3D du système racinaire « en plateau » d'un pin maritime des landes. L'enracinement diffère (a,b,c) en fonction du vent et de la nature du sol. (Danjon(2005))

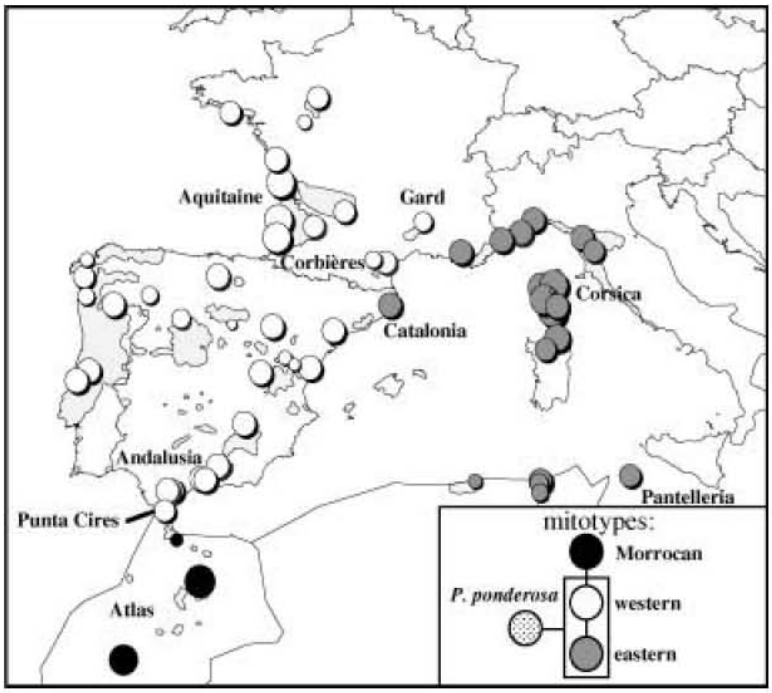


Figure B-7 : Distribution des mitotypes de pin maritime et lien de phylogénie basé sur les variations de nad1 intron 2 d'après Burban et *al.* (2003).

dans un sol un peu compact ou occupé par le chevelu racinaire de graminées comme la molinie. De ce réseau de racines principales partent des racines fines d'une taille de 10 à 15 cm pour un diamètre proche du millimètre formant une sorte de manchon. Ces racines, souvent mycorhizées, présentent des excroissances couvertes de poils absorbants. Elles ont une durée de vie brève de 1 à 3 ans puis se décomposent. Leur fonction principale est de capter activement les éléments nutritifs ce qui nécessite de l'oxygène. Le sol doit pour cela être humide sans toutefois être gorgé d'eau ce qui provoquerait une anoxie. Le système traçant a donc un rôle d'ancrage et d'alimentation de l'arbre en éléments nutritifs. C'est sur cet organe qu'ont porté nos investigations.

B.1.3 Répartition géographique

Dans une publication récente, les auteurs Burban et Petit (2003) décrivent des données de phylogéographie obtenues sur le Pin maritime en utilisant des marqueurs chloroplastiques. En se servant des marqueurs à hérédité uniparentale, ils ont ainsi pu préciser la structuration génétique de l'espèce décrite dans la littérature et basée sur d'autres marqueurs (terpènes, protéines totales, isoenzymes, microsatellites nucléaires et AFLP). Le pin maritime est séparé en 3 groupes : un groupe atlantique comprenant les populations de l'ouest de la France et de la péninsule ibérique, un groupe méditerranéen constitué de toutes les populations de l'est de l'Europe, de la partie nord africaine et des populations de Catalognes, et un groupe marocain (Fig. B-7). Nous nous intéresserons plus particulièrement à l'écotype « Tamjout » originaire du Maroc. Cette station est la plus orientale des provenances marocaines. Localisée au sud de Taza à 1650 m d'altitude en exposition nord ouest, le climat se situe dans l'étage méditerranéen subhumide. Les précipitations atteignent 650 mm par an répartis en 250 mm l'hiver, 210 mm au printemps, 40 mm en été et 150 mm en automne. Le sol est calcaire avec une forte proportion de carbonate et de calcaire actif (Guyon 1980).

En France, l'aire naturelle du pin maritime se situe à faible distance de la mer dans les Landes de Gascogne, la Corse, les Maures, l'Estérel et les Corbières. Par comparaison avec la station marocaine, les précipitations annuelles dans le département des Landes atteignent environ 980 mm avec un minimum de 200 mm pour la saison estivale. Le pin maritime a aussi été introduit dans d'autres régions, notamment en Bretagne, Pays de Loire, Centre, Charente, Aquitaine et Cévennes. Dans le reste du monde, le pin maritime a été introduit au Chili (100 000 ha), en Australie occidentale (500 000 ha en cours de plantation), en Afrique du Sud (40 000 ha), en Argentine, en Nouvelle-Zélande (3 000 ha) et en Grèce (10 000 ha).

B.1.4 Habitat

Le pin maritime est bien adapté aux climats maritimes très tempérés, à température douce et régulière. Cette espèce exige une légère humidité de l'air, mais supporte la sécheresse estivale.

Il tolère des températures de 40°C voire plus quelque soit la provenance. Par contre, la variabilité est importante pour la résistance au gel. En France, les arbres originaires du Portugal ou de Galice ont montré une forte sensibilité lors des hivers de 1956 et 1985 tandis que les variétés landaises ont bien supporté des températures de -25°C (Maugé 1987).

Le niveau des précipitations importe peu si l'accès à une ressource en eau est possible, mais comme nous l'avons vu auparavant, l'humidité de l'horizon superficiel aura beaucoup d'importance sur l'assimilation des éléments nutritifs et donc la production de bois.

Cette essence héliophile exige beaucoup de lumière. C'est une espèce pionnière qui supporte mal la compétition avec une autre végétation, surtout lors de son installation.

Le pin maritime préfère les stations riches, avec des sols légers et profonds, bien pourvus en eau tant qu'il n'y a pas de compétition. Mais il se contente très bien le plus souvent de sols acides et pauvres (essence frugale) qui constituent la majorité de son aire de répartition. Ces sols présentent toutefois régulièrement des carences en phosphore.

Il se développe également bien sur des sols très pauvres en éléments nutritifs -comme les dunes- grâce à son système traçant et aux mycorhizes avec lesquelles il entretient des relations symbiotiques.

Cette espèce est souvent dite calcifuge en raison de son affinité pour les pH acides. Or à ce pH, les argiles sont dispersées, peu pénétrables et le sol compact, ce qui ne permet pas au système traçant de se développer.

B.2 Les plantes et la sécheresse

B.2.1 L'eau et la plante

B.2.1.1 L'eau du sol

La plante tire l'eau et les éléments nutritifs du sol.

Un sol est un milieu complexe, constitué de particules solides (pierres, graviers, sables ou limons), de colloïdes et d'une phase liquide. Les colloïdes sont des macromolécules très hydrophiles qui peuvent être d'origine minérale –ce sont des argiles– ou d'origine organique –les humus-. La combinaison de ces colloïdes forme un complexe argilo-humique. La phase liquide est directement accessible par les racines et constituée d'eau et de substances dissoutes. Elle n'est cependant pas libre et plus le sol se dessèche, plus il est difficile d'en extraire de l'eau qui est liée aux constituants du sol par les forces osmotiques et les forces matricielles.

Les forces osmotiques ont pour origine la concentration en ions et molécules organiques des solutions. Les solutés de deux solutions séparées par une membrane perméable –comme ceux du sol et ceux des racines- iront spontanément de la solution la plus concentrée vers la solution la moins concentrée. De la même manière les molécules d'eau iront de la solution où le potentiel chimique de l'eau est élevé vers celle où il est le plus bas. Le potentiel chimique représente l'enthalpie libre d'une mole d'eau et dépend de la « concentration de l'eau ». En cas de sécheresse, la « concentration en eau » étant plus faible, celle des ions et molécules organiques est donc plus élevée. Les forces osmotiques peuvent alors atteindre 20 bars.

Les forces matricielles englobent les forces d'imbibition et les forces capillaires. Les forces d'imbibition ont pour origine les attractions électrostatiques entre les charges négatives des colloïdes et les charges positives des atomes d'oxygène des molécules d'eau. Lorsque le sol se dessèche les forces mises en œuvre atteignent plusieurs centaines de bar. Les forces capillaires sont liées à des phénomènes de tension superficielle et retiennent l'eau dans les interstices fins.

La notion de potentiel hydrique Ψw est couramment employée pour évaluer les capacités d'une plante à prélever de l'eau du sol.

Pour la plante, le potentiel hydrique dépend essentiellement du bilan de deux forces opposées : le potentiel osmotique Ψs qui dépend de l'osmolarité de la solution baignant les

cellules des racines et la turgescence Ψp exercée par les parois cellulaires. Les flux hydriques résulteront de cet équilibre entre forces osmotiques et mécaniques.

$$\psi_W = \psi_S + \psi_P$$

Le potentiel hydrique d'un sol est égal mais de signe opposé, à l'énergie qu'il faut lui appliquer pour en libérer un gramme d'eau.

Pour maintenir un flux d'eau du sol vers les racines, la plante doit maintenir le potentiel hydrique des racines inférieur à celui du sol. Dans le cas contraire, il y a perte de turgescence et la plante perdra de l'eau. En situation de stress hydrique, le potentiel hydrique du sol s'abaisse, la plante doit alors répondre en abaissant également le potentiel hydrique des racines et peut pour cela, diminuer la pression mécanique exercée par les parois (début de plasmolyse) ou moduler la pression osmotique en augmentant l'osmolarité de la solution cellulaire par accumulation de solutés ou osmoticum, ce mécanisme est appelé osmorégulation.

B.2.1.2 L'eau dans la plante

L'eau captée par les racines va occuper l'ensemble des cellules et des espaces intercellulaires du végétal. La mesure d'un autre paramètre, la teneur en eau relative, permet de connaître le niveau de saturation en eau ou de turgescence de la plante. Il est déterminé de la façon suivante : $RWC = [(MF - MS)/(MT - MS)] \times 100$ RWC=Teneur Relative en Eau

MF, MS, MT= Masse de Matière Fraiche, Sèche et Turgescente

Au niveau des racines, l'eau est rapidement canalisée dans le système conducteur xylémien constitué de cellules différentiées en vaisseaux conducteurs et en trachéides s'arrêtant dans les organes foliaires. La circulation de l'eau s'effectue du sol vers les feuilles où elle va être évaporée sous l'action du rayonnement solaire. Les mouvements d'eau dans la plante sont donc globalement axiaux du sol à l'atmosphère. Un arbre évapore ainsi plus de 200 litres d'eau par jour (Wullschleger *et al.* 1998). Cependant, sous l'action de la photosynthèse, les chloroplastes produisent des hydrates de carbones comme le saccharose. Ces assimilas gagnent un système conducteur phloémien formé de tubes criblés où circule la sève élaborée qui va alimenter l'ensemble de la plante. Le sens de circulation est principalement opposé à celui de la sève brute bien que la sève alimente également les zones de croissance active que sont les bourgeons.

La montée de l'eau depuis le sol jusqu'aux feuilles à travers le système vasculaire nécessite un système de forces. Outre les pressions osmotiques déjà citées, la principale est expliquée

par la théorie de la tension-cohésion foliaire et la transpiration énoncée par Dixon et Joly en 1894 et reformulée par Zimmermann en 1983. Elle repose d'une part sur l'existence d'une continuité hydraulique entre le sol, la plante et l'atmosphère et d'autre part sur la forte cohésion des molécules d'eau entre-elles et avec les parois des trachéides et des vaisseaux. La plante peut être comparée à une colonne d'eau, l'évaporation de l'eau au niveau des feuilles provoque un appel d'eau qui se transmet à toute la colonne et se traduit par une force de succion au niveau racinaire. Les parois des éléments vasculaires doivent donc supporter des tensions fortes et d'autant plus importantes que la disponibilité en eau du sol est faible. La résistance et les caractéristiques hydrauliques de ces parois auront une influence sur les propriétés de réponse à des situations de stress hydrique (Steudle 2001).
Les forces de cohésions supportent des tensions pouvant atteindre 20 bars mais la présence d'une rupture dans la colonne comme la présence d'une bulle d'air provoque une cavitation. L'élément conducteur se remplit d'air puis de vapeur d'eau et la circulation d'eau est stoppée.

La transpiration est observée sur toute la surface de l'appareil foliaire mais elle est localisée essentiellement au niveau des stomates sur la face inférieure des feuilles ou sur les deux faces dans le cas des aiguilles. Il s'agit de dispositifs formés par deux cellules de garde contrôlant la fermeture d'un espace appelé chambre sous-stomatique qui communique avec l'extérieur par l'ostiole. Ils sont le lieu d'échange du CO_2 et de l'H_2O nécessaire à la photosynthèse. Chez les arbres, le CO_2 diffuse vers les cellules du mésophile où il se combine sous l'action de la Rubisco au Ribulose 1,5 bisphosphate pour donner l'Acide 3-phosphoglycérique qui sera utilisé dans le cycle de Calvin. Si l'apport de ces gaz et de l'eau est indispensable pour permettre la réaction chimique et donc la croissance, les stomates laissent également passer l'eau sous forme de vapeur. La plante module l'ouverture des stomates en fonction des conditions de lumière, de l'état hydrique de la plante et du déficit de pression de vapeur (VPD). Le système stomatique a été largement étudié et est un des principaux acteurs de régulation de l'état hydrique de la plante (Li *et al.* 2000; Schroeder *et al.* 2001; Laporte *et al.* 2002; Luan 2002). Les stomates ont donc un rôle majeur dans la réponse au stress hydrique (Buckley 2005). Quant à la transpiration de l'épiderme, elle va dépendre de la nature de celui-ci et la présence d'une cuticule (Burghardt and Riederer 2003).

B.2.1.3 L'eau et les racines

L'appareil racinaire est l'organe spécialisé dans l'absorption de l'eau et des substances minérales en régulant comme nous venons de le voir les propriétés osmotiques de ses tissus. Il

apparaît donc comme une cible d'étude particulièrement intéressante dans le cadre de réponses à des déficits hydriques. Cependant sa morphologie complexe et ramifiée rend les observations difficiles en terre. Les expériences en hydroponie permettent de simuler l'état de sécheresse d'un sol en modulant l'osmolarité de la solution par l'ajout de macromolécules comme le polyéthylène glycol.

Nous avons vu que l'eau était canalisée dans des éléments vasculaires. Avant de parvenir à ces structures différentiées, l'eau est absorbée par les cellules épidermiques puis doit franchir l'endoderme. Le transfert passif des solutés de cellule à cellule reste faible à cause de la faible perméabilité aux solutés des membranes. L'eau par contre circule plus facilement au travers des membranes. Elle peut emprunter différentes voies de circulation (Steudle 2000) (Fig. B-8):

- L'eau peut circuler dans le symplasme, c'est-à-dire de cytoplasmes en cytoplasme sans transit par la vacuole, par l'intermédiaire des plasmodesmes qui sont des communications intercellulaires.
- Elle peut également circuler dans l'apoplasme qui englobe l'ensemble des parois pectocellulosiques, lacunes et méats, en communication directe avec le milieu extérieur.
- Lorsque le transport s'effectue au travers des membranes plasmiques, il peut être facilité par des aquaporines. Il s'agit de la voie transcellulaire.

Dans les deux premiers cas, aucune membrane n'est traversée alors que dans la circulation transcellulaire, deux membranes doivent être traversées par couche cellulaire. La régulation de la perméabilité membranaire par la présence d'aquaporines va pouvoir moduler la circulation hydrique. Les flux apoplasmiques vont dépendre de la composition et de l'épaisseur des parois et de la présence de barrières imperméables telles que la bande de Caspari ou une lamelle de subérine que l'on retrouve au niveau de l'endoderme ou de l'exoderme (Fig. B-8). La présence de ces deux structures va modifier considérablement les propriétés hydrauliques du tissu considéré.

Le trajet de l'eau va pouvoir être combinée entre les voies décrites. La circulation symplasmique et transmembranaire ne vont pas pouvoir être distinguées en pratique et forment la circulation de cellule à cellule.

Les racines sont donc des structures composites et ne peuvent pas être résumées à la simple membrane parfaite d'un osmomètre.

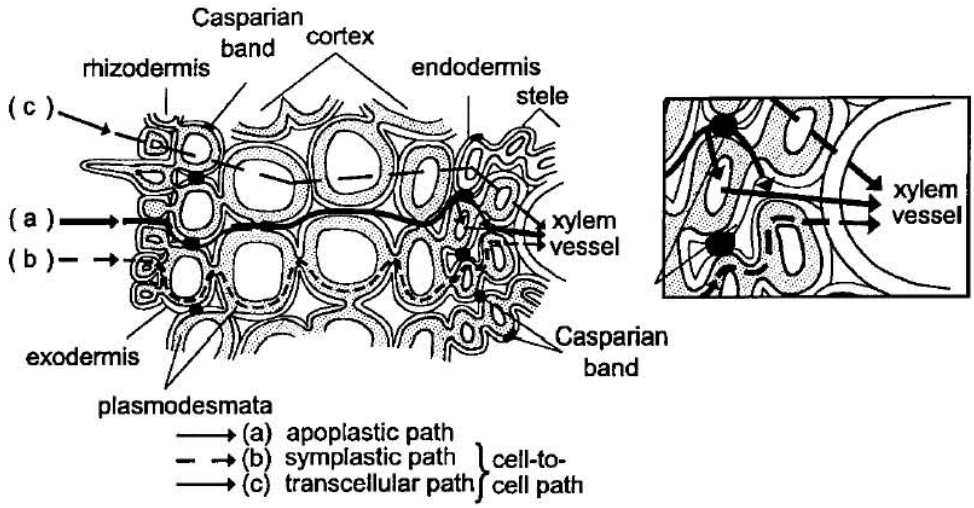


Figure B-8 : Trajet de l'eau dans les racines : a) voie apoplasmique, b) voie symplasmique, c) voie transcellulaire de vacuole à vacuole. La voie apoplasmique est interrompue par le cadre subérifié (Bande de Caspari) d'après Steudle *et al.* (2000)

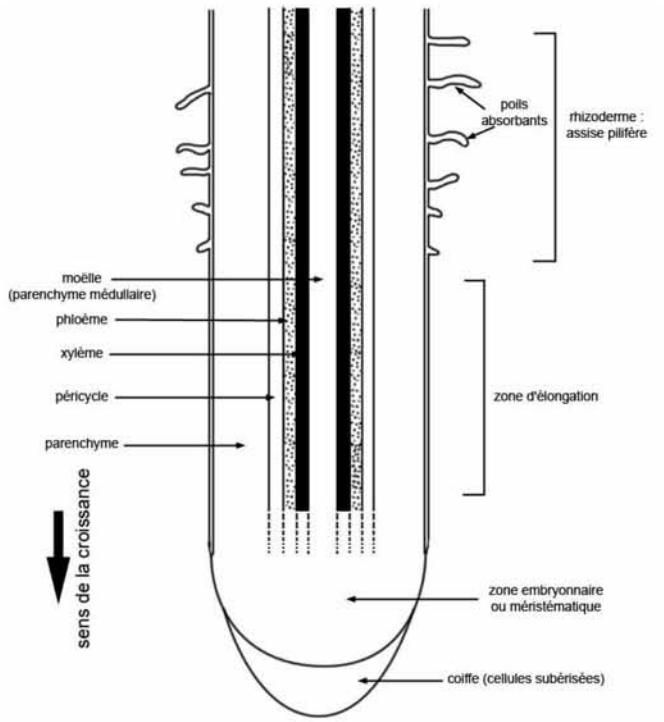


Figure B-9a :
Schéma d'une coupe longitudinale de la partie distale d'une racine.

http://fr.wikipedia.org/wiki/Racine_(botanique)

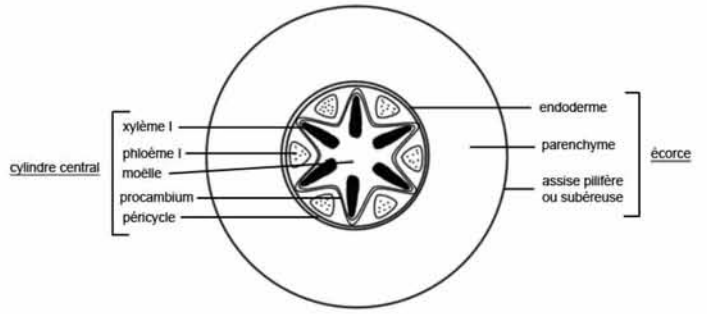


Figure B-9b :
Schéma d'une coupe transversale de la zone subéreuse d'une racine.

http://fr.wikipedia.org/wiki/Racine_(botanique)

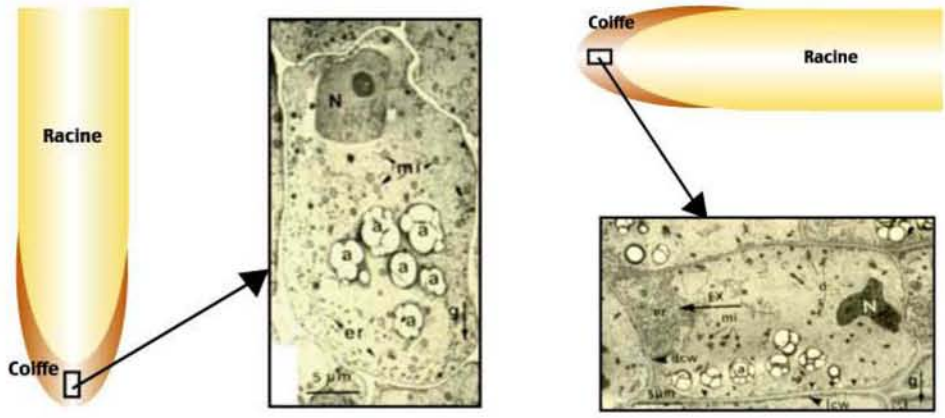


Figure B-10 : Perception de la gravité dans une racine de Lens culinaris.
A Gauche : Photographie d'un statocyte lorsque la racine est en position verticale
A Droite : Photographie d'un statocyte lorsque la racine est en position horizontale
N = Noyau, a = Amyloplastes. (D'après http://www.snv.jussieu.fr/bmedia/gravitropismeBM/perception.html)

Une racine peut être divisée en quatre zones (Fig. B-9a) :

- La zone subéreuse qui correspond à la partie la plus âgée de la racine. Elle porte les racines secondaires ou radicelles.
- La zone pilifère qui émet des poils absorbants. Ce sont eux qui permettent l'absorption de l'eau et des sels minéraux.
- La zone d'accroissement située derrière la coiffe. Elle est responsable de la multiplication cellulaire.
- La coiffe qui termine et protège la racine. Elle permet la pénétration dans le sol.

Au niveau anatomique, la zone subéreuse des racines est constituée de plusieurs tissus différenciés (Fig. B-9b). Au centre, une racine comporte une couche cellulaire appelée endoderme enfermant un cylindre central constitué du péricycle et des tissus conducteurs du phloème et du xylème transportant la sève brute et élaborée. Cette structure ou stèle est entourée par le parenchyme cortical et l'épiderme formant le cortex.
La zone d'accroissement, suivie de la coiffe entourant l'extrémité de la racine, est constituée de cellules en division au métabolisme très actif.
La coiffe protège le méristème apical où se déroulent les divisions cellulaires qui permettent la croissance de la racine. Des cellules centrales ou centre quiescent sont le siège de divisions cellulaires à l'origine des cellules initiales de l'endoderme et des cellules de la coiffe. Elles assurent le renouvellement des cellules pendant que les cellules apicales les plus externes se desquament progressivement. Les parois de ces cellules se transforment en un mucilage qui permet un glissement doux de la jeune racine. La coiffe joue un rôle majeur dans l'orientation de la croissance de la racine. Dans la région centrale de la coiffe (Columelle), les cellules riches en amidon servent de capteur de la gravité par la position des amyloplastes et permettent de modifier le sens de croissance de la racine en fonction de sa position (Fig B-10).

Une extraction des ARN de racines entières regroupera donc les niveaux d'expression des transcrits correspondant à différents types tissulaires. Les variations d'expression observées pourront donc correspondre entre autre, à des modifications de croissance, d'initiation de racines secondaire ou de poils absorbants, de division cellulaire ou encore de régulation de l'osmolarité. Les nouvelles méthodes de microdissection cellulaire comme celles utilisant des lasers permettent de focaliser l'étude sur un tissu ou un type cellulaire précis (Nakazono *et al.* 2003; Casson *et al.* 2005; Day *et al.* 2005). Cette approche est en

cours de développement au laboratoire et devrait permettre, conjointement à des méthodes d'amplification linéaire des ARNm extrait, d'affiner les études d'expression à l'échelle cellulaire.

B.2.2 Le stress hydrique

B.2.2.1 Les stratégies de réponses

Les végétaux rencontrent fréquemment des variations de leur environnement. Dès lors que les facteurs externes sortent des conditions optimales de croissance de la plante, ils provoquent un stress qui a des effets sur la croissance, le développement et la productivité. Ces stress peuvent être provoqués par des organismes vivants –on parle de stress biotique– ou par des excès physiques ou chimiques de l'environnement –on parle de stress abiotique-.

Parmi les stress abiotiques, le stress hydrique édaphique peut se définir comme un déficit de disponibilité en eau pour la plante. L'origine de ce déficit peut venir d'une salinité excessive du sol, d'une sécheresse ou du gel qui par cristallisation des molécules d'eau diminue sa disponibilité. Mais en général, différents stress sont combinés et un stress ayant pour origine la sécheresse peut également devenir un stress thermique (par augmentation de la température des organes) et un stress oxydatif (par altération des structures cellulaires). Les réponses de la plante sont multiples et si certaines sont spécifiques d'un stress, d'autres sont des réponses communes aux agressions extérieures (Fig. B-11).

Chaque espèce va répondre au stress par des stratégies de **résistance** différentes lui permettant de survivre. Chaves *et al*. (2003) détaillent les mécanismes de résistance à la sécheresse. Ils sont groupés en trois catégories même si ces mécanismes ne sont pas exclusifs et qu'une plante combine toute une gamme de types de réponses:

- L'**échappement** concerne les espèces capables d'accomplir leur cycle de vie avant que le déficit physiologique en eau n'intervienne. C'est le cas des plantes des régions arides effectuant leur cycle reproductif après une pluie et dont les graines supportent de longues périodes de sécheresse. Cette stratégie ne s'applique pas aux plantes longévives comme les arbres.
- L'**évitement** consiste à maintenir le potentiel hydrique et à limiter au maximum les pertes d'eau par transpiration pour se prémunir de l'exposition au stress. Certaines plantes ferment très rapidement leurs stomates, d'autres limitent la masse foliaire, présentent des feuilles modifiées ou conservent les feuilles âgées pour se protéger du rayonnement solaire. Les feuilles sont aussi parfois munies de cuticules épaisses. Une

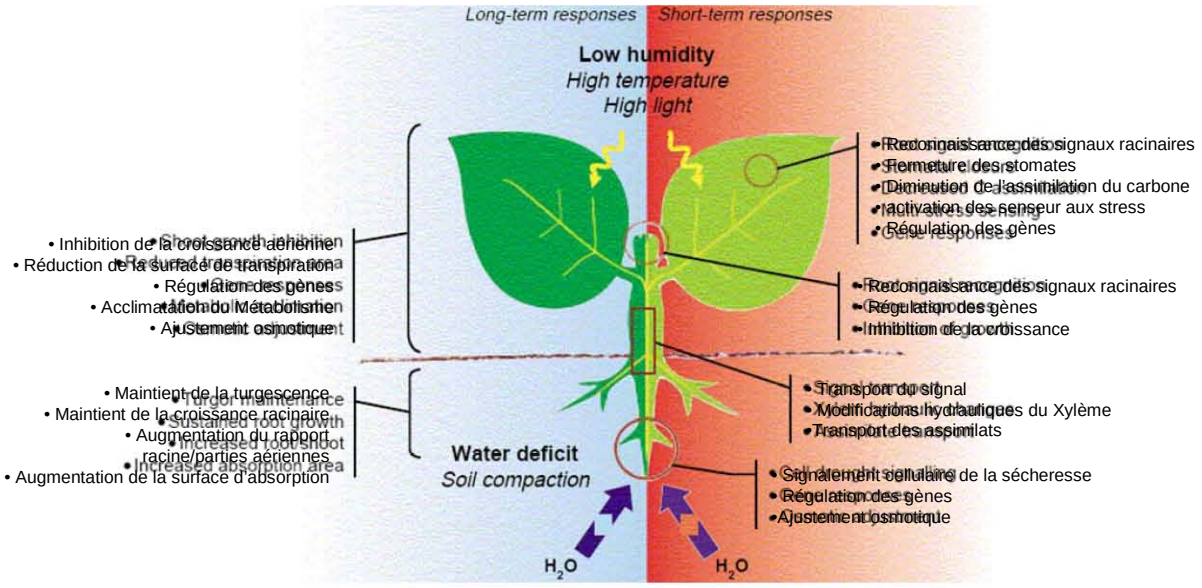


Figure B-11 : Réponses de la plante à la sécheresse. A gauche, réponse à long terme ou acclimatation ; A droite, réponses à court terme. D'après Chaves *et al.* (2003)

allocation du carbone privilégiant un système racinaire profond et développé permet également d'optimiser l'accès aux ressources en eau.

- Les plantes **tolérantes** mettent en œuvre des mécanismes leur permettant de supporter le stress. En diminuant le potentiel hydrique elles maintiennent la turgescence par l'ajustement du potentiel osmotique, obtenu par accumulation d'ions minéraux et/ou de composés organiques.

Mais les plantes ont la possibilité de répondre au stress par une gamme importante de processus que nous détaillerons ci-après.

Les espèces ne présentent pas toutes la même plasticité de réponse au stress. Certaines caractéristiques sont déterminées génétiquement qu'il y ait présence du stress ou non, comme un système racinaire profond, un système foliaire constitué d'aiguilles ou un fonctionnement photosynthétique du type CAM. De telles **adaptations** morphologiques et physiologiques donnent un avantage sélectif à une population au cours de l'évolution.

D'autres mécanismes de résistance concernent les réponses d'un individu aux changements de son environnement. Une telle **acclimatation** va permettre à la plante d'ajuster son homéostasie pour établir un nouvel équilibre compatible à la survie dans ces nouvelles conditions. La résistance d'une plante à un stress sera une combinaison des caractères constitutifs de l'espèce et de sa plasticité propre. Une plante pourra donc être considérée comme appartenant à une espèce évitante (ex : morphologie constitutive) mais d'une variété tolérante (ex : capacité induite d'ajustement du potentiel osmotique).

B.2.2.2 Les systèmes modèles

Pour répondre à une question de recherche et identifier les processus sous-jacent, il est préférable de travailler sur des organismes présentant des aptitudes particulières quant au caractère étudié. Dans le cas de la résistance au stress hydrique, les espèces considérées peuvent se répartir en trois groupes. Nous verrons tout d'abord le cas des systèmes tolérants, puis les modèles génétiques et enfin les plantes de grande culture.

B.2.2.2.1 Systèmes tolérants

Avec une teneur en eau de 10%, la graine est une structure bien adaptée au manque d'eau. Elle permet à la plupart des espèces de survivre à la dessiccation et aux basses températures mais seulement sous cette forme. Cette dessiccation est indispensable au cycle reproductif. Elle est suivit par une réhydratation massive lors de la germination et par le

développement des cellules embryonnaires. Une étude portant sur le pois (*Pisum sativum*), le maïs (*Zea mays*) et le soja (*Glycine max*) a montré que cette tolérance à la dessiccation était progressivement perdue en cas d'imbibition des graines (Koster and Leopold 1988). Les auteurs enregistrent une corrélation entre le pourcentage de levée des graines après traitement (imbibition suivie d'une dessiccation) et la teneur en sucre en fonction du temps d'imbibition. Ingram et Bartel (1996) rapportent que les graines de nombreuses espèces ont fait l'objet d'études approfondies sur le transcriptome et le protéome, telles que *Arabidopsis thaliana*, le coton (*Gossypium spp.*), l'orge (*Hordeum vulgare*), le maïs (*Zea mays*) et le riz (*Oriza sativa*). Un des problèmes majeurs de ce modèle d'étude est de dissocier les gènes impliqués dans la tolérance de ceux impliqués dans le développement de l'embryon. Il a néanmoins permis d'identifier une famille de protéine, les LEA (Late embryogenesis Abundant), que l'on retrouve également fortement exprimée lors de la réponse au stress hydrique. Les LEA sont des protéines hydrophiles qui s'accumulent dans les stades avancés de la maturation de la graine puis disparaissent après la germination (Galau *et al.* 1986; Roberts *et al.* 1993). Le maintien de l'état dormant est fortement lié à la présence d'Acide abscissique (ABA) dont le rôle d'inhibiteur de la germination (Bewley 1997) a été démontré chez le soja, la tomate ou le blé mais pas chez le tournesol.

Le pollen est également une structure à très faible teneur en eau. Comme les graines, les mécanismes du stress sont confondus avec ceux du développement. Mais l'étude des transcrits de pollen a permis d'identifier des gènes impliqués dans la dessiccation. C'est le cas du transcrit LLA23 cytoplasmique identifié sur *Lilium longiflorum* présent uniquement dans le pollen et dont la séquence est proche de celle d'autres protéines impliquées dans le stress (Lp-3, DS2 et ASR) (Huang *et al.* 2000).

Un autre système d'étude regroupe certaines espèces remarquables, capables de tolérer des dessiccations presque totales et de reprendre une turgescence après réhydratation. Ce sont les plantes reviviscentes ou poïkilohydriques parmi lesquelles on peut citer *Xerophyta viscosa* (Baker) et *Craterostigma plantagineum* (Bartels and Salamini 2001) pour laquelle Peter Scott (2000) souligne les modifications du métabolisme intervenant au cours du cycle déshydratation-réhydratation. Les feuilles réduisent leur surface de 85% avec une teneur en eau comprise entre 5 et 10%, les cellules sont fortement concentrées en sucres et des protéines LEA sont fortement exprimées. L'auteur souligne aussi que très peu d'études ont été réalisées sur les racines. Ces systèmes possèdent des mécanismes de protection des cellules très développés et permettent de découvrir et d'étudier de nombreuses protéines du stress. C'est le

cas de *Xerophyta viscosa* (Mowla *et al.* 2002), pour laquelle ces auteurs décrivent une nouvelle enzyme XvPer1 homologue à des antioxydants, ou encore de *Craterostigma Plantagineum* qui a récemment permis de découvrir une nouvelle famille de gènes de réponse à la déshydratation et à l'ABA nommée PTP (Plastid Targeted Protein) (Phillips *et al.* 2002) ainsi que des facteurs de transcription régulant cette réponse tels que CpR18 (Hilbricht *et al.* 2002).

D'autres espèces sont capables de croître à la fois sur des sols à très forte teneur en sel et de survivre à des températures très basses comme *Mesembryanthemum crystallinum*. Elles constituent également de bons modèles d'étude des systèmes de tolérance à la dessiccation (Barkla *et al.* 1999).

Si les mécanismes de réponse sont exacerbés chez ces plantes, ils sont cependant communs aux autres végétaux. Contrairement au système « graines », ils permettent de s'affranchir des effets développementaux.

B.2.2.2.2 Modèles génétiques

Une autre approche porte sur l'utilisation des plantes modèles. L'effort de recherche de la communauté scientifique sur ces espèces a permis de générer une base de connaissances conséquente en physiologie et en génomique. La séquence complète du génome est désormais disponible chez *Arabidopsis thaliana* ("A. thaliana genome" 2000; Initiative 2000), le riz (Yu *et al.* 2002) et le peuplier. Pour d'autres espèces comme le blé, le maïs et la tomate, le séquençage du génome est en cours mais des cartes génétiques fines et des bases de données d'EST et de BAC sont disponibles, rendant possible les approches de clonage positionnel et l'étude de l'expression quasi exhaustive de l'ensemble des gènes. Des collections de mutants ou de variétés permettent d'identifier des individus sensibles, résistants et d'identifier les gènes et voies métaboliques concernées. En 2003, sur *Arabidopsis*, plus de 88'000 insertions par *Agrobacterium* avaient été localisées précisément sur le génome ce qui représentait plus de 21'700 gènes pour lesquels un mutant était désormais disponible (Alonso *et al.* 2003). Hirochika *et al.* (2004) ont recensé les ressources en mutant disponibles chez le riz dont le nombre dépasse le million. Ils soulignent l'intérêt de mettre à disposition de la communauté une banque de séquences bordant les mutations (FST, Flanking Sequence Tag) dont 1373 ont été identifiées en 2005 (Enckevort *et al.* 2005). En effet, la connaissance de la séquence des régions flanquant le transposon permet d'effectuer des recherches électroniques d'homologie pour un gène donné, d'identifier un mutant pour ce gène et de commander les semences pour étudier l'impact phénotypique de cette mutation.

Des mutants ultra sensible au sel (mutants sos) ont permis de déterminer les gènes régulés spécifiquement lors du stress (Gong *et al.* 2001). Les auteurs ont identifié une liste de 89 gènes régulés par soustraction de banques d'ADNc d'*Arabidopsis* sauvage contrôle vs stressé, puis ont comparé leur expression avec les mutants sos. Six gènes ont pu alors être identifiés.
De nombreux autres mutants ont servi à identifier des gènes et protéines impliqués dans la réponse à la sécheresse (Chen *et al.* 2005). Notamment les gènes régulé par les hormones et en particulier l'ABA et les gènes ABI (ABA-insensitive) associés à la réponse à cette hormone (Campalans *et al.* 1999; Bianchi *et al.* 2002; Dong *et al.* 2005).

B.2.2.2.3 Plantes de grande culture

Avec l'agriculture intensive, la question de l'alimentation en eau est devenue primordiale. Les exigences en eau des cultures rendent souvent l'irrigation incontournable alors que les ressources du milieu sont limitées. Les semenciers proposent une large gamme de variétés dont la résistance à la sécheresse est variable. Les surfaces plantées, les climats différents et les degrés de tolérances disponibles font des plantes de grandes cultures un objet d'étude intéressant pour des études de génétique quantitative en conditions naturelles. Les programmes de recherche des grands groupes agroindustriels ont d'ores et déjà brevetés des variétés résistantes ou modifiées génétiquement. Leurs recherches s'orientent actuellement sur des variétés résistantes à la sécheresse (ex : http://www.monsanto.com/monsanto/layout/sci_tech/prod_pipeline/productpipeline.asp).

B.2.2.3 Réponses morphologiques et physiologiques

B.2.2.3.1 Régulation de la croissance

L'impact d'un déficit en eau se traduit par une diminution de la croissance de la plante. Les rendements au champ peuvent chuter très rapidement. Une analyse des pertes de rendement par application de stress hydriques chez le riz a été réalisée par Pantuwan (2000). Selon l'intensité du traitement, le stade de développement et la modalité d'application, des chutes de 18 à 81% ont été enregistrées par rapport au témoin irrigué de façon optimale. L'eau est indispensable à la photosynthèse et à la régulation thermique par évaporation de l'appareil foliaire. En situation de stress, le défaut d'alimentation en eau entraîne une chute de la photosynthèse. En outre, l'eau du sol en solubilisant les éléments nutritifs facilite leur transport jusqu'aux racines et leur assimilation. En situation de sécheresse les flux hydriques sont plus faibles et la nutrition diminuée.

Au niveau cellulaire, le grandissement commence par un relâchement des parois accompagné par une absorption de solutés permettant de maintenir le potentiel osmotique. Sous l'action de la pression osmotique, et du relâchement des tensions des parois, l'augmentation de volume est initiée. Nous le verrons par la suite, la modification des propriétés physiques des parois est sous le contrôle de gènes spécifiques. Ces phénomènes de croissances cellulaires ont été étudiés par Triboulot (1996) dans le cadre de sa thèse.

La relation entre la consommation en eau et la production de composés organiques lors de la photosynthèse est évaluée par l'efficience d'utilisation de l'eau : rapport entre la quantité de biomasse produite et l'eau utilisée par la transpiration. Ce paramètre est pour une part déterminée génétiquement et constitue une caractéristique de l'espèce ou de l'individu et pour une part sous contrôle environnemental (Brendel *et al.* 2002). La perte en eau et l'assimilation du CO_2 sont régulés par la conductance stomatique à la vapeur d'eau et au CO_2, ainsi que par le gradient de concentration de ces deux éléments entre l'intérieur (w_i, c_i) et l'extérieur (w_a, c_a) de la feuille. Une plante va pouvoir augmenter son efficience d'utilisation de l'eau en limitant les pertes en eau par fermeture des stomates et en augmentant sa capacité photosynthétique pour le CO_2 ce qui signifie une diminution de la concentration sous stomatique en CO_2 (c_i bas). Une telle plante consommera moins d'eau à masse de composés organiques produite égale.

La discrimination isotopique du carbone est une mesure indirecte de cette efficience d'utilisation de l'eau. En effet, le carbone existe sous la forme ^{12}C ou ^{13}C dans le CO_2 atmosphérique. Le ^{13}C, de part sa masse plus élevée, diffusera plus difficilement à travers les stomates et sera plus difficilement incorporé lors des réactions biochimiques. Son incorporation finale dans les composés organiques dépendra donc de la conductance stomatique et de la concentration en CO_2 de la feuille, et par conséquent à l'efficience d'utilisation de l'eau. Mais ce ne sont pas les seuls paramètres à intervenir (état hydrique du sol, appareil racinaire, discrimination du carbone d'autres sources..) et la relation directe n'est pas forcément mise en évidence entre espèces chez les conifères (Froux 2002).

En terme de performances, l'avantage d'une efficience d'utilisation de l'eau élevée vis-à-vis de la sécheresse est très discutable, et détaillée dans l'ouvrage édité par Bacon (2004). Les variétés à forte efficacité d'utilisation de l'eau ne pourront pas toujours tirer avantage de cette caractéristique en conditions d'alimentation en eau non limitante et la biomasse produite pourra être, dans ces conditions, inférieure à celle d'une variété à faible efficacité d'utilisation de l'eau. C'est le cas des écotypes landais et marocains. Sur les tests de provenance plantés dans les Landes, la provenance marocaine se caractérise par une vigueur moindre en hauteur

et diamètre par rapport à la provenance landaise. Les mesures de $\delta^{13}C$ sur des cernes de bois ont révélées une valeur plus basse pour la provenance « Tamjout » qui peut être considérée comme évitante en comparaison à la provenance landaise (Nguyen-Queyrens *et al.* 1998).

La sensibilité à la sécheresse en conditions extrêmes peut aussi provenir d'un risque élevé à l'embolie. Il s'agit d'une cavitation des vaisseaux conducteurs qui se remplissent d'air ou de vapeur d'eau. La sève circule sous tension, l'introduction accidentelle d'une bulle d'air –par un insecte, l'abscission d'une feuille, une blessure ou lors d'un gel/dégel- ou une tension trop forte, remplit l'élément conducteur et provoque l'arrêt de la circulation. La propagation du phénomène entre éléments conducteurs voisin va dépendre de la taille et de la morphologie des pores de communication entre vaisseaux. Des pores larges laisseront passer une bulle d'air sous des tensions plus faibles. Les pores des gymnospermes sont protégés par des torus qui fonctionneraient à la manière d'une soupape si une différence de pression est établie de part et d'autre de la ponctuation (Fig. B-12). Il existe une variabilité inter et intra-spécifique à l'embolie. F. Froux (2002) observe des réponses et sensibilités différentes entre *C. sempervirens*, *C. atlantica*, *P. halepensis* et *P. nigra* où *P. nigra* est l'espèce la plus vulnérable et *P. halepensis* la moins vulnérable. Une régulation stomatique efficace semble jouer un rôle en permettant de ne pas atteindre des tensions critiques à l'origine de l'embolie dans une stratégie d'évitement. Cependant, pour les deux espèces du genre *Pinus*, la marge de sécurité entre le potentiel hydrique de fermeture des stomates et celui d'initiation de la cavitation, était pratiquement nul alors que ces deux espèces ont des vulnérabilités différentes. D'autres facteurs interviennent et l'auteur souligne l'importance des parties racinaires où l'embolie peut se produire mais où des mécanismes de « désembolisation » -poussée racinaire nocturne- et de pousse de nouvelles racines latérales est possible.

La conductance hydraulique peut également chuter si les parois des éléments conducteurs ne résistent pas aux pressions négatives élevées. Les parois du xylème peuvent se collapser et bloquer la circulation. Ce phénomène réversible lors de la réhydratation a été mis en évidence par Cochard *et al.* (2004) sur des aiguilles de pins soumises à des déshydratations poussées. L'implication de ce nouveau système de réponse reste à étudier plus précisément.

Le métabolisme et l'expression des gènes spécifiquement impliqués dans ces mécanismes seront une autre source de distinction des variétés face à la résistance au stress hydrique.

B.2.2.3.2 Croissance racinaire

Comme évoqué précédemment, pour une station donnée, la morphologie du système racinaire va déterminer la disponibilité en eau accessible par la plante. Lors d'un stress

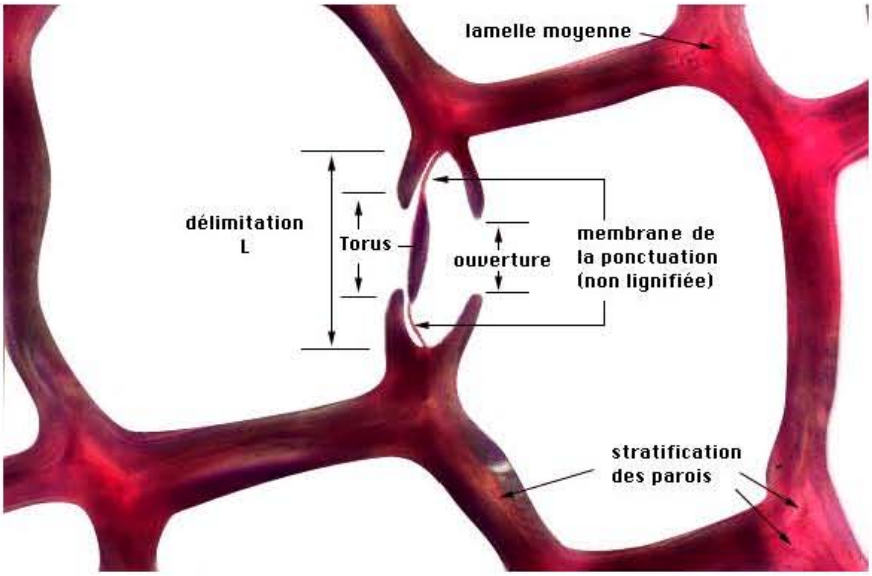


Figure B-12 : Trachéides du xylème secondaire d'une tige de pin blanc (*Pinus strobus*) : détail d'une ponctuation aréolée avec torus. (http://sylva.for.ulaval.ca/foret/anatomie/types_cell/trach_transv_torus.html)

hydrique, la croissance des parties aériennes est généralement réduite mais la plante peut opter pour une stratégie de croissance racinaire. La réallocation du carbone dans cette voie peut lui permettre d'atteindre de nouvelles ressources en eau. Des expériences menées sur *Arabidopsis* confirment cette stratégie avec une augmentation de la biomasse racinaire lors de stress modérés (van der Weele *et al.* 2000). C'est aussi le cas chez le pin maritime pour lequel une valeur plus élevée du ratio de biomasse racinaire sur parties aériennes est observé pour l'écotype marocain (Sarrauste 1982; Loustau *et al.* 1995; Nguyen-Queyrens *et al.* 1998).

Dans une revue récente (Ueda *et al.* 2005) sur le développement des apex racinaires, les auteurs ont classé les gènes connus selon leur rôle dans la destinée et la morphologie du chevelu racinaire, la séparation de l'endoderme et du cortex, ou la spécification des cellules quiescentes et celles de la coiffe. La première catégorie regroupe la majorité des gènes découverts avec par exemple GLABRA2(GL2) qui est une protéine à domaine homéobox jouant un rôle clé dans l'initiation d'un chevelu racinaire et régulée par une liste de facteurs de transcription favorisant sa régulation (WER, TTG1, GLABRA3(GL3), EGL3) ou au contraire l'inhibant (CPC, TRY, ETC1). Les auteurs rapportent également d'autres gènes impliqués comme ACTIN2(ACT2), KOJAK(KjK), OXI1 ou RabA4b, et citent de nombreux gènes dont l'expression est liées à la présence d'une hormone végétale dont nous parlerons ci-après : l'Auxine. Les gènes impliqués dans les mécanismes des deux autres catégories sont pratiquement inconnus à ce jour mais le fonctionnement des apex racinaires est primordial puisque c'est à ce niveau que se décide l'axe de croissance de la future racine.

Jusqu'à présent, très peu d'études sur l'hydrotropisme ont été réalisées. L'orientation de la croissance est en effet également fortement dépendante du gravitropisme et les effets se confondent. Eapen *et al.* (2005) rapportent les résultats obtenus lors de cultures horizontales d'*Arabidopsis* sur des milieux gélosés avec un gradient de potentiel hydrique. La croissance des racines des phénotypes sauvages présente un changement brutal de direction pour éviter la région à potentiel hydrique bas ce qui n'est pas le cas du mutant nhr1 caractérisé par un gravitropisme exacerbé et de volumineux amyloplastes (Eapen *et al.* 2005). Les auteurs citent également d'autres expériences sur des mutants touchant le transport polarisé basipète de l'auxine et insensibles au gravitropisme. Les résultats indiquent qu'une part du gravitropisme aurait pour origine le gradient en auxine de la coiffe racinaire à la zone d'élongation. Ils avancent l'hypothèse que l'auxine interviendrait dans l'hydrotropisme avec une voie similaire à celle du gravitropisme. L'acide abscissique (ABA) pourrait jouer également un rôle au vu de la sensibilité à l'hydrotropisme de mutants aba1-1 et abi2-1 décrit par McCourt (1999). Le comportement de ces mutants montre qu'il y a bien un contrôle génétique de l'hydrotropisme

mais les gènes impliqués ne sont pas encore connus et rien ne nous renseigne sur la nature des senseurs hydriques de cette région. Cependant, les hormones contrôlent fortement le développement racinaire.

L'architecture du système racinaire est basée sur le développement de racines latérales sur des racines existantes. L'application de ce modèle réitératif aboutit à une morphologie complexe. Pour observer l'effet d'un stress osmotique sur le développement des racines latérales, Deak et Malamy (2005) ont placé des plantules d'*Arabidopsis* dans des milieux gélosés contenant des sels, du Chlorure de Potassium ou du Mannitol. Ils ont pu observer une inhibition presque totale du développement de racines latérales. L'initiation des divisions du péricycle n'était pas altérée mais le développement des primordiums était stoppé. Le stress osmotique régule donc le développement des racines latérales. L'Acide abscissique étant connu pour avoir un rôle clé dans les réponses osmotiques et pour bloquer le développement latéral des racines, les auteurs ont testé le comportement des mutants ABA-déficients aba2-1 et aba3-1. Ils ont pu observer un développement du système latéral réduit de 52.2 et 28.4% par rapport au type sauvage présentant une réduction de 97.7% dans un milieu additionné de mannitol ce qui confirme le rôle de l'hormone dans le contrôle de la croissance des racines latérales. Les auteurs ont également analysé un autre mutant –lrd2– leur permettant de découvrir un gène de régulation du développement des racines latérales nommé LRD2 et rapportent d'autres gènes impliqués dans ce phénomène et décrits dans la littérature comme ARN1, LIN1, PDR2 ou NIT3.

A l'inverse, certains facteurs comme l'auxine ou l'oxyde d'azote (NO), favorisent le développement des racines latérales. Sur des plants de tomates, Correa-Aragunde *et al.* (2004) ont étudié le rôle de l'oxyde d'azote sur la présence de racines latérales. Ils montrent qu'un traitement avec un donneur de NO oriente le développement racinaire vers un système latéral tandis qu'un traitement avec un inhibiteur de NO induit un phénotype sans racine latérale. De plus, le dosage de la teneur en NO endogène leur a permis de détecter une accumulation dans les cellules lors du développement de primordiums de racines latérales. Ils confirment ainsi l'implication du NO déjà identifié dans les réponses aux hormones végétales.

B.2.2.3.3 Les hormones

L'acide abscissique et l'éthylène sont couramment qualifiés d'hormones du stress de par leur implication dans les réponses de la plante aux perturbations de son environnement. Si de nombreux composés ont une action physiologique au niveau cellulaire ou tissulaire, une catégorie de molécules, les hormones végétales présentes en très faible quantité ont un effet

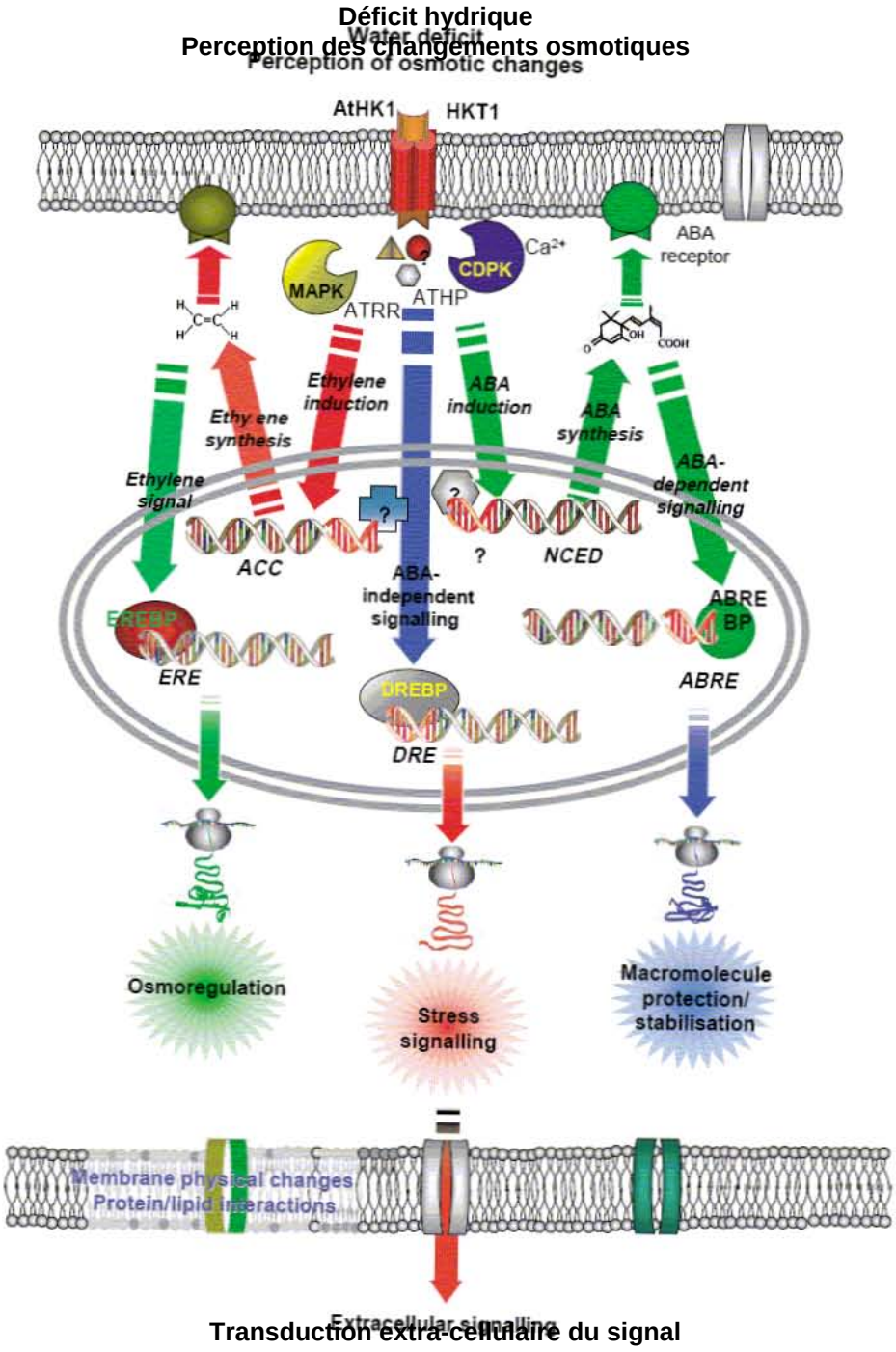


Figure B-13 : Détection, transmission du signal et réponses à la sécheresse au niveau cellulaire (d'après Chaves *et al.* (2003)). Abréviations : ABRE : ABA-responsive element ; ABRE BP : ABRE binding protein ; ACC : 1-aminocyclopropane-1-carboxylic acid ; AtHK1 : Arabidopsis thaliana histidine kinase; ATHP : un phospho-relais intermédiaire; ATRR : un régulateur de la réponse; CDPK : Ca2+-dependent kinases; DRE : dehydratation responsive element; DREBP : DRE binding protein; ERE : ethylene-responsive element; EREBP : ERE binding protein; HKT1 : un transporteur à forte affinité au K+; MAPK : mitogen-activated protein kinase.

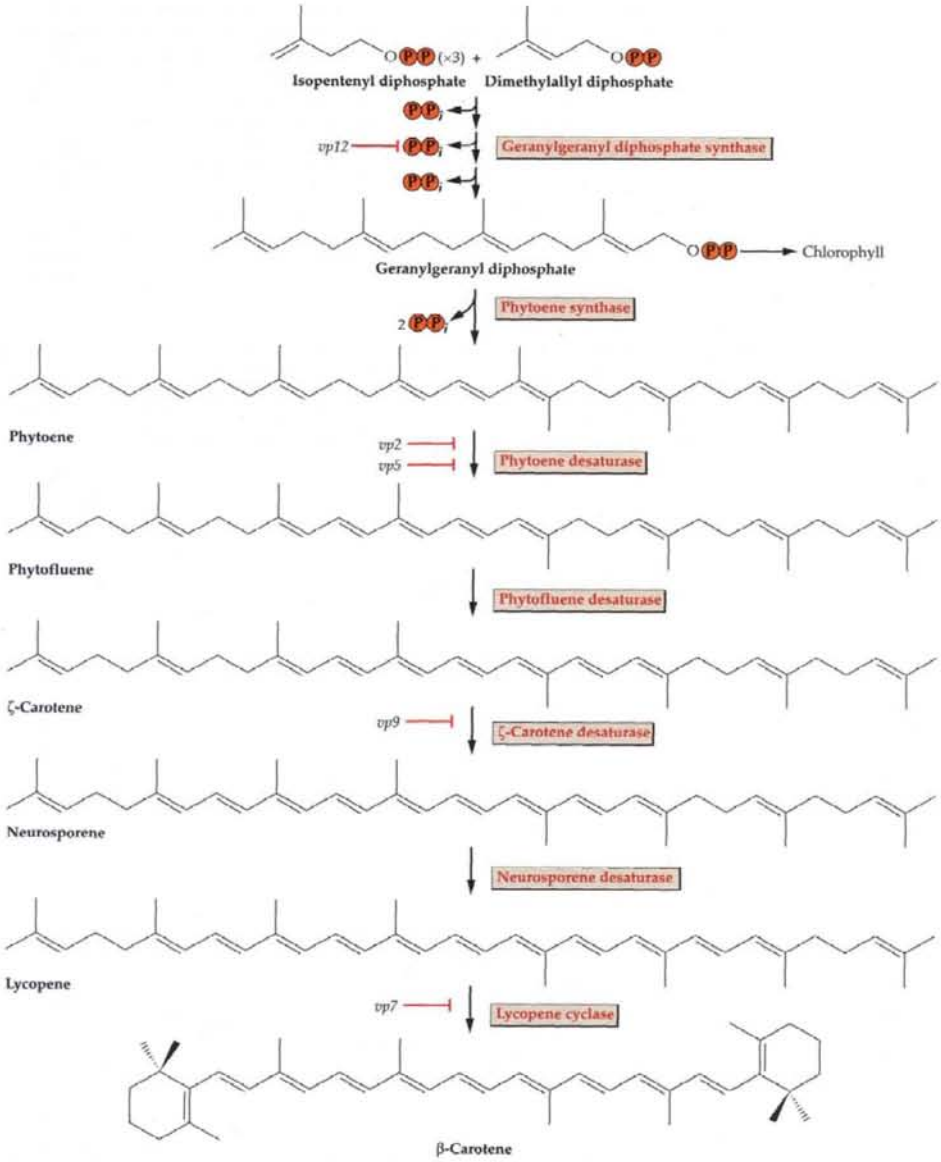


Figure B-14 : Etapes précoces de la voie de biosynthèse indirecte C40 de l'ABA. Production de GPP et synthèse du β-carotène. Les enzymes déficients des mutants vp sont indiqués. (Crozier *et al.*, 2000)

plus global et peuvent circuler au sein de la plante pour modifier l'état physiologique de celle-ci. Elles vont : propager une information du lieu de perception d'un stimulus extérieur vers les zones de réponse de la plante à ce signal, réguler la physiologie des cellules cibles et déclencher des voies métaboliques entières. Chaves *et al.* (2003) structurent ces voies en trois groupes : celles à médiation éthylène contrôlant l'osmorégulation, les ABA dépendantes à l'origine de macromolécules de protection/stabilisation et les ABA indépendantes impliquées dans le signalement du stress (Fig. B-13). Nous passerons en revue dans les paragraphes suivants les caractéristiques et les rôles des différentes hormones végétales.

B.2.2.3.3.1 Acide abscissique

Nous avons déjà évoqué l'implication de l'acide abscissique (ABA) dans la dormance des graines, l'hydrotropisme et l'inhibition du développement des racines latérales. Cette hormone a pour rôle physiologique principal communément admis l'**inhibition de la croissance** et du **développement** des parties aériennes sur des plants bien alimentés en eau, elle agit comme une anti-gibbérelline. Mais de récentes études sur le maïs en condition de stress hydrique indiquerait que ces caractères serait en fait sous le contrôle d'une **balance** entre l'ABA et l'éthylène (Sharp and LeNoble 2002). Elle a également un rôle important en cas de **déficit hydrique** en agissant sur la **fermeture des stomates**.

Sur le plan biochimique, il fait partie des isoprènes aussi appelés terpènes qui dérivent d'un précurseur : l'Isopentenyl pyrophosphate (IPP). La voie de synthèse est bien connue chez les bactéries mais diffère de celle des plantes qui a été récemment corrigée. En effet, pendant longtemps, il était admis que tous les terpènes dérivaient de l'acide mévalonique (MVA) mais une autre voie –voie MVA indépendante ou voie MEP- permettant de synthétiser l'IPP a été découverte chez les plantes supérieures.

Nambara *et al.* (2005) décrivent dans une revue détaillée la voie de synthèse de l'ABA qui est en fait formé par clivage de caroténoïdes (C_{40}) et non à partir du Farnesyl diphosphate (C_{15}). Le β-Carotène est obtenu à partir du Geranylgeranyl diphosphate (GGPP), lui-même issu de la condensation d'IPP et de Dimethylallyl diphosphate sous l'action d'une enzyme clé : la Geranylgeranyl diphosphate synthase (GGPP synthase) (Fig. B-14). Des mutants vp2, vp5, vp9 et vp7 touchant des enzymes de cette voie ont été identifiés chez *Arabidopsis*. Après métabolisme, le β-Carotène est d'abord transformé en Zeaxanthin puis successivement en Antheraxanthin et Violaxanthin sous l'action d'une zeaxanthin epoxidase (ZEP) pour laquelle des mutants caractérisés par une réduction en ABA et une accumulation en Zeaxanthin sont également disponibles. Le clivage de la Violaxanthin et de la Neoxanthin en Xanthoxin est

sous le contrôle d'une famille de 9 enzymes appelées NCED identifiées chez *Arabidopsis* (Bray 2002). La Xanthoxin donne finalement un aldéhyde abscissique et enfin l'ABA sous l'action de 2 enzymes produites par les gènes AtABA2 et AAO3. (Fig B-15). Une liste des gènes connus chez *Arabidopsis* est fournit par Bray (2002).

Une liste des mutants identifiés et des enzymes touchés a été établie par Taylor *et al.* (2000) et Leung *et al.* (1998).Les auteurs décrivent les phénotypes des plants ainsi que l'impact en condition de stress hydrique. D'après la littérature citée par les auteurs, des études sur *Nicotiana sp*. mettent en évidence une augmentation de la concentration en ZEP des racines de plants exposés à la sécheresse. En ce qui concerne les NCED, les racines de plants de tomate et de haricot en condition de stress hydrique présentent de fortes augmentations de concentrations suivies par une teneur en ABA augmentée. Chez *Arabidopsis*, les gènes AtNCED3, AAO3, AtABA3 et AtZEP sont induit lors de la déshydratation mais le gène AtABA2 n'est pas affecté et son expression est constitutive (Taylor *et al.* 2000).

La teneur en ABA, comme celle de tout composé, n'est pas déterminée uniquement par les gènes de sa biosynthèse mais par l'équilibre entre les voies de biosynthèse et de dégradation. Nambara *et al.* (2005) mettent l'accent sur ce point en rapportant le rôle de l'ABA 8'-hydroxylase codée par les gènes CYP707A régulant une des 3 voies de dégradation de l'ABA et induit lors de stress osmotiques ou de la réhydratation (Fig. B-16).

L'ABA est synthétisée dans les racines et dans les feuilles. Sauter *et al.* (2001) reprennent les lieux de synthèse, de circulation et les effets de l'acide abscissique endogène sur des plantes en situation de stress hydrique (Fig. B-17). Lors d'un stress hydrique, l'ABA s'accumule en grandes quantités dans les racines. Il gagne le xylème et circule ensuite jusqu'à l'appareil foliaire où il modifie les flux d'ions K^+ et Cl^- de la vacuole vers le cytoplasme des cellules de garde (Leung and Giraudat 1998). Les stomates se ferment et la conductance stomatique chute. L'acide abscissique peut diffuser hors du système conducteur en empruntant les voies symplasmiques ou apoplasmiques mais la présence d'obstacles comme les bandes de Caspari vont freiner la diffusion. La synthèse d'ABA dans les feuilles apparaît lorsque le potentiel hydrique de la feuille n'assure plus la turgescence mais l'ABA présent dans la sève peut diffuser jusqu'aux stomates avant d'atteindre ce point critique. Une expérience sur le passage de l'ABA dans les tiges tend à soutenir l'hypothèse selon laquelle le transport de l'ABA serait modifié en fonction du pH du milieu cellulaire qui varie si la plante est stressée ou non (Sauter *et al.* 2001; Wilkinson and Davies 2002). Un pH de 6.3, rencontré en conditions normales, favoriserait le passage dans les cellules et donc le captage de l'ABA par les tissus ce qui baisserait la concentration dans le xylème. A l'opposé, un pH de 7.2

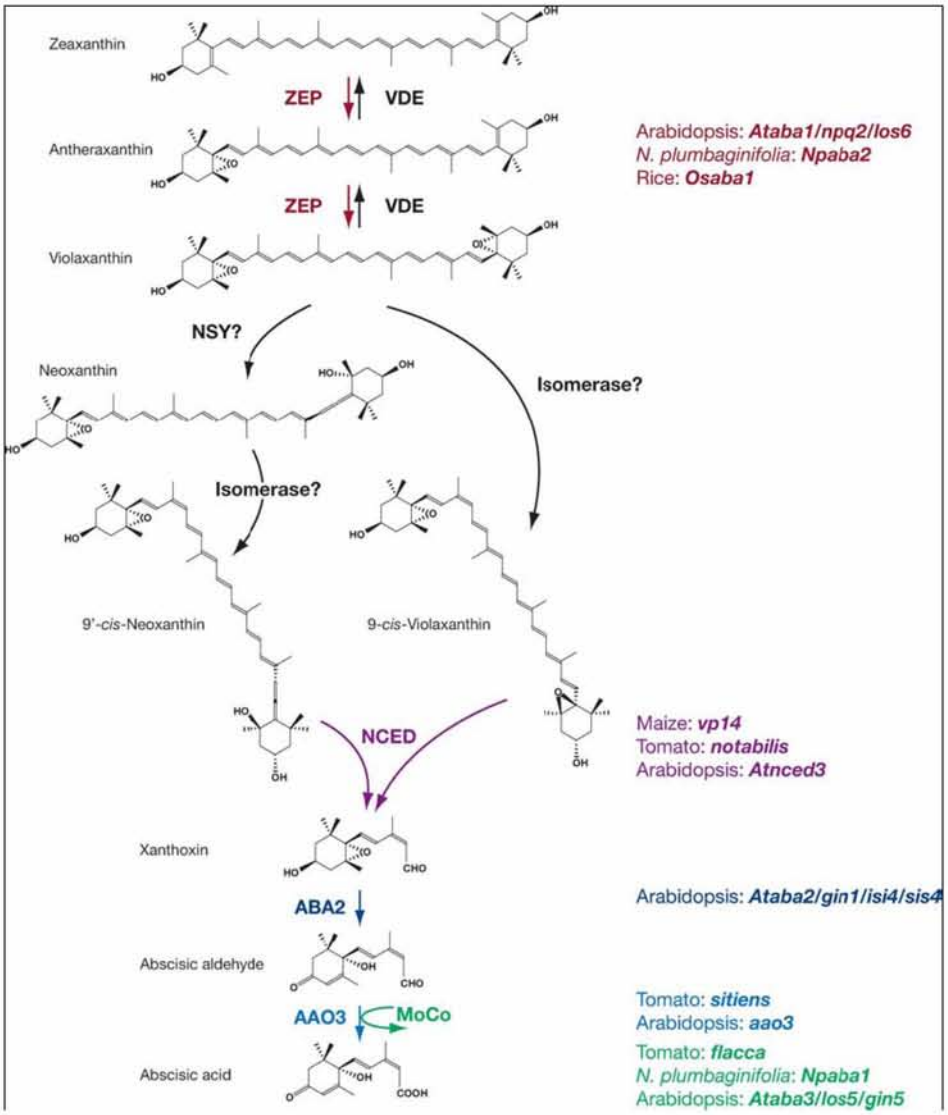


Figure B-15 : Voie de biosynthèse de l'ABA. Les enzymes déficientes pour des mutants de différentes espèces sont indiquées. Abréviations : ZEP : zeaxanthin epoxidase ; VDE : violaxanthin de-epoxidase ; NSY : neoxanthin synthase ; NCED : 9-cis-epoxycarotenoid dioxygenases ; ABA2 : alcool déshydrogénase de chaînes courtes ; AAO3 : abscissic aldehyde oxidase ; MoCo : Molybdenum cofactor sulfurase. (Nambara *et al.*, 2005)

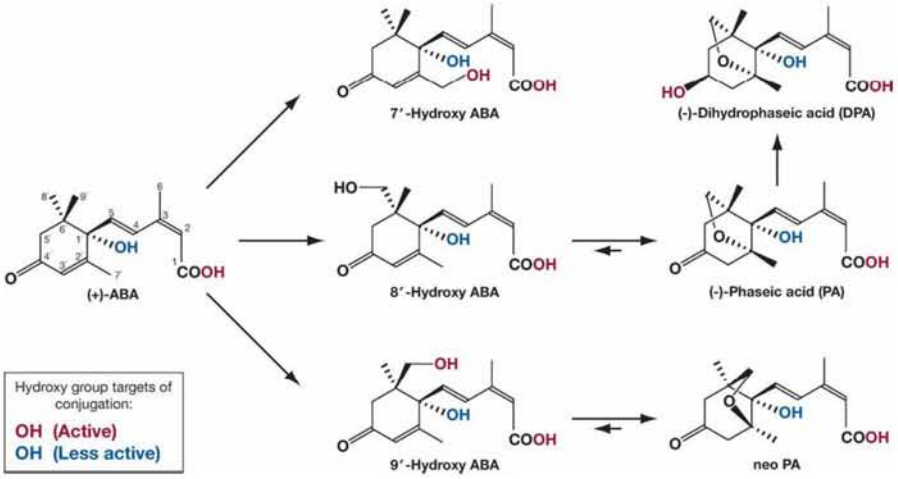


Figure B-16 : Voies de dégradation de l'ABA. Trois voies différentes d'hydroxylation sont présentées. La voie de la 8'-Hydroxylation semble être la voie prédominante de dégradation. (Nambara *et al.*, 2005)

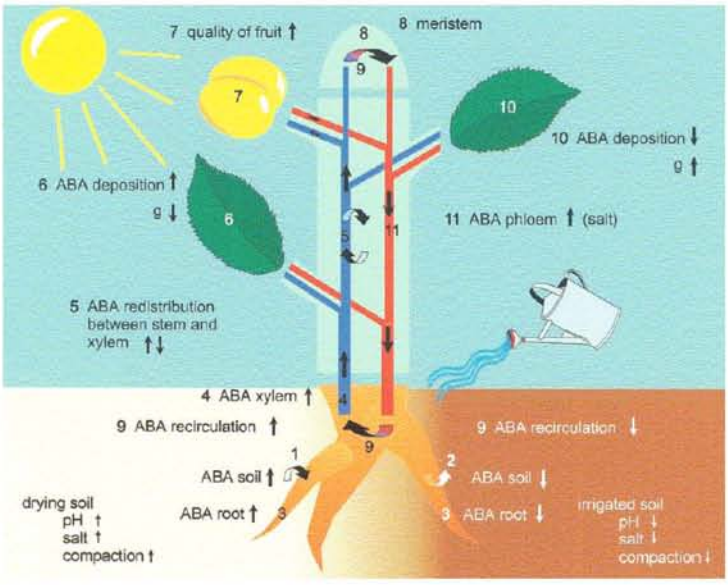


Figure B-17 : Effet de l'état d'hydratation sur l'ABA dans la plante. (Sauter *et al.*, 2001)
1-ABA dans la rhizosphère, 2-Perte d'ABA dans le milieu environnant, 3-Accumulation d'ABA dans les racines par augmentation de la biosynthèse, 4-Transport radial de l'ABA dans les racines, 5-Redistribution de l'ABA entre les tissus de la tige et les vaisseaux du xylème, 6-10-ABA dans les feuilles, 7-ABA dans les fruits, 8-ABA dans les méristèmes, 9-Recirculation de l'ABA.

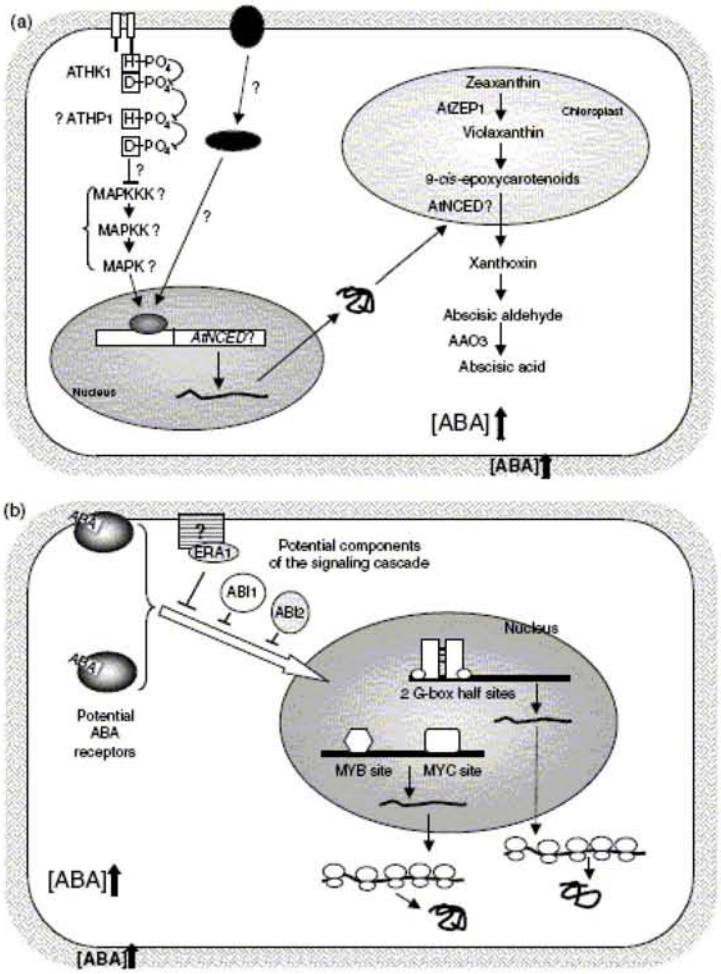


Figure B-18 : Modèle de réponse d'une cellule d'*A. thaliana* soumise à un stress hydrique. a) mécanisme possible de perception du stress avec accumulation d'ABA. b) proposition d'une voie de transmission du signal avec reconnaissance de l'ABA et régulation de l'expression de gènes. (Bray *et al.*, 2002)

correspondant à une plante stressée aurait un captage moindre, une concentration supérieure dans le xylème et l'apoplasme ce qui favoriserait l'accès de l'ABA aux stomates. En diffusant, l'ABA peut circuler également dans le phloème qui va transporter l'ABA synthétisé au niveau des feuilles et des racines. En arrivant dans les racines, il peut re-circuler et représenter une part non négligeable du flux racines vers feuilles.

La perception du signal permettant l'initiation de la biosynthèse de l'ABA est mal connue mais certains mécanismes sont proposés au vu des mécanismes observés chez *E. coli* ou *S. cerevisiae* (Bray 2002). Un ARNm d'*Arabidopsis* codant pour une protéine ATHK1 similaire aux protéines du complexe de perception des stress osmotiques des bactéries ou levures pourrait faire partie d'une cascade de perception du signal. Le senseur activerait cette ATHK1 qui pourrait interagir avec une autre protéine ATHP1 puis une cascade de type MAP kinases par l'intermédiaire de phosphorylations. Il y aurait alors activation de la transcription des ARNm codant pour les enzymes de la voie de biosynthèse de l'ABA telle que les AtNCED déjà citées.

Les mécanismes de régulation des gènes par l'ABA restent également flous. Les récepteurs à cette hormone ne sont pas encore identifiés mais les mutants Abi1 et Abi2 évoqués plus haut supposent l'existence d'une cascade de transmission du signal reliant des récepteurs à la régulation de la transcription par l'intermédiaire d'Era1 (farnésyl-transférase), Abi1 et Abi2 (Fig. B-18). L'hypothèse de l'existence de cette cascade de signalisation semble corroborée par des résultats obtenus en exposant des graines d'*Arabidopsis* à la protéine HrpN d'un pathogène connue pour stimuler la croissance et la résistance aux pathogènes. Sur le génotype sauvage, cette exposition stimule la résistance à la sécheresse de plants en situation de stress hydrique en activant la synthèse d'ABA qui provoque une fermeture des stomates. Par contre, les mutants Abi1-1 et Abi2-1 ne présentent pas ces caractéristiques. L'expression des gènes ABI1 et ABI2 répond de la même façon que le génotype sauvage mais l'expression du gène ABH1 n'est pas stimulée. ABH1 interviendrait après ABI1 et ABI2 dans la cascade de régulation (Dong *et al.* 2005).

Les motifs cibles de la régulation de la transcription par l'ABA sont par contre mieux identifiés. Il existe des éléments promoteurs de régulation en CIS ; nous citerons les ABRE, les motifs MYC et MYB. Les ABRE (ABA responsive element) sont des séquences particulières –« ACGT containing G-box » (Leung and Giraudat 1998)- sur lesquelles se fixent les facteurs de transcription bZIP qui sont des protéines constituées d'un domaine de fixation à l'ADN suivit d'une « leucine zipper ». Les motifs MYC -ACACATGT- et MYB –YAAC(G/T)G- se retrouvent chez les gènes induits lors de stress sévères (Bray 2002). La

recherche des motifs possibles se poursuit et de nouveaux motifs sont identifiés, ils sont regroupés sous le nom de ABFs (ABRE binding factors) (Choi *et al.* 2000).
Les gènes régulés spécifiquement par l'ABA sont difficilement identifiables car confondus avec des voies de régulations indépendantes de l'ABA ou déclenchées par d'autres hormones comme l'éthylène lors du stress. Néanmoins, une approche utilisant des mutants déficients pour la voie de l'ABA en condition de stress a permis à Chen *et al.* (2000) d'identifier par électrophorèse bidimensionnelle des polypeptides différentiellement exprimés dans les racines d'*Arabidopsis* mais le séquençage des polypeptides et leur identification n'a pas été réalisé (Chen *et al.* 2005).

Le nom de l'hormone vient de sa découverte lors de recherches sur l'abscission des fruits du cotonnier par Addicott *et al.* (1963). Mais l'abscission est maintenant attribuée à l'action de l'éthylène. Lors d'une étude sur l'exposition simultanée à un stress hydrique (sécheresse) et à l'ABA de deux génotypes contrastés de peuplier, Chen *et al.* (2002) ont observé une abscission des feuilles marquée chez le génotype sensible contrairement au génotype tolérant. Le suivi des concentrations en ABA, en éthylène et en polyamines –une autre classe d'hormones– révèle que la concentration foliaire en ABA enregistrée chez le génotype sensible est accompagnée d'une augmentation de la teneur en éthylène et d'une baisse en polyamines. Ces voies sont bien corrélées par une compétition notamment pour un précurseur commun (S-adenosylmethionine), mais c'est l'éthylène qui est à l'origine du phénomène d'abscission (Chen *et al.* 2002).

B.2.2.3.3.2 Ethylène

Cette hormone a tout d'abord une action sur les graines en étant capable de **lever la dormance**. Elle favorise également la **maturation des fruits** et déclenche **l'abscission des feuilles** qui est contrôlée par une balance Auxine/Ethylène (Heller *et al.* 1993). La migration de l'auxine dans les tissus serait ralentie par l'éthylène. Son action sur la croissance se traduit par une inhibition de l'**élongation racinaire** au profit de la croissance radiale (Crozier *et al.* 2000). Sur les racines, l'application d'éthylène induit la formation d'un **chevelu racinaire**. Des expériences de transformation génétique permettant de diminuer la teneur en éthylène des plants ont permis de mettre en évidence d'autres propriétés de cette hormone. La diminution de la teneur a pu être réalisée en dégradant le précurseur direct à l'aide de l'ACC désaminase. Ainsi, une tolérance au nickel a été observée chez des plants d'*Arabidopsis* transformés par ce gène ce qui souligne sont implication dans la réponse aux stress (Stearns *et al.* 2005). La transformation de bactéries n'exprimant pas initialement ce gène a permis de favoriser la

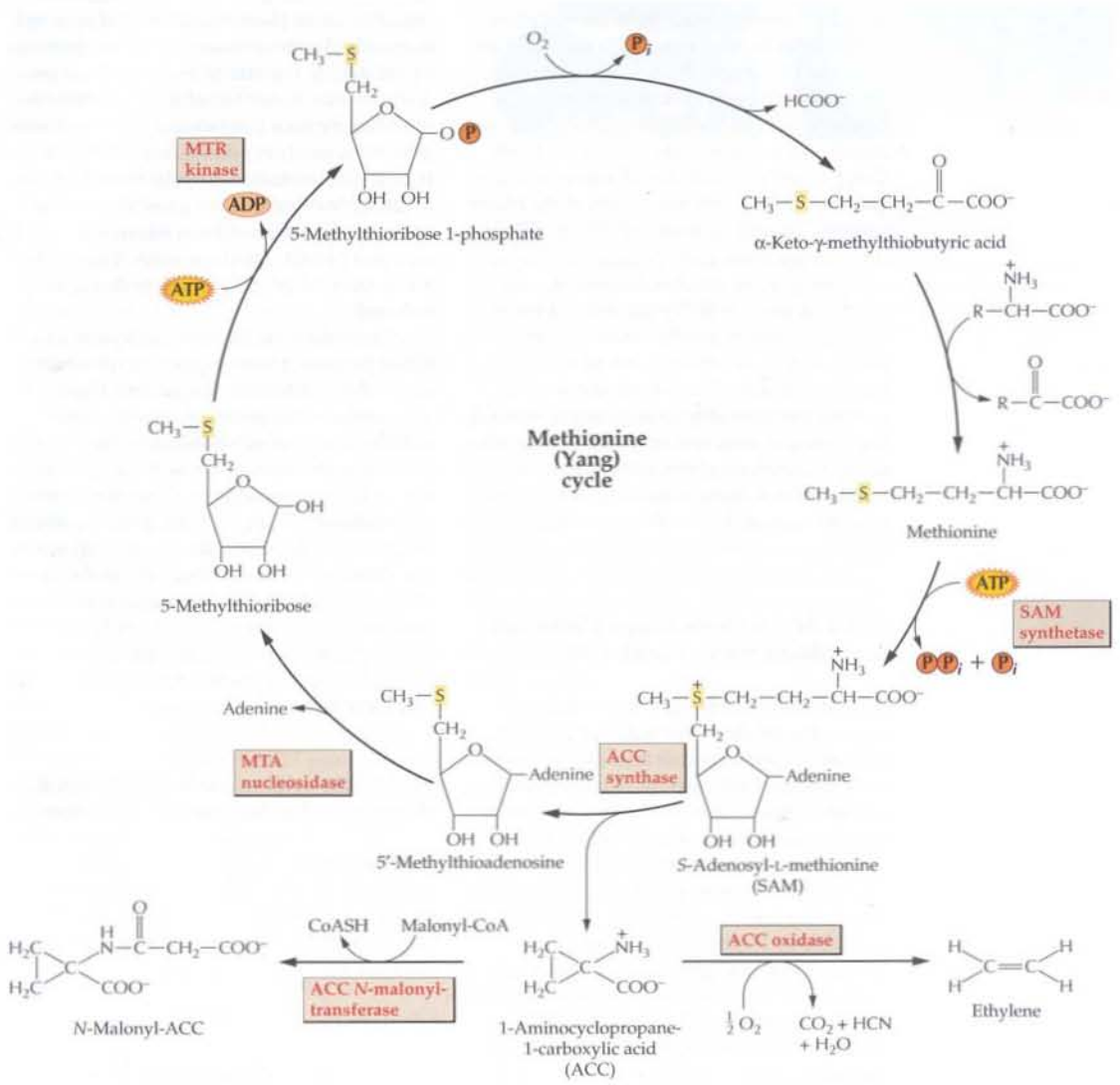


Figure B-19 : Cycle de la méthionine et biosynthèse de l'éthylène. L'éthylène est synthétisé à partir de la méthionine par l'intermédiaire de la SAM et de l'ACC. Les enzymes qui catalysent ces trois étapes sont l'ATP methionine S-adenosyl-transférase (SAM synthase), la S-adenosyl-L-methyl-thioadenosine-lyase (ACC synthase) et l'ACC oxydase. D'après Crozier *et al.* (2000).

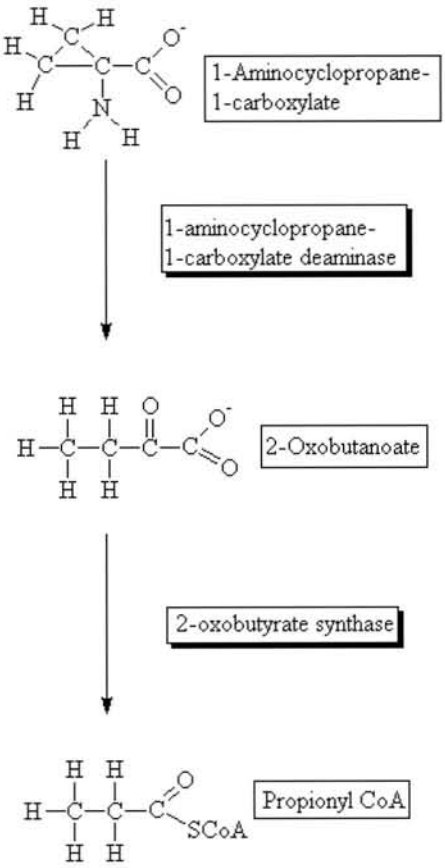


Figure B-20 : Action de l'enzyme 1-Aminocyclopropane-1-carboxylate désaminase (ACC deaminase) sur le précurseur (ACC) direct de l'éthylène.

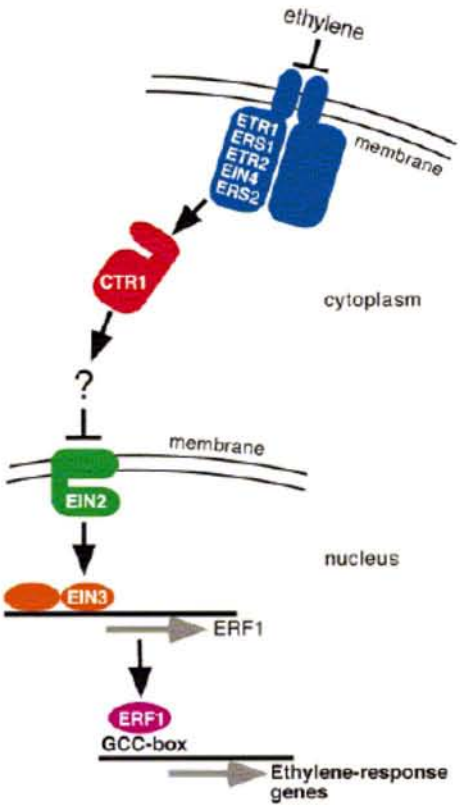


Figure B-21 : Schéma de la voie de transduction du signal. Cinq récepteurs à l'éthylène (ETR1, ERS1, ETR2, EIN4 et ERS2) transmettent le signal éthylène à la protéine kinase CTR1 qui présente des similarités avec la famille des MAPKKKs. EIN2 agit en tant que domaine membranaire. La localisation spécifique des récepteurs à l'éthylène et de EIN2 sont inconnus. Le facteur de transcription EIN3 favorise la transcription de ERF1 qui code pour un membre de la famille des facteurs de transcription EREBP. ERF1 va se lier aux promoteurs à boîte GCC pour activer la transcription des gènes de réponse à l'éthylène. (Chang *et al.*, 2001)

nodulation de plants de pois et de trèfle par respectivement *Rhizobium leguminosarum* et *Sinorhizobium meliloti* (Ma *et al.* 2003; Ma *et al.* 2004). Le rôle de l'éthylène dans la réponse au stress n'est pas clairement mis en évidence car les effets de l'ABA et de l'éthylène sont étroitement liés. Les expériences de Sharp et LeNoble (2002) soulignent ce phénomène et suggèrent une inhibition de la croissance des pousses et des racines de maïs en condition de stress. Elle aurait pour origine une concentration élevée en éthylène induite par des promoteurs de croissance de ces jeunes plants et une concentration relative en ABA faible.

L'éthylène est un gaz, sa formule à 2 carbones est très simple C_2H_4. Il est synthétisé à partir de la méthionine (Fig. B-19) qui sous l'action de la SAM synthase donne de la **S-Adenosyl-L-méthionine** ou **SAM.** Ce composé intervient dans la synthèse des polyamines mais permet aussi d'obtenir de l'**ACC** (acide 1-Aminopcyclopropane-1-carboxylique) par une réaction catalysée par l'**ACC synthase** qui fait partie du cycle de la méthionine, appelé également cycle de Yang. L'ACC produit peut alors donner différents composés dont l'éthylène sous l'action d'une **ACC oxydase,** mais aussi du 2-Oxobutanoate ou α-ketobutyrate puis du Propionyl CoA dans une réaction catalysée par l'**ACC désaminase** (Fig. B-20), ou encore du N-Malonyl-ACC avec l'aide d'une ACC N-malonyl-transférase. La régulation de la production de l'éthylène est principalement dépendante de l'ACC synthase. Des régulations peuvent cependant intervenir en aval par compétition sur l'ACC qui est le précurseur direct de l'éthylène en donnant l'α-ketobutyrate et l'ammonium par l'ACC désaminase ou pour la synthèse des polyamines. Cette voie est bien identifiée chez les bactéries du sol capables d'exprimer une ACC désaminase pour contrôler la production d'éthylène et favoriser leur interaction avec la plante (Ma *et al.* 2003; Ma *et al.* 2004). Cet enzyme est souvent détecté lors de l'étude des promoteurs de croissance des rhizobactéries (PGPR) (Penrose and Glick 2003). De nombreuses expériences de transformations génétiques de plantes exposées à des stress et exprimant une ACC désaminase d'origine bactérienne démontrent la diminution de la production d'éthylène (Hontzeas *et al.* 2004; Stearns *et al.* 2005). Si la plupart des hormones nécessitent une voie de dégradation, l'éthylène peut être éliminé simplement sous forme de gaz. Les gènes codant pour les enzymes de la voie de biosynthèse sont bien décrits dans la littérature et Johnson *et al.* (1998) fournissent une liste des gènes identifiés à ce jour.

L'éthylène agit sur des récepteurs membranaires dérivés de la famille Histidine-Protéine kinases qui par phosphorylation transmettent le signal à un aspartate (Chang and Stadler 2001) (Fig. B-21). Parmi les récepteurs d'*Arabidopsis* classés en sous familles, on citera les

récepteurs ETR1 (ethylene resistant), ERS1, ETR2, EIN4 et ERS2 (Chang and Bleecker 2004). Le mutant etr1 ne présente pas un certain nombre de réponses à l'éthylène comme l'activation de la germination, la stimulation de l'activité peroxydase ou l'accélération de la sénescence (Kieber 1997). Chang *et al.* (2004) rapportent que les études récentes indiquent que des facteurs de transcription EIN3 et EIL1 sont nécessaires à la transmission du signal capté par les récepteurs dans la phase lente de réponse à l'éthylène (par opposition aux réponses initiées dans les premières minutes suivant exposition au gaz). Une protéine membranaire EIN2 découverte récemment agirait en amont de EIN3. EIN2 serait un élément clé, les mutations de ce gène bloquent totalement les réponses à l'éthylène. Les mutants etr1 ou ein2 ont une teneur basale en éthylène 5 à 10 fois supérieure ce qui suggère qu'une voie de régulation négative de la biosynthèse serait déficiente. La liste des gènes identifiés et leur position supposée dans la cascade de transduction du signal est représentée dans la figure B-22. Les mutants se répartissent en deux classes, la première regroupe ceux qui surproduisent de l'éthylène (Eto) et la seconde regroupe les mutants conservant leur réponse en présence d'inhibiteurs de la synthèse et de la fixation de l'éthylène ce qui suggère plutôt une perturbation de la transduction du signal (Kieber 1997). Un élément clé de cette transduction est la protéine kinase CTR1 (Johnson and Ecker 1998).

A l'issue de cette cascade de transmission du signal, les gènes régulés possèdent des séquences particulières identifiées en étudiant les gènes PR (Pathogenesis-related). Ces séquences ERE possèdent un motif AGCCGCC ou « GCC box » nécessaire et suffisant à la régulation des gènes porteurs. Puis un domaine de fixation à l'ADN appelé domaine AP2 a été identifié. Les protéines capables de reconnaître ces séquences sont appelées EREBPs (ERE-binding proteins) et pourraient être les cibles de régulations initiales de l'éthylène.

Au niveau racinaire le gène EIR1, présent chez des mutants insensibles à l'éthylène, détermine l'agravitropisme qui est un caractère rencontré chez des mutants pour la sensibilité à l'Auxine. Cela suggère un rôle carrefour de EIR1 entre les voies de l'éthylène et de l'auxine.

B.2.2.3.3.3 Polyamines (PA)

Cette catégorie d'hormones est plus abondante dans la plante que les gibbérellines ou les cytokinines. Ce sont des polycations de faible masse moléculaire. Les membres les plus couramment rencontrés sont la putrescine, la spermidine et la spermine (Fig. B-23). La putrescine est synthétisée à partir de l'arginine par deux voies, l'une faisant intervenir la L-Ornithine et l'Ornithine décarboxylase, et l'autre catalysée par l'Arginine décarboxylase passe

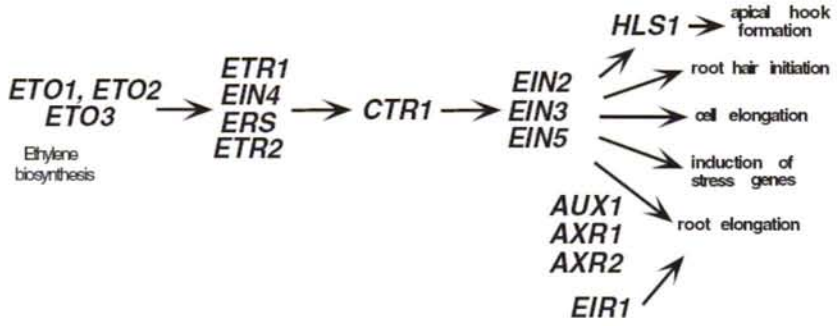


Figure B-22 : Proposition de l'organisation des gènes impliqués dans la production et la perception de l'éthylène. A droite, différentes réponses possible à l'éthylène sont indiquées.

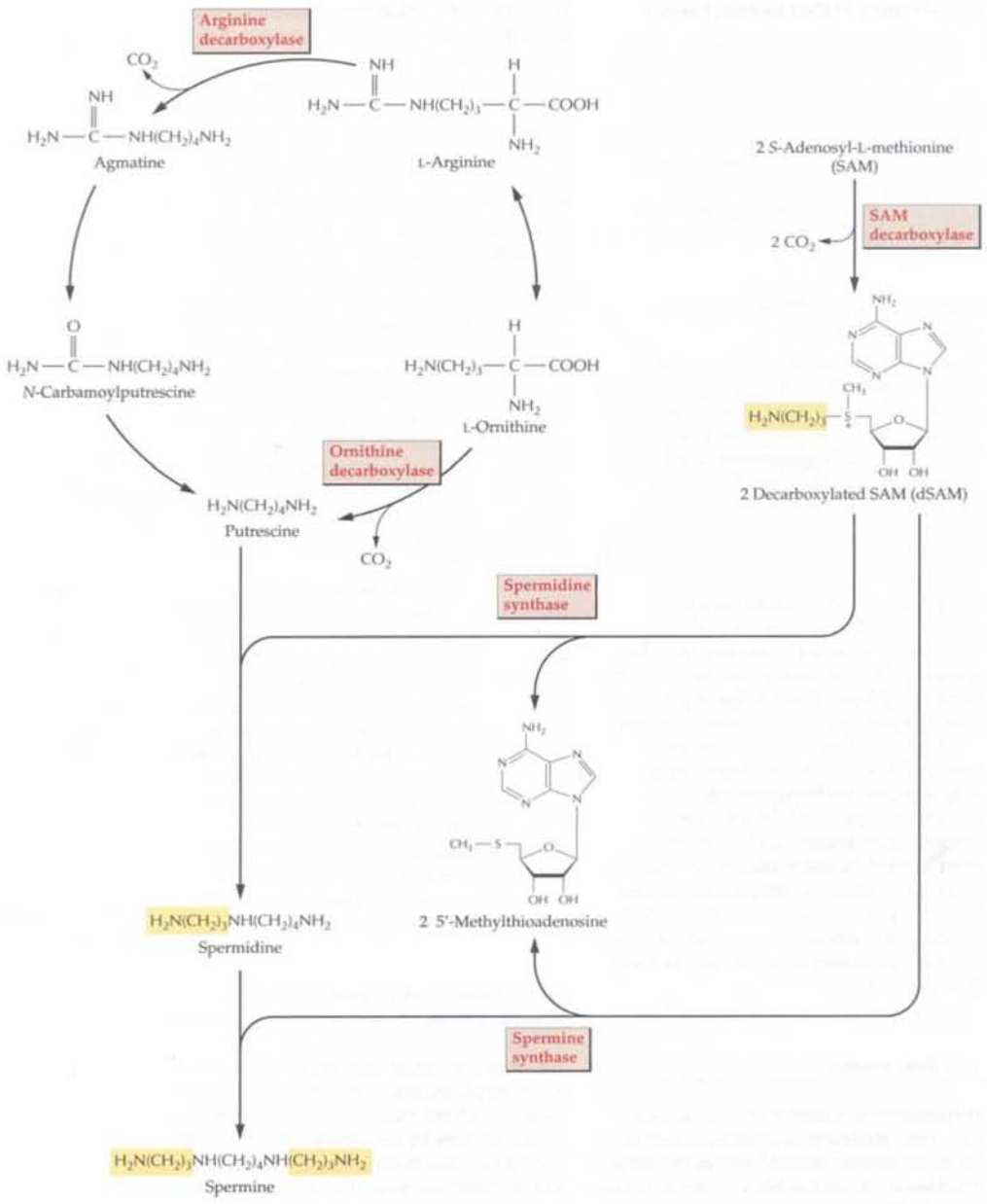


Figure B-23 : Biosynthèse de la putrescine, spermidine et spermine. Chez la plupart des plantes, l'arginine décarboxylase est plus active que l'ornithine décarboxylase. (Crozier *et al.*, 2000)

par l'Agmatine et le N-Carbamoylputrescine. La Spermidine et la spermine dérivent de la SAM par action de la SAM décarboxylase puis de la Spermidine synthase ou de la Spermine synthase (Martin-Tanguy 2001). Leur rôle exact n'est pas complètement connu. Cependant, l'utilisation de mutants des gènes de la voie de biosynthèse a permis d'obtenir de nombreux phénotypes tels que : croissance ralentie, entre-nœuds courts, branchaison des tiges ou petites feuilles. Les mutants présentaient également une teneur en éthylène fortement augmentée ce qui confirme la compétition pour un précurseur SAM commun (Walden *et al.* 1997). Il a été montré que lors de stress osmotiques, les plants tolérants ont des concentrations en polyamines supérieures aux plants sensibles (Tiburcio *et al.* 1994). Une des hypothèses rapportées par Gallardo *et al.* (2003) quant à leur rôle serait que leur biosynthèse pourrait être un mécanisme de détoxification de l'ammonium en conditions de stress. La nature polycationique de ces molécules leurs confère un fort pouvoir de liaison aux molécules chargées négativement au pH physiologique telles que les acides nucléiques, les protéines ou les membranes phospholipidiques. En condition de stress, ils interagissent avec les membranes et les stabilisent (Roberts *et al.* 1986) et pourraient aussi agir comme destructeur des radicaux libres. L'exposition de plants de pois chiches à des stress osmotiques et à des stress froid a permis à Nayyar et Chander (2004) d'observer une augmentation de la teneur en PA lors du stress. Par contre la nature des PA n'était pas la même pour les deux stress. En stress froid il s'agissait de putrescine tandis qu'en stress osmotique cela concernait la spermidine, ce qui confirme la littérature. Les auteurs ont également observé que l'application exogène de PA -plus particulièrement la putrescine et la spermidine, dans une moindre mesure la spermine- favorisait la croissance des plants stressés. L'application d'inhibiteurs des enzymes de biosynthèse de ces PA s'est traduit par des dommages marqués lors du stress (Nayyar and Chander 2004). Une revue de Couée *et al.* (2004) met l'accent sur l'implication des PA dans le développement racinaire. Les auteurs nous renseignent sur la nature des PA détectées dans les racines. En plus de la putrescine, de la spermidine et de l'Agmatine, il existe des formes conjuguées de PA comme la Caffeoylputrescine et la Feruloylputrescine, et des formes liées insolubles. Les PA jouent également un rôle dans le développement racinaire primaire, latéral et adventif. Couée *et al.* (2004) rapportent des études montrant une corrélation entre la quantité de PA et la formation des racines adventives, la croissance des racines primaires, la prolifération et le cycle cellulaire.

Après avoir revu en détail les hormones les plus impliquées lors d'un stress hydrique, nous allons décrire succinctement les trois hormones végétales majeures impliquées dans un

grand nombre de régulations physiologiques. Elles ne sont pas en soit caractéristiques de la réponse au stress mais interviendront dans la balance hormonale de la plante.

B.2.2.3.3.4 Les auxines

Synthétisées dans les apex des tiges (ou dans les tiges chez certains ligneux comme le Frêne), elles migrent dans la plante jusque dans les racines. Leur action principale est de favoriser la **croissance cellulaire** en stimulant l'élongation des cellules. Cependant, au niveau racinaire l'action est opposée, avec une **inhibition de la croissance**. L'action mitogène est particulièrement marquée dans les zones génératrices libéro-ligneuses c'est-à-dire les cellules cambiales. Elles ont également un **pouvoir rhizogène** fort dans le cas d'une action limitée dans le temps car l'effet inhibiteur sur la croissance racinaire gène le développement des jeunes ébauches. Elles interviennent également sur le développement des bougeons et sont responsables du phénomène de dominance apicale.

L'auxine joue également un rôle de stimulation de la production d'éthylène qui lui-même réduit la migration de l'auxine. L'importance de cette balance Auxine/Ethylène se retrouve au niveau de la zone d'abscission des feuilles des espèces caduques. La chute des feuilles est une des stratégies de réponse au déficit hydrique prononcé. La réduction de la surface foliaire permet de limiter les pertes par évapotranspiration et de réduire les risques de cavitation. La biosynthèse de l'AIA (acide indole-3-acétique) s'effectue à partir du Tryptophane et plusieurs voies sont possibles. La principale passe par la synthèse d'acide indole-pyruvique et d'indole acétaldéhyde. Dans une revue, Bartel (1997) reprend en détail la biosynthèse de l'auxine.

B.2.2.3.3.5 Les Gibbérellines

L'effet le plus remarquable de ces hormones est le gigantisme induit lors de son application. Les gibbérellines (GA) provoquent l'**élongation des entre-nœuds** du à une élongation et une **prolifération des cellules** de la tige. A forte concentration on peut également observer une augmentation de la **taille** des feuilles. Elles induisent la formation des cônes chez les conifères. Elles ont aussi une action opposée à celle de l'ABA sur la dormance des graines et des bourgeons. Les GA peuvent donc être qualifiées de **régulateur de croissance**.

Comme l'ABA, ce sont des composés terpéniques mais à 19 ou 20 carbones. La synthèse est très active dans les apex des jeunes pousses et les racines. Elle s'effectue à partir d'Isopentenyl-pyrophosphate (IPP) qui par polymérisation donne finalement du Geranylgeranyl-pyrophosphate (GGPP) (Fig. B-24). Le GGPP se cyclise en Kaurène, sous

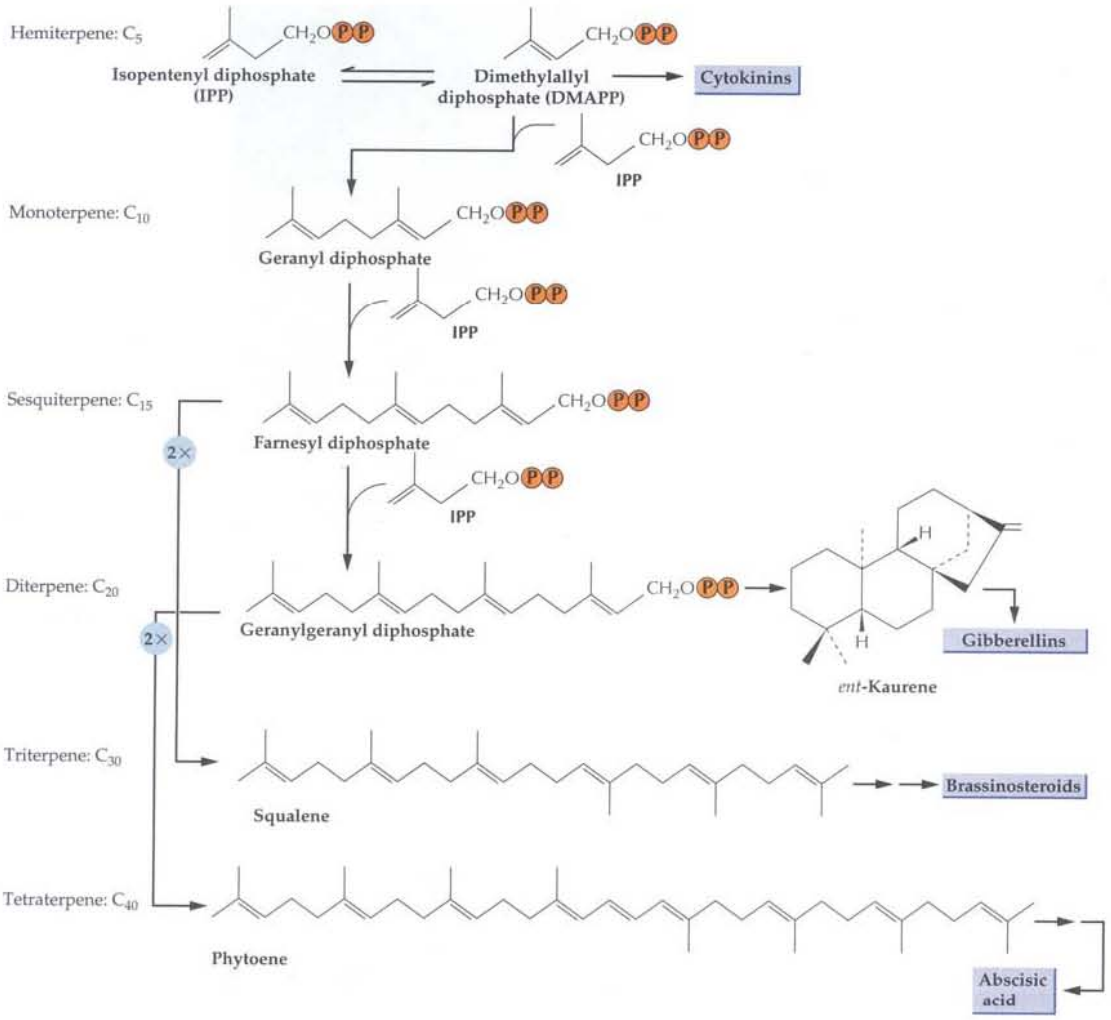


Figure B-24 : Voie de biosynthèse des terpènes montrant l'origine terpénique de la biosynthèse des Gibbérellines, Cytokinines et de l'ABA. (Crozier *et al.*, 2000)

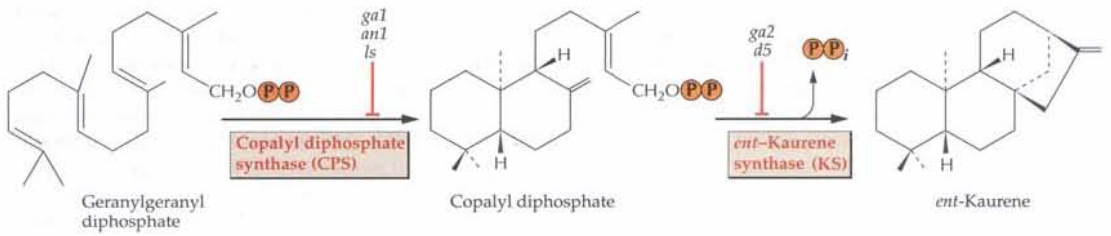


Figure B-25 : Biosynthèse du Kaurene, précurseur de synthèse des Gibbérellines, à partir du GGPP. Les mutants d'Arabidopsis et du pois pour les enzymes catalysant les réactions sont indiqués. (Crozier *et al.*, 2000)

l'action de la Copalyl diphosphate synthase (CPS) et de l'ent-Kaurène synthase (KS) (Fig. B-25). Le Kaurène est le point de départ de toutes les gibbérellines. Pour une revue plus précise des voies de biosynthèse, des récepteurs et mutants identifiés et des effets physiologique, se référer à la publication de Richards *et al.* (2001).

B.2.2.3.3.6 Cytokinines

Les cytokinines regroupent un ensemble de molécules qui provoquent entre autres réactions, la division des cellules et la néoformation des bourgeons conjointement à l'auxine. Le ratio Cytokinine/Auxine va déterminer le développement des pousses –ratio élevé- ou le développement racinaire –auxine seule-. Elles se retrouvent à l'état libre dans les tissus ou fréquemment associées à du ribose pour former des ribosides. Leur lieu de synthèse de prédilection semble être la région apicale des racines. Aloni *et al.* (2004) ont pu montrer par immunolocalisation et gène rapporteur GUS cette localisation apicale. De plus, ils ont mis en évidence que le gravitropisme des racines serait sous le contrôle des auxines et des cytokinines. En effet, la distribution asymétrique de la zéatine dans la zone apicale correspond à la direction de croissance adoptée par la racine ce qui s'explique par l'action de régulation négative sur la croissance racinaire par les cytokinines (Aloni *et al.* 2004). En agissant sur la croissance racinaire, les cytokinines peuvent jouer un rôle dans le développement des racines en réponse au stress hydrique. La kinétine (Amasino 2005) et la zéatine sont des cytokinines naturelles. Biochimiquement, elles dérivent de l'adénine avec une base isopentenique attachée à l'azote N_6 du groupement aminé.

D'autres molécules que nous n'aborderons pas ici ont également un rôle de régulation dans la plante comme les brassinostéroïdes, l'acide jasmonique et l'acide salicyliques (Yang *et al.* 2004).

B.2.2.4 Taille du génome et adaptation au stress hydrique

Avec 2n = 2x = 24 chromosomes, le caryotype est remarquablement uniforme à ceux des autres pins du genre Pinus, tant sur le plan de la morphologie que du nombre (Murray 1998). Les chromosomes sont de grande taille mais en nombre limité et représentent env 50pg/2C pour *Pinus pinaster* (Murray 1998; Chagne 2004). La totalité du génome n'est pas codante et la proportion d'ADN en simple copie représente uniquement 14-24%. Une étude réalisée sur *Pinus strobus* par Elsik *et al.* (2000), divise son génome en fonction du niveau de redondance. Les séquences simple copie constituent 14% soit plus de 3000 Mb, ce qui en

taille dépasse déjà le génome des céréales. Les régions répétées sont classées en zones faiblement (20%), moyennement (45%) et hautement (14%) répétées. Une très grande partie du génome est donc constituée de séquences répétées comme par exemple des retro-transposons. Le rôle de cet excès d'ADN répété et un éventuel avantage sélectif pour l'espèce n'est pas encore très bien expliqué. Quelques études ont mis en avant une relation entre la taille des génomes et les caractéristiques adaptatives des plantes et en particulier les gymnospermes à leur habitat (précipitation et température) (Wakamiya *et al.* 1993). La théorie se base sur une relation entre la taille du génome, celle des cellules et celle de la paroi cellulaire qui impliquerait au niveau des cellules des trachéides une différence de conductivité et caractéristique hydraulique. La taille des génomes des pins semble en effet être corrélée à des paramètres physiologiques comme le potentiel hydrique minimum avant flétrissement (Wakamiya *et al.* 1996) qui est directement relié à la capacité de résistance à des potentiels osmotiques du sol bas en cas de sécheresse. Cependant, si une étude à plus grande échelle, portant sur 117 gymnospermes (Murray 1998) confirme ces observations pour certaines espèces, dans d'autres cas l'hypothèse est rejetée. La taille du génome est variable au sein du genre Pinus mais selon l'espèce, des variations intra-spécifiques ou au contraire une taille stable comme chez *Pinus taeda* peuvent être observées. Aucune conclusion générale ne peut donc actuellement être avancée quant à l'implication de la taille du génome sur une aptitude à résister à la sécheresse.

Une étude exhaustive et systématique du génome tel que les programmes de recherche conduit sur *Arabidopsis thaliana* ou *Populus tremuloides* ne sont pas applicable à l'espèce étudiée. Si l'on regarde la taille du génome du genre Pinus (Fig. B-26) et en particulier de *Pinus pinaster*, associé à la forte proportion d'ADN répété, on se rend compte qu'il n'est actuellement pas envisageable d'effectuer un séquençage systématique du génome. Il est toutefois aisément possible d'accéder au génome exprimé dans une condition et pour un organe donné par une extraction des ARN transcrits.

B.2.2.5 Gènes et voies métaboliques régulées au cours du stress

Les techniques de biologie moléculaire et de biotechnologie ont permis ces deux dernières décennies d'identifier de nombreux gènes régulés lors du stress hydrique. Face au stress, nous avons vu que la plante répond en régulant sa physiologie par l'intermédiaire de nombreux mécanismes à médiation hormonale notamment. Au niveau cellulaire, les signaux de stress doivent en premier lieu être perçus par des récepteurs spécifiques. Ils vont permettre d'activer des cascades de transduction du signal qui entraîneront la transcription et

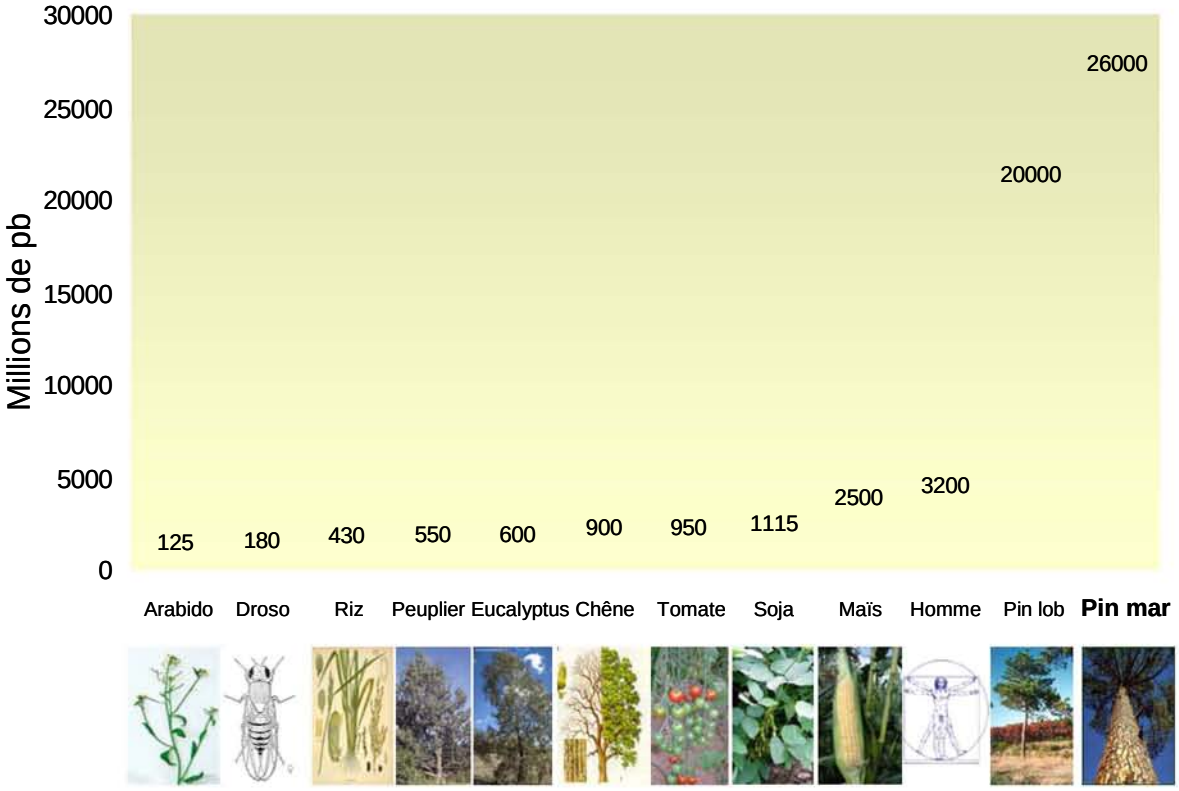


Figure B-26 : Schéma comparatif de la taille physique respective du génome diploïde de *A. thaliana* (125 Mb/C, 0.35pg/2C1, 5 chromosomes), le riz (430Mb/C, 1pg/2C1, 12 chromosomes), le chêne (900Mb/C, 1.85pg/2C, 12 chromosomes)[1], l'homme (3200Mb/C, 6.4pg/2C, 23 chromosomes), le pin maritime (25700Mb/C, 51pg/2C, 12 chromosomes)2.

1: Royal Botanic Gardens, Kew (www.rbgkew.org.uk).
2: Chagné *et al.* (2002)

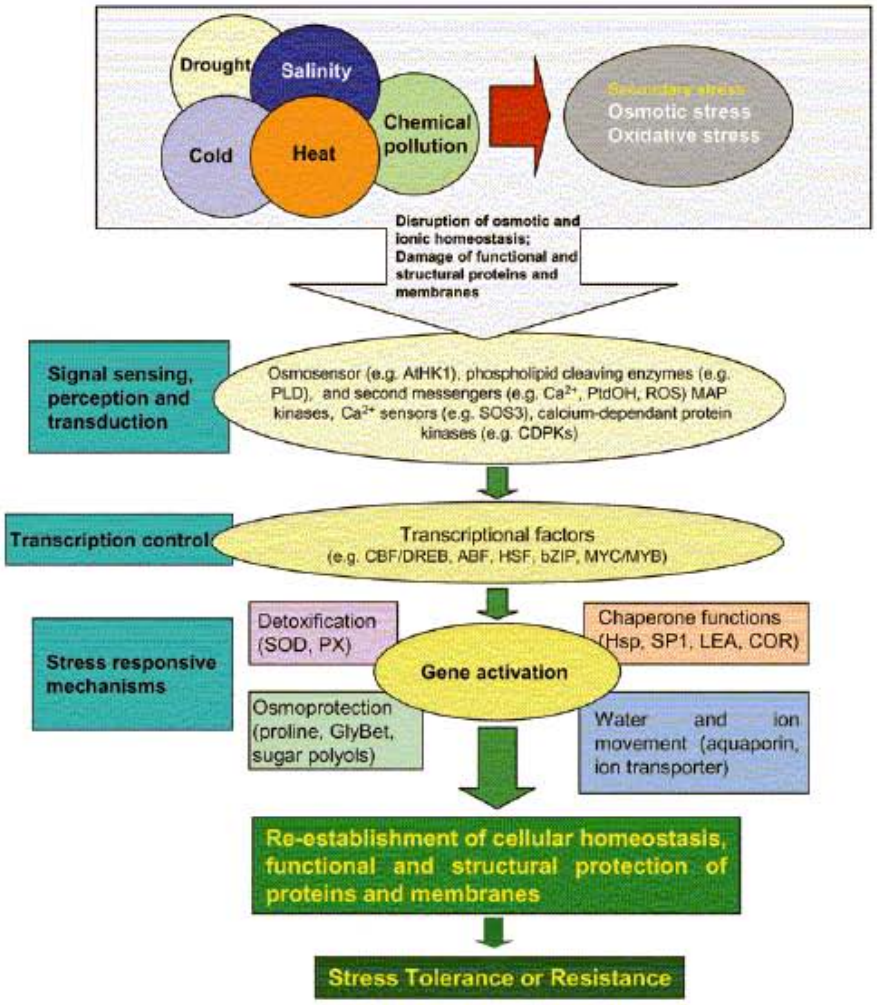


Figure B-27 : Processus de la perception du signal à la mise en place des mécanismes de résistance et/ou tolérance par la régulation de l'expression des gènes impliqués dans la réponse de la plante à des stress abiotiques. Les gènes régulés peuvent intervenir dans la détoxification, l'osmoprotection, les fonctions chaperonnes et les flux d'eau et d'ions. (Wang *et al.*, 2003)

l'expression de gènes spécifiques. La figure B-27 synthétise ce processus et regroupe les gènes régulés en 4 catégories : détoxification, osmoprotection, fonctions chaperonnes ou contrôle des flux ioniques et hydriques. Dans ce chapitre, nous allons reprendre les gènes identifiés et impliqués dans ces différentes étapes ou fonctions. Les gènes cités ne sont pas tous induits spécifiquement par un stress hydrique mais peuvent intervenir pour rétablir l'homéostasie cellulaire quelque soit la source de perturbation.

B.2.2.5.1 Méthodes de détection des gènes régulés

L'identification des gènes régulés lors d'un stress peut se faire à l'échelle d'un gène ou sur un ensemble de gènes.

La première approche implique généralement une connaissance à priori du gène étudié. C'est le cas lorsque la teneur d'un composé est régulée au cours des conditions expérimentales et que la voie de synthèse biochimique et les enzymes concernés sont connus. Il est alors possible de rechercher le(s) gène(s) codant(s) pour les enzymes et d'utiliser des mutants déficients pour le gène d'intérêt. En comparant les phénotypes du mutant et du sauvage, il sera possible de confirmer ou d'infirmer l'implication du gène dans le caractère étudié. Le recours à la transgénèse est une autre option. La sur ou sous-expression du transgène choisi va alors permettre de vérifier l'hypothèse de son rôle dans la réponse au stress appliqué.

Si aucune connaissance sur la fonction n'est disponible sur un gène, des études du niveau d'expression dans les conditions expérimentales sont possibles. La PCR quantitative ou le Northern par exemple, vont permettre de déterminer le niveau d'expression du gène dans chaque condition.

La deuxième approche est plus adaptée si aucun gène n'est privilégié. Il s'agit d'identifier des gènes dont l'expression est régulée et/ou corrélée avec les conditions expérimentales appliquées. De nombreuses techniques permettent de repérer les transcrits régulés au cours du stress au sein de l'ensemble des ARN totaux. Les principales sont les techniques SSH d'hybridations suppressives soustractives –qui permettent d'éliminer les transcrits communs aux différents échantillons-, les techniques d'identification de polymorphisme sur les ADNc (ADNc-AFLP) issus des différentes conditions, et les techniques d'hybridation où le niveau d'expression des gènes dans les différentes conditions est obtenu par hybridation des ADNc sur des séquences d'ADN choisies et fixées sur des membranes (Filtres) ou des lames de verre (puces à ADN). Chaque séquence correspond à un

gène donné. Il est possible par cette méthode d'étudier simultanément plusieurs milliers de gènes.

Les différentes approches adoptées en génomiques pour étudier la tolérance d'une plante au stress ont été reprises dans une revue de Cushman *et al.* (2000).

B.2.2.5.2 Détoxication

Les cellules stressées comme les cellules non stressées produisent des composés oxydants. Les plantes disposent de systèmes de défense pour les éliminer. La production et la dégradation sont en équilibre dans les cellules non stressées. Lorsqu'une plante est exposée à de basses températures, à la sécheresse ou à d'autres stress, la quantité de composés oxydants augmente. Ce sont des produits de réactions impliquant le peroxyde d'hydrogène (H_2O_2) et les ions superoxyde (O_2^-) qui sont regroupés sous le nom d'AOS (active oxygen species). Ces **AOS** ont un pourvoir oxydatif très fort capable de causer des dommages aux cellules qui en réponse synthétisent des composés spécifiques chargés d'éliminer ces AOS. Les principaux acteurs de cette détoxification sont les superoxyde dismutases (**SOD**), les ascorbate peroxydases (**APX**), les catalases (**CAT**), les monodehydroascorbate réductases (**MDAR**), les dehydroascorbate réductases (**DHAR**) et les glutathione réductases (**GR**) (Iba 2002). La SOD va permettre de transformer les radicaux superoxydes en O_2 et H_2O_2 qui sera pris en charge par l'APX et la CAT. Les MDAR, DHAR et GR vont permettre la régénération de l'acide ascorbique utilisé par l'APX.

La famille des SOD est divisée en trois grands groupes en fonction du cofacteur utilisé. On distingue les SOD à Manganèse (Mn SOD) –présentes dans les mitochondries et les peroxisomes-, les SOD à Fer (Fe SOD) –localisées dans les chloroplastes- et les SOD à Cuivre et Zinc (Cu-Zn SOD) qui sont divisées en deux groupes, l'un représenté dans le cytoplasme et les peroxysomes, et l'autre chloroplastique et extracellulaire (Alscher *et al.* 2002). L'augmentation de la SOD lors d'un stress va permettre d'éviter une dégradation des constituants cellulaires par les **ROS** (Reactive Oxygen Species), mais de nombreux gènes codent pour des SOD et tous les gènes ne seront pas régulés par le stress. Chez le maïs il a été identifié 4 Cu-Zn SODs cytosoliques (SOD-2, SOD-4, SOD-4A et SOD-5), 4 Mn SODs mitochondriales et une Cu-Zn SOD chloroplastique. L'étude des gènes SOD4 et SOD4A chez le maïs par Guan *et al.* (1998) montre que ces deux gènes sont similaires en tout point, excepté pour la région promotrice. Les expériences réalisées sur des embryons, des mutants et des feuilles de jeunes plants permettent d'avancer que les deux gènes répondent différemment à l'ABA aux différents stades de développement étudiés. Le gène SOD4 est surexprimé à 8h

et 12h dans les feuilles des plants exposés à un stress osmotique, et il s'agit d'une réponse sous le contrôle de l'ABA. L'expression du gène SOD4A sera dépendante du tissus et du stade de développement (Guan and Scandalios 1998).

Les autres acteurs de la détoxification sont également codés par des familles de gènes complexes régulées par des voies métaboliques qui peuvent être reliées ou non à une réponse au stress. Ainsi, chez *Arabidopsis*, Jespersen *et al.* (1997) distinguent 52 séquences réparties en 7 groupes codant pour des ascorbate peroxydases. Les Catalases sont des assemblages de monomères. Un nombre de gènes faible permet d'obtenir un nombre de combinaison important et l'identification de la régulation de l'activité enzymatique va dépendre de la combinaison des régulations individuelles des gènes codant chaque unité (Frugoli *et al.* 1996).

Les mécanismes de protection de la cellule aux ROS sont interconnectés et difficiles à comprendre. En dehors des composés déjà cités, ils font appel au cycle des xanthophyles qui fait intervenir des agents complexant, des pigments et des composés facilement oxydables comme la glutathionne (GSH) et des enzymes tels que la Glutathionne S-transférase (GST) (Light *et al.* 2005).

B.2.2.5.3 Fonctions chaperonne

Les stress abiotiques sont à l'origine de dysfonctionnements protéiques. La plante va activer de nombreux gènes permettant d'accumuler des protéines parmi lesquelles des **HSP** (Heat-shock protéins), des **LEA** (Late embryogenesis abundant proteins) et autres chaperonnes qui vont jouer un rôle de protection de la cellule. Elles sont responsables du repliement protéique, de l'assemblage, de la translocation et de la dégradation dans de nombreux processus cellulaires. Elles vont stabiliser les protéines et les membranes et permettre de retrouver les conformations protéiques actives en condition de stress (Wang *et al.* 2003).

B.2.2.5.3.1 Les Hsp (Heat-shock proteins)

Les Hsp sont divisées en cinq familles en fonction de la taille des protéines produites : les Hsp70 (DnaK), les Hsp60, les Hsp90, les Hsp100 (Clp) et les sHsp (small HSP) dont la taille varie de 12 à 40 kDa. Si elles ont été mises en évidence lors d'expositions à des températures élevées, on les trouve également surexprimées lors de l'exposition des plants à la sécheresse, aux fortes teneurs en sel, à des agressions chimiques ou des stress oxydatifs. Les Hsp60, Hsp70 et les Hsp90 vont interagir avec de nombreuses protéines co-chaperonnes qui vont réguler leur activité ou faciliter leur action de repliement protéique. Les Hsp vont

également interagir entre elles pour former un véritable réseau de protéines chaperonnes et agir en synergie avec les autres mécanismes de réponse au stress (Wang *et al.* 2004). Wang *et al.* (2004) ont réalisé une revue très complète sur les fonctions connues des différentes familles d'Hsp.

Les Hsp70 vont empêcher la formation d'agrégats protéiques et permettre aux protéines de se replier pour adopter une conformation active. Elles sont aussi impliquées dans les transports protéiques, la transduction du signal et l'activation de la transcription. Les Hsp70 répondent également aux attaques de pathogènes. L'infection de plants d'*Arabidopsis* par le virus TuMV ou le virus TCV induit l'expression des gènes AtHSC70-1, -2, -3 et AtHSP70 mais pas AtHSP70B (Aparicio *et al.* 2005). Les auteurs montrent également que la quantité de protéine HSP70 cytosolique en se liant à des HSF (heat shock factor) exercerait un rétrocontrôle sur l'expression du gène HSP70. Ce type de régulation concernerait aussi les sHsp (Rhoads *et al.* 2005). La surexpression du facteur HSF3 chez *Arabidopsis* se traduit par une accumulation de Hsp en l'absence de choc thermique et confère aux plants une tolérance accrue à la chaleur (Prandl *et al.* 1998).

Les Hsp60 dont le rôle essentiel est de faciliter le repliement protéique, sont essentiellement chloroplastiques et mitochondriales, même si certaines formes se retrouvent dans le cytoplasme.

En plus du rôle de contrôle du repliement, les Hsp90 (Milioni and Hatzopoulos 1997) se distinguent en intervenant dans les réseaux de transduction du signal, le contrôle du cycle cellulaire, la dégradation des protéines et la destination des protéines.

Les Hsp100/clp vont interagir avec les Hsp70 et les sHsp pour désagréger et déplier les complexes protéiques et permettre leur solubilisation puis leur repliement dans la conformation correcte. L'étude de plusieurs mutants d'*Arabidopsis* thermosensibles portant des mutations pour le gène HOT1 codant pour la protéine AtHsp101 et la recherche de suppresseurs de mutations a permis aux auteurs de restaurer la solubilité d'agrégats de sHsp après un stress thermique (Lee *et al.* 2005). Ce travail révèle une interaction entre la famille des Hsp100/Clp et les sHsp.

Les sHsp ont en commun un domaine conservé de 90 acides aminés appelé α-crystallin domain (ACD). C'est la famille la plus exprimée et répandue chez les plantes. Les sHsp ont une très grande affinité pour se lier aux protéines dans leur forme non native et vont empêcher leur agrégation. Elles interagissent alors avec les protéines chaperonnes ATP dépendantes du complexe DnaK pour replier les protéines sous leur forme native. Chez la tomate, la Hsp21 est induite dans les feuilles et les fruits par un stress thermique. L'expression constitutive de

cette protéine chez la tomate protège le Photosystème II des stress oxydatifs combinant l'exposition au froid et à de fortes lumières appliquées lors de l'expérience (Neta-Sharir *et al.* 2005). Les auteurs ont également observé une accumulation de caroténoïdes dans les fruits des lignées transgéniques.

B.2.2.5.3.2 Les LEA (Late Embryogenesis Abundant Proteins)

Les LEA sont des protéines hydrophiles qui s'accumulent très fortement dans les tissus tolérants à la dessiccation -comme les graines– ou en réponse à la déshydratation, les basses températures, la salinité et l'application d'ABA. Distribuées très largement chez les mono- et dicotylédones, c'est initialement sur le coton qu'elles ont été découvertes. Plusieurs groupes -nommés d'après l'identifiant des clones d'ADNc- ont été formés sur la base de leur similarité de séquence. Le Groupe 1 (D19) est caractérisé par un motif commun de 20 acides aminés. Les LEA du Groupe 2 (D11) aussi appelé **Déhydrines** ou **RAB** partage 3 domaines hautement conservés : le segment-K qui est un motif de 15 acides aminés riches en lysine, une chaîne de sérines contigües ou segment-S, le segment-Y qui est un motif DEYGNP. La protéine adopterait également une conformation en hélice α. Les LEA du Groupe 3 (D7) partages un motif de 11 acides aminés permettant une conformation en hélice α. A ces trois groupes ce sont ajoutés : le Groupe 4 (D113), le Groupe 5 (D29) et le Groupe 6 (D34) ayant également des propriétés communes (Ramanjulu and Bartels 2002). Cette classification a été récemment remise en question par Wise (2003). L'auteur a réalisé une étude bioinformatique en prenant en compte la séquence, la distribution en acides aminés, la présence de motifs, domaines ou conformations spécifiques ainsi que les fonctions supposées associées. Les résultats proposent une nouvelle distribution des groupes existants avec le rapprochement possible de certains d'entre eux (Wise 2003). Toutes les LEA, excepté celles du groupe 5, sont très hydrophiles et stable à haute température souvent même après ébullition. Les fonctions des LEA restent largement méconnues. Mais si l'on considère leur haute solubilité associée à de fortes concentrations, une distribution en acides aminés biaisée et la présence de conformation en hélice α, il est peut probable que leur fonction soit enzymatique. Par contre leur thermo stabilité, leur structure, leur accumulation dans les graines et lors de stress hydriques suggèrent que les LEA jouent un rôle fondamental dans la tolérance à la dessiccation. C'est ce que laisse supposer les observations sur la quantité uniformément répartie dans l'embryon et la localisation cytosolique de deux LEA chez le coton (Roberts *et al.* 1993). Des hypothèses ont été avancées sur leur rôle possible de soluté permettant de conserver une hydratation minimum des cellules et de protéger les structures cytoplasmiques.

La fonction de protection des protéines par les LEA a été montrée chez le pois. L'accumulation des transcrits de la PsLEAm -une LEA du groupe 3 localisée dans la matrice mitochondriale– s'effectue dans les dernières étapes de maturation de la graine puis disparaissent avec la réhydratation lors de la germination. L'induction de cette LEA peut être obtenue dans les feuilles en soumettant les plants à un stress hydrique violent. L'expression a pu être maintenue plusieurs heures après germination par un traitement des graines à l'ABA (Grelet *et al.* 2005). Les auteurs ont montré que la PsLEAm protégeait des effets de la dessiccation deux enzymes mitochondriales (Rhodanese et Fumarase) et augmentait même l'activité de la Rhodanese. La régulation de l'expression des LEA est fréquemment mise en évidence sur des plants exposés à la sécheresse ou à des stress osmotiques. Chez le peuplier, l'expression d'une Déhydrine a été mesurée lors d'un arrêt d'arrosage, de l'application de PEG 6000, d'un milieu concentré en sel et de l'exposition aux basses températures (Caruso *et al.* 2002). Le transcrit peudhn1 était induit dans chaque condition. La tolérance au stress n'est pas toujours améliorée par la surexpression d'une LEA. Cependant chez *Arabidopsis*, l'obtention de plants transgéniques surexprimant simultanément 2 Déhydrines (RAB18 et COR47 ou LT129 et LT130) a permis d'obtenir des plants avec une tolérance accrue au gel. Tous les types de LEA ne sont pas encore connus. Récemment, le gène CpEdi-9 codant pour une protéine hydrophile a été découvert chez *C. plantagineum*. La séquence ne donnait aucune homologie nucléotidique ou protéique avec des gènes connus mais la structure de la protéine et certains motifs ont permis de la rapprocher des Déhydrines ou des LEA du groupe 3. Le transcrit de cette nouvelle LEA est fortement exprimé dans les graines matures et dans les cellules des vaisseaux conducteurs du phloème de plants exposés à la déshydratation, au sel ou à l'ABA. Il est également rapidement accumulé dans les feuilles et racines en déshydratation (Rodrigo *et al.* 2004). Les auteurs ont également étudié la partie promotrice du gène. Ils ont trouvé des motifs ABRE, CE1-like, Myb-like, ASCE-like, PB et un élément spécifique du phloème. La construction de gènes rapporteurs GUS chez des plants d'*Arabidopsis* et de tabac leur a permis de confirmer que ce promoteur répondait à la dessiccation, à l'ABA et s'accumulait au niveau des stomates, du système vasculaire et lors de la formation du pollen. Nous verrons plus en détail la régulation de la transcription dans un prochain paragraphe.

B.2.2.5.4 Ajustement osmotique

Nous avons vu que l'ajustement osmotique était un paramètre essentiel de la résistance au stress. L'augmentation du potentiel osmotique peut être obtenue par accumulation d'ions

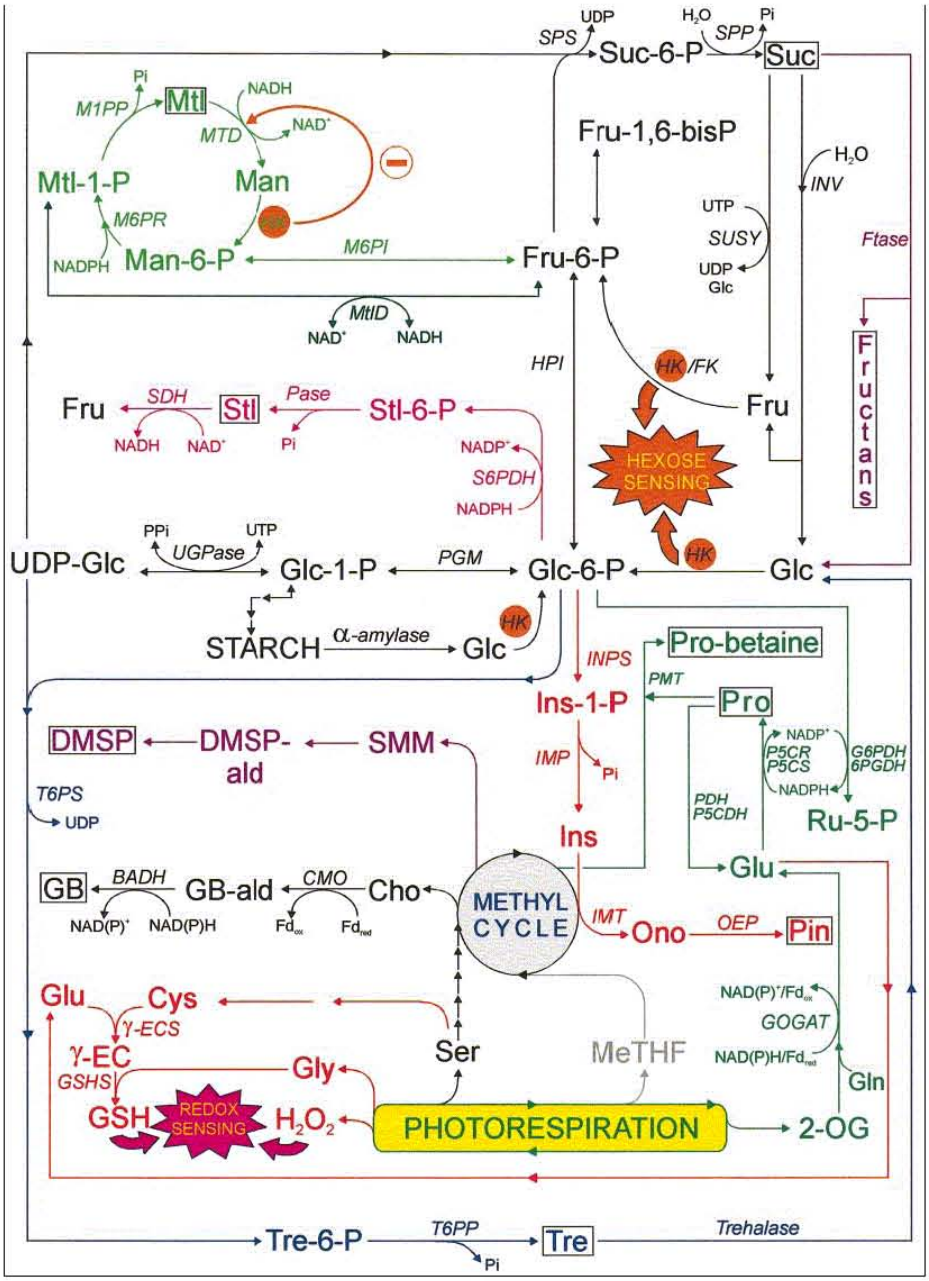


Figure B-28 : Osmolytes et leur métabolisme. Les lignes interrompues indiquent de multiples étapes intermédiaires. (Hare *et al.*, 1998)

Abréviations des composés : Cho, choline; DMSP, 3-dimethylsulphoniopropionate; DMSP-ald, DMSP aldehyde; g-EC, g-glutamylcysteine; Fd, ferredoxin; Fru, fructose; Fru-6-P, fructose-6-phosphate; Fru-1,6-bisP, fructose-1,6-bisphosphate; Glc-1-P, glucose-1-phosphate; Glc-6-P, glucose-6-phosphate; GB, glycine bétaine; GB-ald, glycine bétaine aldehyde; Gln, glutamine; Glu,Glutamate; GSH, reduced glutathione; Ins-1-P, myo-inositol-1-phosphate; Ins, myo-inositol; Man, mannose; Man-6-P, mannose-6-phosphate; Met, methionine; MeTHF, N5-methylenetetrahydrofolate; Mtl-1-P, mannitol-1-phosphate; Ono, ononitol; 2-OG, 2-oxoglutarate; Pin, pinitol; Pro, proline; Ru-5-P, ribulose-5-phosphate; Ser, serine; SMM, S-methylmethionine; Suc, sucrose; Suc-6-P, sucrose-6-phosphate; Stl-1-P, sorbitol-1-phosphate; Tre, trehalose; Tre-6-P, trehalose-6-phosphate; UDPGlc, uridine 5¢-diphosphoglucose.

Abréviations des enzymes (en italique) : BADH, betaine aldehyde dehydrogenase; CMO, choline monooxygenase; g-ECS, g-EC synthetase; FK, fructokinase; Ftase, fructosyltransferase (levansucrase); G6PDH, glucose-6-phosphate dehydrogenase; GOGAT, glutamate synthase; GSHS, GSH synthetase; HK, hexokinase; HPI, hexose phosphate isomerase; IMP, myo-inositol-1-phosphate monophosphatase; IMT, myo-inositol O-methyltransferase; INPS, myo-inositol-1-phosphate synthase; INV, invertase; M6PI, mannose-6-phosphate isomerase; M6PR, mannose-6-phosphate reductase; M1PP, mannitol-1-phosphate phosphatase; mtlD, bacterial mannitol-1-phosphate 5-dehydrogenase; MTD, plant mannitol dehydrogenase; OEP, ononitol epimerase; Pase, phosphatase; PDH, proline dehydrogenase; P5CDH, P5C dehydrogenase; P5CR, P5C reductase; P5CS, P5C synthetase; PMT, proline methyltransferase(s); 6PGDH, 6-phosphogluconate dehydrogenase; PGM, phosphoglucomutase; SPS, sucrosephosphate synthase; SPP, sucrose-6-phosphatase; SDH, sorbitol dehydrogenase; S6PDH, aldose-6-phosphate reductase; SUSY, sucrose synthase; T6PP, trehalose-6-phosphate phosphatase; T6PS, trehalose-6-phosphate synthase; UGPase, UDPGlc pyrophosphorylase.

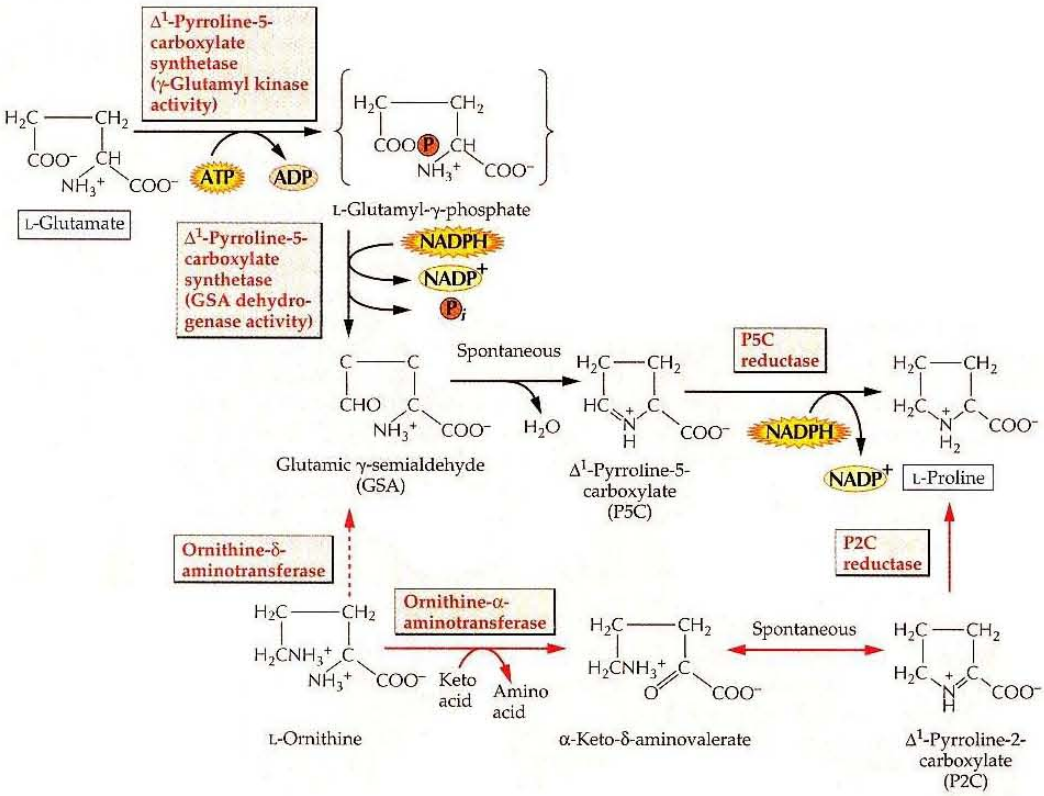


Figure B-29 : Voie de biosynthèse de la proline chez les plantes. La voie dérivée du glutamate est bien établie (flèches noires). En rouge la voie de l'ornithine dont l'importance reste à déterminer. (Bray *et al.*, 2000)

dans la cellule. Cependant, à forte concentration, les charges électriques vont interagir et altérer les propriétés des cofacteurs, substrats, membranes et enzymes. La cellule peut compartimenter ces ions dans la vacuole mais aussi produire des composés organiques qui en plus du rôle d'osmolyte, vont pouvoir offrir une osmoprotection en stabilisant les protéines.
Les principales molécules identifiées sont la proline, la glycine bétaïne et les sucres alcool. Hare *et al.* (1998) ont fait une synthèse sur les osmolytes, leur accumulation et leurs rôles au cours du stress. Il synthétise dans la figure B-28 les principales voies métaboliques et les enzymes impliquées.

B.2.2.5.4.1 Proline

Un des osmolytes le plus répandu est certainement la **proline**. La synthèse de cet acide aminé à partir du **L-Glutamate** est catalysée par un enzyme : le Δ^1-Pyrroline-5-carboxylate synthétase (**P5C synthétase**) dont la fonction est double. Il permet d'abord la conversion en présence d'ATP du L-Glutamate en L-Glutamyl-γ-phosphate qui lui-même est convertit en présence de NADPH en Glutamic γ-semialdehyde (GSA) donnant le **P5C** (Δ^1-Pyrroline-5-carboxylate). Un autre enzyme, le **P5C réductase** permet d'obtenir la proline en présence de NADPH (Fig. B-29). Une autre voie encore mal connue serait possible à partir de l'Ornithine. La proline se retrouve également dans les cellules sous forme d'analogues parmi lesquels le N-methyl-L-proline (MP), le trans-4-hydroxy-Nmethyl-L-proline (MHP), le trans-4-hydroxy-N-dimethyl-L-proline (DHP), ou la stachydrine (proline bétaïne).

L'importance de la proline dans la tolérance au stress a orienté nombre de recherches sur les gènes impliqués dans sa synthèse. Une équipe japonaise a étudié la réponse des deux enzymes cités lors de l'exposition de plants d'*Arabidopsis* à différents traitements (Yoshiba *et al.* 1995). Les résultats de Northern blot montrent que la P5C synthétase est induite par la sécheresse, le stress salin et l'exposition à l'ABA mais pas les stress thermiques. Ils ont pu vérifier que l'expression du gène lors du stress osmotique était accompagnée par une accumulation de proline. Par contre, la P5C réductase n'était pas régulée lors du stress. Dans l'hypothèse que les autres voies de biosynthèse seraient minoritaires, la synthèse de la proline serait donc chez *Arabidopsis* sous le contrôle d'un seul enzyme, la P5C réductase, codé par un gène unique (Savoure *et al.* 1995). Lors de la disparition du stress, la proline fortement accumulée doit être dégradée. Plusieurs voies de dégradations ont été proposées et étudiées. La proline peut être oxydée en glutamate en faisant intervenir la **proline oxydase** ou **proline déshydrogénase** (ProDH). Verbruggen *et al.* (1996) ont isolé un ARNm, At-POX, codant pour cet enzyme chez *Arabidopsis*. Les auteurs ont montré que l'expression de At-POX est

régulée négativement lors d'un stress osmotique et précède l'accumulation de proline. Lors de la levée du stress son expression augmente. Nakashima *et al.* (1998) ont confirmé ces résultats en mesurant une accumulation d'un transcrit ERD5 codant pour un précurseur de ProDH lors de la réhydratation, d'un stress hypo-osmotique ou de l'exposition à de la proline chez *Arabidopsis*. La création d'un système de gène rapporteur GUS avec le promoteur de la ProDH montre une expression dans l'ensemble de la plante avec une coloration plus marquée à l'extrémité des racines (Nakashima *et al.* 1998). L'expression d'un ARN antisens de la ProDH, réalisé sur *Arabidopsis* par Nanjo *et al.* (1999), confirme les résultats précédents en améliorant la tolérance au stress froid et salin des lignées transformées.

Les osmoticums, pour jouer leur rôle physiologique, doivent être compartimentés. La proline synthétisée par la plante fait l'objet d'un transport actif. Ueda *et al.* (2001) ont identifié un ADNc codant pour un transporteur de Proline (HvProT) dont l'expression est fortement induite au niveau des racines mais pas des feuilles lors du stress salin. La spécificité du transporteur pour la proline en présence de compétiteurs tels que d'autres acides aminés, la glycine bétaïne ou la choline, a été vérifiée (Ueda *et al.* 2001). Des ADNc codant pour des transporteurs spécifiques de la proline et induits lors de stress osmotiques ont été identifiés chez *Arabidopsis* (ProT2, ProT1, famille des AAP), la tomate et le riz (Rentsch *et al.* 1996; Igarashi *et al.* 2000).

B.2.2.5.4.2 Glycine bétaïne

En tant qu'osmolyte, la glycine bétaïne ne se retrouve accumulée principalement que chez les plantes hautement tolérantes au sel ou au froid. Elle a pu être détectée chez l'épinard, la betterave à sucre ou encore le maïs. Elle est synthétisée en deux étapes à partir de la choline dans une réaction catalysée par la choline monooxygénase et l'aldéhyde bétaïne déshydrogénase. Une expérience de résistance au stress froid a été réalisée sur le tabac (Parvanova *et al.* 2004). Les auteurs ont utilisé des lignées transgéniques GB9 exprimant le gène bactérien codA qui permet l'accumulation de glycine bétaïne. La mesure de marqueurs du stress tels que l' « electrolyte leakage », la teneur en Malondialdehyde (MDA) –produit final de la peroxydation des lipides-, l'H_2O_2, les catalases et les peroxydases, indique que les plants riches en glycine bétaïne sont plus tolérants au stress froid que le phénotype sauvage.

B.2.2.5.4.3 Sucres et dérivés

Si les sucres sont les éléments carbonés primaires synthétisés et exportés dans toute la plante lors de la photosynthèse, ils ont également un rôle non énergétique d'osmorégulateur et

d'osmoprotectant. L'augmentation de la teneur en sucres ou dérivés, et leur rôle dans la réponse aux stress hydrique a été montré depuis longtemps chez de nombreuses espèces (Gorham *et al.* 1981). L'accumulation de saccharose chez les plantes reviviscentes est particulièrement nette et prendrait une large part dans l'acquisition de la tolérance à la dessiccation. *C. plantagineum* accumule de grandes quantités de 2-octulose qui servirait de source à la conversion en saccharose pour une part supérieure à 80% (Scott 2000). Lors d'une dessiccation sévère une saturation en sucres permet une vitrification du cytoplasme de la cellule et évite une cristallisation des solutés qui causerait des dommages cellulaires irréversibles. Les propriétés physiques obtenues sont proches de celles d'un solide et sont à l'origine de la tolérance à la dessiccation des graines. Koster *et al.* (1988) ont rapporté une corrélation entre la teneur en sucres solubles –stachyose et raffinose- des graines et la capacité germinative du soja, du pois et du maïs.

Le sucre principal intervenant dans la croissance est certainement le saccharose mais d'autres rôles lui sont attribués dans les mécanismes de régulation de l'expression des gènes. Il peut être synthétisé à partir d'UDP-glucose et de Fructose-6-Phosphate ou de Fructose. Dans le premier cas l'enzyme impliqué est la sucrose-phosphate synthase (SPS), dans le second, il s'agit du sucrose synthase (SuSy) mais in vivo cette enzyme catalyse plutôt la réaction inverse de dégradation du saccharose. Dans une synthèse sur le rôle et la régulation de la SPS chez les plantes supérieures, Huber *et al.* (1996) rapportent que cet enzyme est activé par les stress osmotiques chez l'épinard et les tubercules de pomme de terre.

L'osmoprotectant le plus efficace en termes de concentrations est sans doute le **Tréhalose**. Ce disaccharide non réducteur a la particularité d'être particulièrement stable à la température, au pH et aux réactions avec d'autres molécules comme les acides aminés. Présent chez de nombreuses bactéries, champignons et quelques plantes tolérantes à la dessiccation, il n'est pratiquement pas détecté chez les plantes supérieures. Cependant des séquences homologues aux enzymes de sa voie de biosynthèse ont été découvertes chez *Arabidopsis* (Goddijn and Dun 1999). Il s'agit de la tréhalose-6-phosphate synthase (TPS) qui permet de convertir l'UDP-Glucose et le Glucose-6-P en Tréhalose-6-P, et de la tréhalose-6-phosphate phosphatase (TPP) qui convertit le Tréhalose-6-P en Tréhalose (Fig B-30). Des essais de transformation de plants de tabac avec des TPS de levures ou de bactéries ont permis d'obtenir des plants avec une meilleure tolérance à la sécheresse, une capacité photosynthétiques sous stress améliorée, une baisse de la teneur en saccharose et une

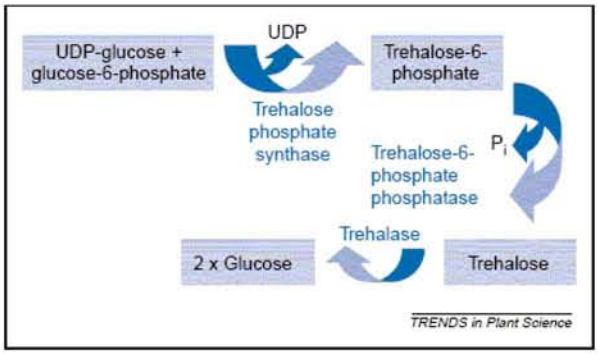


Figure B-30 : Biosynthèse et métabolisme du tréhalose chez les plantes. (Penna *et al.*, 2003)

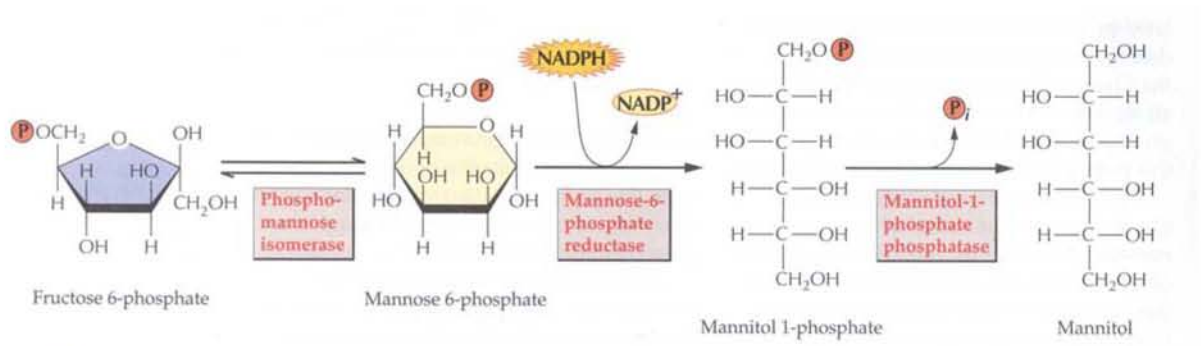


Figure 22.10
The biosynthetic pathway for mannitol.

Figure B-31 : Voie de biosynthèse du Mannitol. (Bray *et al.*, 2000)

croissance retardée. Les quantités de tréhalose accumulées restaient cependant faibles (≤ 0.5 $\mu molg^{-1}$) (Penna 2003).

Le **Mannitol** est la forme réduite du mannose. Ce sucre alcool se retrouve très largement distribué chez les plantes. Son accumulation est souvent mise en évidence lors de stress salins. Le mannitol est synthétisé à partir de Fructose-6-phosphate (Fig. B-31). Ce F6P est convertit en Mannose-6-phosphate par la Mannose-6-phosphate isomérase puis une réaction catalysée par la Mannose-6-phosphate réductase en présence de NADPH permet d'obtenir du Mannitol-1-phosphate qui donne le Mannitol sous l'action de la Mannitol-1-phosphate phosphatase ou Mannitol-1-phosphatase. Chez les bactéries, il existe un gène Mt1D -codant pour une Mannitol-1-Phosphate Déshydrogénase– qui est nécessaire à l'assimilation du Mannitol et du Glucitol par la bactérie (Watanabe *et al.* 2003). L'enzyme produite, catalyse la réaction de conversion entre le D-fructose 6-phosphate et le D-mannitol 1-phosphate en présence de NADH. L'expression combinée de ce transgène Mt1D et d'un autre gène GutD –codant une glucitol-6-phosphate déshydrogénase– chez le pin *taeda* s'est traduite par une surexpression des transcrits correspondants et une accumulation de mannitol et glucitol. Les plants obtenus présentaient une meilleure tolérance au stress salin (Tang *et al.* 2005). Le **Glucitol** est aussi appelé sorbitol, c'est un sucre présent à de très fortes concentrations dans les fruits des *Rosaceae* où il peut représenter 60 à 90% des composés carbonés provenant des feuilles. On le retrouve également impliqués dans la régulation osmotique (Lo Bianco *et al.* 2000). L'expression d'un transgène de la pomme codant pour une sorbitol-6-phosphate déshydrogénase chez le tabac permet de mesurer une accumulation de sorbitol. Les teneurs enregistrées seraient suffisantes pour avoir un effet sur le potentiel osmotique. Mais les concentrations élevées de ce composé chez une plante dont le métabolisme n'y est pas adapté, ont entraîné des perturbations comme un développement racinaire réduit, des nécroses des feuilles ou une croissance réduite. Seul les plants ayant des teneurs faibles en sorbitol ont conservé le phénotype sauvage et l'impact sur la tolérance au stress était difficile à interpréter (Sheveleva *et al.* 1998). Cette expérience met en évidence les difficultés à estimer l'implication d'un composé ou d'un gène dans la réponse à un stress. Les voies de biosynthèse sont interconnectées et l'augmentation de l'activité d'un enzyme, en convertissant de façon accrue un composé va perturber les voies de biosynthèse pour lesquelles ce composé est une source. Dans le cas des sucres cités jusqu'ici, les sources sont des sucres intervenant directement dans la glycolyse qui est la voie directe de production d'énergie dans la cellule. De plus l'accumulation d'osmoprotectant, en stabilisant les protéines peut gêner les réactions biochimiques.

Il existe une autre catégorie d'osmoticums regroupés sous le nom **d'Oligosaccharides de la famille du Raffinose (RFO family)**. Leur rôle essentiel dans le processus de dessiccation des graines a été rapporté chez de nombreuses espèces. La synthèse du **raffinose** et du **stachyose** nécessite du galactinol qui est produit par une galactinol synthase (**GoIS**) à partir d'UDP-galactose et de myo-inositol (Fig. B-32). La raffinose synthase produit le raffinose à partir de galactinol et de saccharose tandis que la stachyose synthase catalyse la synthèse de stachyose à partir de galactinol et de raffinose. La GoIS va donc réguler l'étape initiale de biosynthèse des RFO et son activité ainsi que l'expression des gènes codant la GoIS sont induits lors du stress froid. Taji *et al.* (2002) ont étudié l'accumulation du Galactinol et des RFO chez des plants d'*Arabidopsis* exposés au froid, au sel et à la déshydratation. Les teneurs en galactinol et raffinose mesurées étaient fortement augmentées. Les auteurs ont identifié 7 gènes codant la GoIS chez *Arabidopsis*. Trois des 7 gènes étaient surexprimés pendant le stress. Les gènes AtGoIS1 et 2 répondaient au stress salin et à la sécheresse ; ils possèdent des promoteurs DRE ou DRE-like et ABRE-like et sont légèrement induits par l'ABA. AtGoIS3 répondait au froid ; il possède des promoteurs DRE, DRE-like et ABRE mais n'est pas induit par l'ABA. L'utilisation de lignées transgéniques a permis de confirmer d'une part, que seul AtGoIS3 était induit par des facteurs de transcription DREB1A et d'autre part que la surexpression d'AtGoIS conférait au plant transgénique une nette tolérance à la sécheresse et une transpiration réduite de leurs feuilles (Taji *et al.* 2002). Les auteurs émettent l'hypothèse d'une induction de la production d'ABA par le galactinol et le raffinose et ont mis en évidence le rôle d'osmoprotectant sur les membranes des chloroplastes.

Les *Pinaceae*, les *Fabaceae* et les *Caryophylaceae* accumulent du **Pinitol**. Les concentrations de ce sucre alcool sont plus fortes chez les espèces halophytes ou adaptées à la sécheresse. Chez *Mesembryanthemum crystallinum* exposés à de fortes concentrations en sel, le Pinitol peut constituer plus de 70% des sucres solubles (Nelson *et al.* 1999). Ce constituant majeur des plants de soja est répartit dans la plante selon un gradient de concentration ; les teneurs les plus élevées correspondant aux feuilles des nœuds supérieurs les plus exposées aux rayonnements solaires (Streeter *et al.* 2001). Une étude sur des variétés introduites en chine a mis en évidence que les plants présentant les concentrations les plus fortes correspondent aux variétés sélectionnées pour leur rendement dans les régions les plus sèches (Streeter *et al.* 2001). Ces résultats soulignent l'implication du Pinitol dans la tolérance à la sécheresse chez le soja. Le Pinitol, le **myo-Inositol** et l'**Ononitol** font partie de la famille des cyclitols caractérisés par un cycle à 6 carbones. Les enzymes intervenant dans sa voie de biosynthèse à

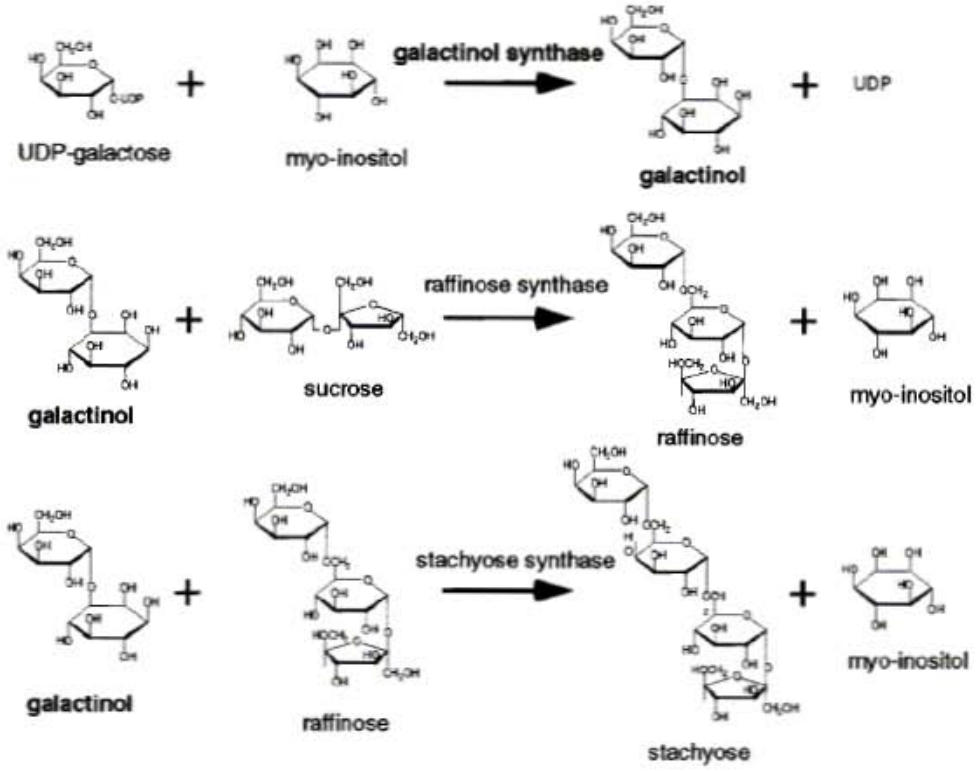


Figure B-32 : Voies de biosynthèse des oligosaccharides de la famille du Raffinose (RFO family) (Taji *et al.*, 2002)

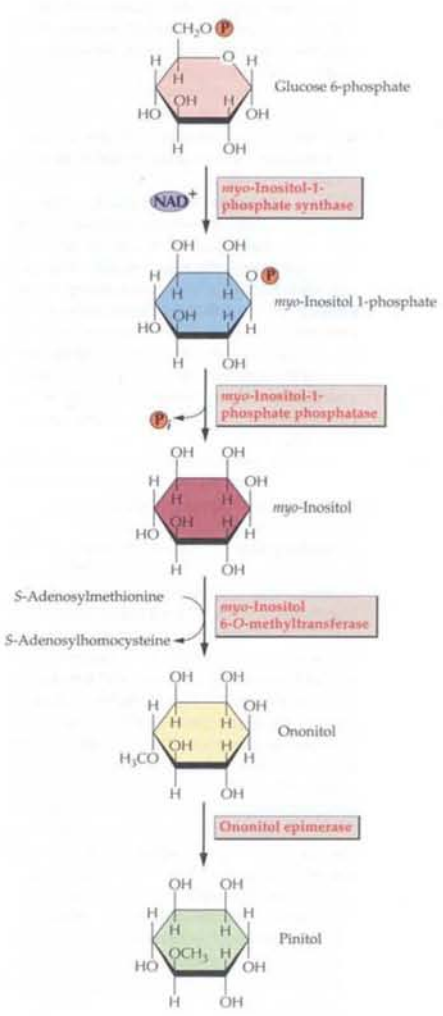


Figure B-33 : Voie de biosynthèse du pinitol. (Bray *et al.*, 2000)

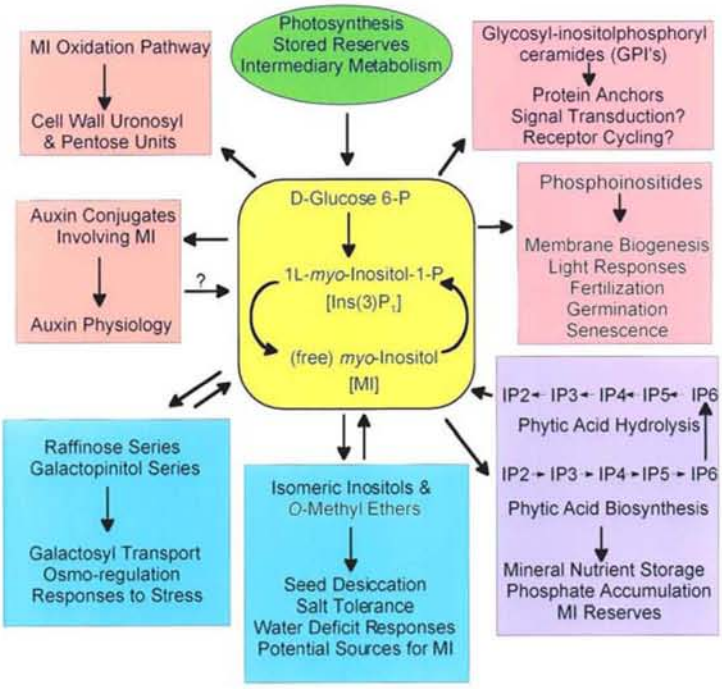


Figure B-34 : Rôle central de l'Inositol et voie de biosynthèses associées. (Loewus *et al.*, 2000)

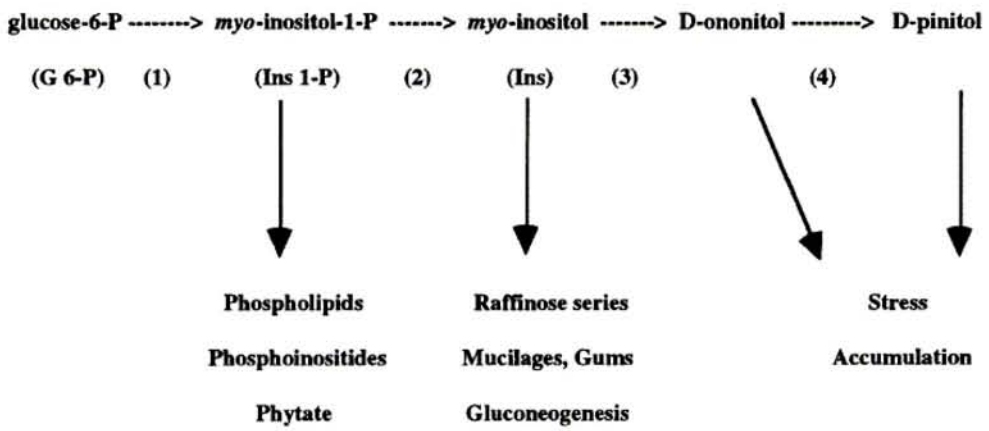


Figure B-35 : Voie de Biosynthèse de l'Inositol en relation avec la biosynthèse des osmolytes. (1) myo-Inositol-1-phosphate synthase (INPS1) ; (2) myo-Inositol-1-phosphate-phosphatase (IMP1) ; (3) myo-Inositol-O-methyltransferase (IMT1) ; (4) Ononitol épimérase (OEP1) (Ishitani *et al.*, 1996)

partir du glucose-6-phosphate sont au nombre de quatre (Fig B-33). La première, la myo-Inositol-1-phosphate synthase (INPS1), convertit le G6P en myo-Inositol 1-phosphate (Ins 1-P) qui sert de précurseur à la synthèse des phospholipides, phosphoinositides et phytates. L'Ins 1-P est déphosphorylé par la myo-Inositol-1-phosphate phosphatase (IMP1) et donne le myo-Inositol (Ins) qui occupe un rôle clé pour de nombreuses voies métaboliques comme la synthèse des dérivés du Raffinose (Fig. B-34). La méthylation du myo-inositol par la myo-inositol-O-methyltransferase (IMT1) aboutit à l'Ononitol dont l'épimérisation par l'Ononitol épimérase (OEP1) donne le Pinitol (Fig.B-35). Plusieurs études ont montré que les transcrits de l'INPS1 et de l'IMT1 étaient fortement induits lors de l'exposition de *Mesembryanthemum crystallinum* au sel (Ishitani *et al.* 1996). Dans cette plante, l'accumulation du myo-inositol, de l'ononitol et du pinitol se fait dans le cytosol et des mécanismes de transport actif d'inositol sont mis en place pour accumuler le composé au niveau des feuilles au détriment des racines (Nelson *et al.* 1999). Le rôle du myo-Inositol dans la plante a été développé de façon très complète dans une revue de Loewus *et al.* (2000). L'auteur souligne que la seule voie de synthèse de novo du myo-Inositol repose sur l'IMP1 et décrit les principales réactions dans lesquelles il est impliqué. La tolérance à la sécheresse et au milieu salin peut être apportée par l'Ononitol. La surexpression de l'IMT1 par des plants de tabac leur permet d'accumuler l'Ononitol et de limiter l'inhibition de l'assimilation photosynthétique lors du stress, les plants obtenus sont plus tolérants au stress (Sheveleva *et al.* 1997). Nous verrons plus loin l'importance du pinitol dans la réponse au stress hydrique chez le pin maritime.

B.2.2.5.5 Transport de l'eau et des solutés

Les flux d'eau dans les cellules vont dépendre de la perméabilité des membranes cellulaires à l'eau mais aussi à des ions et autres molécules dont le gradient de concentration entre les compartiments intra et extra cellulaires influe sur les potentiels osmotiques.

Une famille de protéines intra membranaires, les **MIP** (Major Intrinsic Protein) va jouer un rôle important dans la conductance de ces membranes. Les **aquaporines** (**AQP**) font partie de ces MIP et ont été localisées dans la membrane plasmique (**PIP** /Plasma Intrinsic Protein) et dans la membrane de la vacuole (Tonoplaste)(**TIP** /Tonoplast Intrinsic Protein). Ces protéines de 23 à 29 kDa ont 6 domaines transmembranaires en hélice α et de nombreuses régions conservées. Les AQP laissent passer sélectivement l'eau et les gaz dissous. Le niveau d'expression des gènes codant pour des aquaporines n'est pas similaire quelque soit le tissus. Les racines et les fleurs sont en général les deux organes où l'expression est la plus forte. Chez *Nicotiana glauca*, les ARNm codant pour des MIP s'accumulent dans les racines ou les

tiges et leur expression est faible dans les cellules du mésophile, les feuilles entières ou les cellules de garde (Smart *et al.* 2001). On observe également une régulation du niveau d'expression des gènes codant pour des aquaporines. Une répression forte de l'accumulation des ARNm de NgMIP2, NgMIP3 et NgMIP4 a été observée chez *N. glauca*. L'étude de 5 TIP chez le tournesol a montré que lors d'une déshydratation, la quantité de transcrits dans les racines augmentait pour SunTIP7, diminuait pour SunTIP18, augmentait transitoirement pour SunTIp20 et ne fluctuait pas pour SunRb7 et SunγTIP (Sarda *et al.* 1999). La régulation de l'expression des aquaporines va influer sur les propriétés de perméabilité à l'eau des membranes. Leur fonction a été étudiée plus avant chez le tabac en se servant d'une technique d'ARNm antisens. Siefritz *et al.* (2002) ont comparé le phénotype de plants transgéniques sous exprimant le gène NtAQP1 codant pour une PIP1 avec le phénotype sauvage. Les auteurs ont observé une perméabilité plus forte des membranes des cellules exprimant le gène. Les plants transgéniques présentaient une conductivité hydraulique réduite et une résistance plus basse lors de l'exposition à des stress osmotiques (PEG) violents. Chez *Arabidopsis*, le génome contient 35 gènes codant des aquaporines dont 13 sont homologues à des PIP. Deux études récentes ont porté sur l'étude de l'expression de ces gènes. Il apparaît que l'expression des AQP est dépendante de l'organe. AtPIP1;1 et AtPIP2;2 sont exprimés préférentiellement dans les racines ce qui est le cas de la plupart des PIP étudiées. La plupart des gènes sont régulés négativement dans les feuilles en condition de sécheresse excepté AtPIP1;4, AtPIP2;5 surexprimés et AtPIP2;6, AtSIP1;1 non régulés (Alexandersson *et al.* 2005). Ces auteurs ont pu vérifier que la variation de quantité de transcrits des PIP suivait bien celle des protéines. La réponse à différents types de stress abiotiques –osmotique, froid, haute salinité- et à l'ABA a été testée par Jang *et al.* (2004). Le traitement froid provoque une sous expression de pratiquement toutes les PIP. Le stress osmotique induit des régulations positives ou négatives marquées. Le stress salin est à l'origine d'une légère surexpression concernant pratiquement toutes les PIP. L'application d'ABA induit des réponses variables. Cependant certaines PIP ont des régulations opposées entre feuilles et racines. Les auteurs proposent un schéma des traitements régulant les différentes PIP (Fig. B-36) (Jang *et al.* 2004).

La sélectivité des aquaporines pour l'eau n'est pas absolue. Tyerman *et al.* (2002) décrivent les autres rôles de ces canaux. Les AQP peuvent être perméables au CO_2, à l'ammoniaque, l'ammonium, à l'acide borique et au peroxyde d'hydrogène.

Le contrôle des concentrations en Na^+ et K^+ est indispensable au fonctionnement de la cellule. Des protéines spécialisées vont assurer le transport membranaire de ces ions. Il a été montré que l'expression des gènes codant pour ces canaux membranaires est régulée lors des stress.

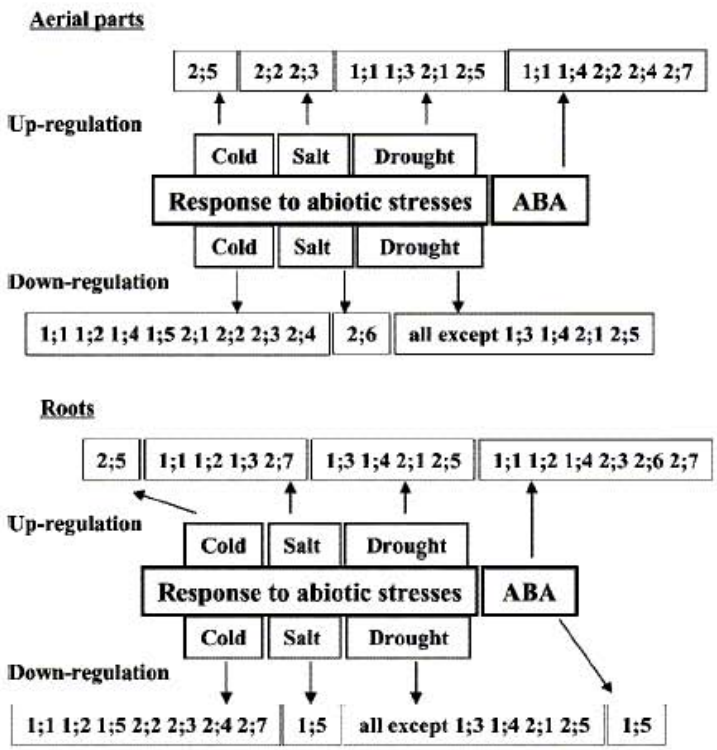


Figure B-36 : Schéma de l'implication des gènes codant des PIP (indiqués par es chiffres) dans les réponses aux stress abiotiques ou à l'ABA. Les gènes indiqués sont régulés positivement ou négativement au moins par un facteur 3 dans les racines ou les parties aériennes chez Arabidopsis lors des traitements indiqués. (Jang *et al.*, 2004)

Shi *et al.* ont montré que le gène SOS1, codant pour un antiporteur membranaire Na+/H+ d'*Arabidopsis*, était essentiel à la tolérance aux fortes salinités (Shi *et al.* 2000). L'expression constitutive de SOS1 chez des plants transgéniques permet une meilleure croissance du système racinaire en condition de stress salin (Shi *et al.* 2003). Ces plants, ont une accumulation de Na^+ réduite car l'efflux de Na^+ est augmenté. L'accumulation de solutés dans les vacuoles des cellules peut être obtenue par la surexpression de gènes codant pour des pompes à proton telles que AVP1. Des plants transgéniques exprimant cette H^+-pyrophosphatase vacuolaire permet aux plants d'acquérir une tolérance à la sécheresse et aux stress salin (Gaxiola *et al.* 2001).

Il existe d'autres type de transporteurs spécialisés dans le transport d'autres composés comme les acides aminés (Chen and Bush 1997; Igarashi *et al.* 2000)ou les monosaccharides (Truernit *et al.* 1996; Weber *et al.* 2000). L'expression des gènes codant ces protéines peut être régulée et tissus dépendante (Kwart *et al.* 1993; Truernit *et al.* 1996). Le produit transporté peut avoir un rôle dans la réponse au stress hydrique. Il a été montré que les transporteurs pour la proline pouvaient être induit par les stress (Rentsch *et al.* 1996; Ueda *et al.* 2001).

B.2.2.5.6 Modification structurale de la cellule

Au cours de la croissance, du développement et de stress biotiques et abiotiques, la structure de la paroi cellulaire est continuellement modifiée sous l'action d'enzymes. Lors d'un stress hydrique, le maintien d'une croissance racinaire accompagné d'un arrêt de croissance des parties aériennes est souvent observé. La croissance nécessite l'expansion et/ou la division cellulaire. La paroi joue un rôle essentiel dans ce mécanisme.

Basiquement, les parois des cellules végétales sont constituées d'un réseau de microfibrilles de cellulose reliées entre elles par des glycans. Mais les parois sont des entités complexes faites d'un amalgame d'eau, de polysaccharides, de protéines, de lignine et de substances telles que cutine, subérine, cires auxquelles s'ajoutent le Bore, le Calcium et des enzymes. Les protéines localisées dans la paroi se distinguent par la richesse en un ou deux acides aminés, la présence de domaines hautement répétitifs dans leur séquence et leur forte ou faible glycosylation. Les cinq classes principales sont les **hydroxyproline-rich glycoproteins** (HRGPs) ou **extensines**, les **glycine-rich proteins** (GRPs), les **proline-rich proteins** (PRPs), les **arabinogalactane proteins** (AGPs) et les lectines (Showalter 1993). La structure, la localisation et la composition de ces protéines a été développée par Cassab (1998). Les gènes codant pour ces protéines appartiennent à de grandes familles multigéniques (Imoto *et al.*

2005). En dehors des protéines citées, on retrouve également toute une série d'enzymes intervenant dans les modifications, la dégradation et la synthèse des parois.

Les **GRPs** contiennent jusqu'à 70% de glycine avec une glycosylation faible à modérée. Elles pourraient se diviser en deux groupes distincts suivant la présence ou non d'un peptide signal qui déterminerait leur localisation membranaire ou non.

Les **PRPs** présentent de grandes homologies avec les extensines et sont identifiées par leurs répétitions Pro-Pro. Leur rôle dans la structure de la paroi reste encore très peu connu mais la présence des GRP et PRP dans les parois (Ryser 2003) et leur régulation dans différentes conditions (Showalter 1993; Akiyama and Pillai 2003) laisse supposer un rôle possible dans la réponse au stress hydrique.

Les **extensines** sont des protéines nettement plus étudiées que l'on retrouve dans différents tissus dont les racines. Leur motif caractéristique est la répétition d'un pentapeptide $Ser(Hyp)_4$ constitué de Sérine galactosylé et d'Hydroxyproline qui porte de 1 à 4 arabinose. Ces extensines sont sécrétées par la cellule au niveau de la paroi où elles se fixent par des liaisons inter et intra moléculaires. Leur expression est régulée par le stress et tissu spécifique. Le gène AtExt1 est ainsi exprimé dans le chevelu racinaire et les zones d'élongation des racines primaires (Merkouropoulos and Shirsat 2003).

Les **AGPs** sont des protéoglycanes, caractérisée par leur très forte glycosylation qui peut atteindre 90% de la masse du produit formé. Elles ne sont pas liées de façon covalente à la paroi et se retrouvent dans des vésicules dérivant du Golgi, dans la membrane plasmique et la paroi. Leur composition suggère des rôles possibles de glue, lubrifiant ou humectant, mais mis à part leur composition commune, aucune fonction certaine ne leur est encore attribuée. Les transcrits codant des arabinogalactanes sont régulièrement mis en évidence dans les réponses à divers stress. Chez le pin *taeda*, l'analyse de l'expression de 11 transcrits d'AGPs a montré que toutes les AGPs étudiées, exceptée ptaAGP5, étaient réprimées dans les racines de plants exposés à la sécheresse (Yang *et al.* 2005).

Harrak *et al.* (1999) ont identifié sur la tomate une Proline- Thréonine- et Glycine-rich Protéine (PTGRP) dont le gène est réprimé lors de la sécheresse et activé lors de la réhydratation. Les auteurs avancent que cette protéine répondant à la sécheresse aurait une localisation dans la paroi.

B.2.2.5.6.1 Modification de la paroi

L'expansion de la cellule nécessite un relâchement de la paroi. Les microfibrilles de cellulose ne sont pas extensibles. La structure de la paroi peut être modifiée par l'action de

complexes enzymatiques. Cela nécessite de couper les liaisons entre les microfibrilles de cellulose et les glycanes puis de rétablir les liaisons après glissement des fibres les unes par rapport aux autres. Les **expansines**, les **Xyloglucan endotransglucosylase/hydrolases** (XTH) et les **Pectines estérases** vont intervenir dans ce processus. Les expansines induisent directement l'extension de la paroi tandis que les XTH, les pectinases, et les glycanases ne le font généralement pas. L'analyse des génomes du riz et de *Arabidopsis* a permis d'identifier respectivement 80 et 38 ORF répartis en 4 classes d'expansines (Li *et al.* 2003). Sur des plants de maïs en condition de stress osmotique, le maintien de l'élongation des cellules de la région apicale des racines primaire et des feuilles est associé à l'augmentation de l'activité et des transcrits des expansines (Wu *et al.* 2001; Sabirzhanova *et al.* 2005). Les XTH vont cliver et reformer les liaisons entre les chaines de xyloglucan qui forment l'hémicellulose majoritaire de la paroi primaire. On dénombre 33 ORF chez *Arabidopsis* et 29 ORF chez le riz (http://labs.plantbio.cornell.edu/XTH/overview.htm). L'appellation de XTH est une nouvelle nomenclature remplaçant les anciennes appellations XET, XEH et EXGT. Au vu de leur rôle supposé, l'activité des XTH serait associée à l'élongation des parois. Si cette hypothèse n'est pas toujours validée, de nombreuses études tendent à la confirmer. L'analyse de l'expression des 33 gènes d'*Arabidopsis* montre que les membres ont des profils d'expression différents selon le tissu et les hormones appliquées (Yokoyama and Nishitani 2001). Chez des plants de soja exposés à un faible potentiel hydrique, l'activité en XET dans la zone d'élongation de l'hypocotyle était associée aux variations d'extensibilité mesurées (Wu *et al.* 2005).

B.2.2.5.6.2 Dégradation de la paroi

Les principales enzymes intervenant dans ce mécanisme sont les **Pectate lyases** -qui vont cliver la pectine pour donner des oligosaccharides-, les **endo-** et **exo-polygalacturonases**, et les **β-galactosidases** auxquelles se rajoutent les **Endo-1,4-β-glucanase** ou **cellulases.** L'expression de ces enzymes est également affectée par le stress hydrique (Bray 2004).

B.2.2.5.6.3 Biosynthèse de la paroi

La synthèse de paroi nécessite l'intervention de nombreux composés. Les protéines et enzymes déjà citées vont intervenir dans la formation d'une nouvelle paroi. Deux enzymes du complexe de synthèse de la cellulose seront indispensables à la synthèse de l'hémicellulose : la **Cellulose synthase** et les **Galactosyl/xylosyl transférases**. L'étude de mutants d'*Arabidopsis* a permis de mettre en évidence qu'un gène de la cellulose synthase jouait un

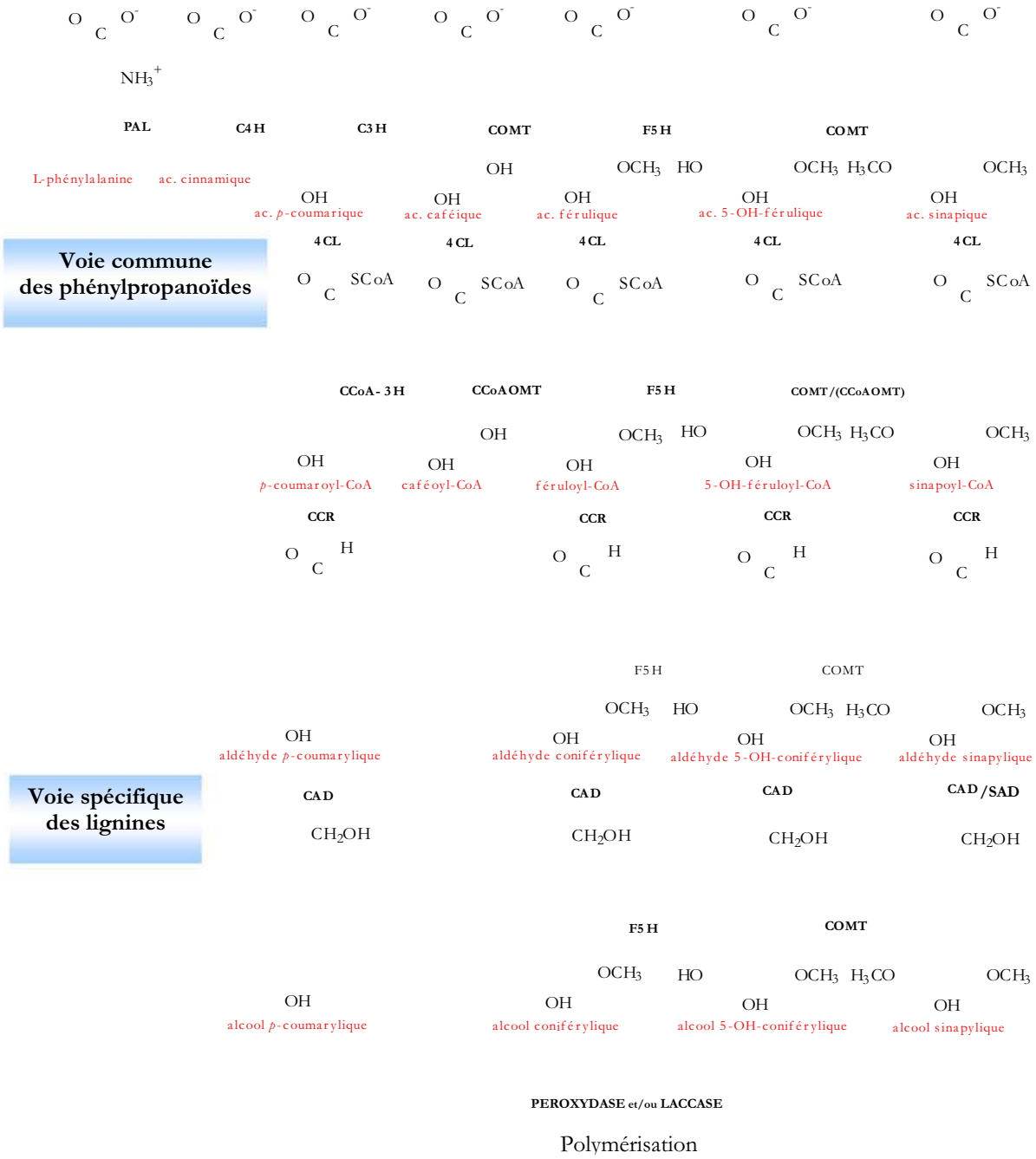


Figure B-37 : Voie de biosynthèse des lignines d'après Rech (2002).
Les enzymes mentionnées sont : la phénylalanine ammonia-lyase (PAL), la cinnamate–4-hydroxylase (C4H), la coumarate-3-hydroxylase (C3H), l'acide caféique/acide 5-hydroxyférulique 3-O-méthyltransférase (COMT), la férulate-5-hydroxylase (F5H), la 4-hydroxycinnamate CoA ligase (4CL), la p-coumaroyl-CoA 3-hydroxylase (CCoA-3H), la caféoyl-CoA 3-O-méthyltransférase (CCoAOMT), la cinnamoyl CoA réductase (CCR), l'alcool cinnamylique déshydrogénase (CAD) et l'alcool sinapylique déshydrogénase (SAD) .

rôle important dans la réponse à la sécheresse. En effet, les mutants allèliques lew2-1 et lew2-2 sont plus résistants aux stress hydriques (sécheresse, sel, osmotique) que le phénotype sauvage. Le gène touché est AtCesA8/IRX1 qui encode pour une sous unité du complexe de synthèse de la cellulose (Chen *et al.* 2005). La lignine est également un composé essentiel des parois des cellules végétales. Ce polymère modifie les propriétés physiques en apportant hydrophobicité et résistance mécanique. La voie de biosynthèse (Fig. B-37) a pour précurseur la L-phénylalanine. Les principaux enzymes sont la PAL, la C4H et la 4CL qui interviennent également dans la synthèse des composés phénoliques, et les C-COMT, CCoAOMT, C3H, F5H, CAD et CCR impliqués plus spécifiquement dans la régulation de la synthèse des polymères de lignine (cf. Fig. B-37 pour les abréviations). Dans une étude sur le maïs, Vincent *et al.* (2005) ont montré, en étudiant en particulier l'accumulation des protéines de **COMT**, que dans la feuille, la zone de lignification se situe immédiatement sous la zone d'élongation. En réponse au stress hydrique, cette zone se décale vers la base de la feuille et le niveau de lignine diminue. Les auteurs avancent que la diminution de la lignification serait une réponse adaptative à la sécheresse pour éviter une lignification des zones d'élongation. Dans sa thèse sur le pin maritime, Costa (1999) avait détecté l'accumulation de CCoAOMT dans les aiguilles lors du stress. C. Dubos (2001) avait détecté lors de sa thèse une diminution de l'expression d'un transcrit homologue à la COMT tandis que les transcrits de la PAL, la 4CL et la CAD étaient induits lors du stress.

B.2.2.5.6.4 Cytosquelette

Les microtubules interviennent entre autre dans la morphologie, la mobilité de la cellule ainsi que dans les transports intracellulaires de molécules. Une étude de Nyporko *et al.* (2003) souligne l'importance des microtubules dans la résistance au froid. L'exposition de cellules apicales de racines de blé à un stress osmotique violent provoque leur désintégration. Il semble qu'ils soient remplacés par des macrotubules de tubuline et des paracristaux de tubuline (Komis *et al.* 2002).

B.2.2.5.7 Régulation de la transcription

La recherche de gènes régulés permet d'améliorer nos connaissances sur les fonctions et voies métaboliques impliquées dans la réponse de la plante à son environnement. Chaque gène ne participe toutefois que pour une part limitée dans le processus de tolérance au stress. Certains gènes vont produire des facteurs de transcription (FT) qui vont activer un ensemble de gène. L'identification de gènes qui codent pour ces facteurs de transcription permet de

découvrir des réseaux de régulation et les gènes associés. Plusieurs familles de FT induits lors de la déshydratation ont déjà été identifiées. Ils sont classés en fonction de leur structure protéique, du type de stress les induisant et des éléments auxquels ils vont se lier sur le promoteur des gènes cibles (Ramanjulu and Bartels 2002).

B.2.2.5.7.1 proteines (b-ZIP)

Les facteurs b-ZIP (Basic leucine zipper) ont en commun un domaine de liaison à l'ADN riche en acides aminés basiques et adjacent à un domaine leucine zipper. La famille des **ABF** (ABRE binding factors) appartient à cette catégorie. Ils sont capables de reconnaître des éléments de régulation spécifique, les **ABRE** (ABA-responsive elements), situé en cis sur le promoteur du gène. Il existe deux motifs pour les éléments ABRE. Les **G/ABRE** partagent une séquence consensus (C/T)ACGTGGC que l'on retrouve dans l'élément Em1a (GGACACGTGGC) du gène Em du blé. L'expression du gène Em découvert au niveau de l'embryon a été étudié sur plusieurs espèces (Marcotte Jr *et al.* 1989; Bostock and Quatrano 1992; Vicient *et al.* 1998). Il est induit par l'ABA, les stress osmotiques, et sert de modèle dans les études de régulation de l'expression (Guan and Scandalios 1998). Les **C/ABRE** ont en commun la séquence CGCGTG. Les voies de régulations faisant intervenir ces deux types d'éléments ne sont pas les mêmes. Une revue détaillée sur les différents membres de la famille des ABF a été présentée par Choi *et al.* (2000). Parmi les facteurs de transcription capables de reconnaître les éléments G/ABRE, on peut citer EmBP-1, TAF-1 et GBF3.

B.2.2.5.7.2 Protéines Myb-like et Myc-like bHLH

Les FT de ces familles présentent une structure à deux hélices séparées par une boucle pour les Myc et un coude pour les Myb. La structure des **Myc** comportent une région hélice-boucle-hélice basique à motif leucine zipper. Le site de reconnaissance est du type CACATG (Abe *et al.* 1997). La séquence à l'origine de cette conformation pour les **Myb** est formée par la répétition imparfaite de trois motifs avec conservation de la position d'un tryptophane tous les 18-19 acides aminés. Il existe un site reconnaissance conservé dont la séquence est (T/C)AACTG (Urao *et al.* 1993). L'expression de ces facteurs est liée à la déshydratation. Chez *Arabidopsis* le gène Atmyb2 est induit par la déshydratation et les transcrits disparaissent avec la réhydratation. L'exposition au stress salin et à l'ABA provoque également l'expression de ce gène présentant une très forte homologie avec les protéines Myb (Urao *et al.* 1993). Il est possible de vérifier l'affinité d'une protéine à se fixer sur une séquence d'acides nucléiques spécifique. Dans le cas de Atmyb2, les auteurs ont utilisé un

vecteur d'expression pour produire la protéine et tester sa spécificité à se lier sur des acides nucléiques contenant des séquences de reconnaissance pour Myb. La découverte de gènes codant pour des FT spécifiques d'un type d'élément est également possible. Sur *Arabidopsis*, Abe *et al.* (1997) ont étudié le gène rd22 induit lors de la déshydratation ou de l'application d'ABA. Leur but était de vérifier qu'une région de 67 pb du promoteur, comportant un site de reconnaissance Myb et un site Myc, suffisait à la régulation du gène. Si un gène codant pour un facteur de transcription Myb était connu (ATMYB2), un gène codant pour une protéine Myc était nécessaire. Les auteurs ont utilisé une banque d'expression d'ADNc qu'ils ont criblé en utilisant la séquence du site de reconnaissance Myc. Ils ont alors pu démontrer par des constructions promoteur-GUS que le gène rd22 mis en évidence dans la réponse au stress hydrique pouvait être régulé à la fois par des FT des familles Myc et Myb.

B.2.2.5.7.3 Protéines AP2(APETALA2)/EREBP

Ce domaine est une caractéristique des facteurs du type **DREB** (Dehydratation-responsive element binding factor) et **CBF** (C-repeat binding factor)**.** Ceux ci vont pouvoir reconnaître et se lier à un motif « C-repeat » (CCGAC) des éléments de régulation **CRT** (C-repeat) ou DRE (Dehydratation-responsive element)(A/GCCGAC). On retrouve la présence de ces éléments dans les promoteurs des gènes COR (cold-regulated) induits par le froid.

Les éléments DRE ont été mis en évidence par Yamaguchi-Shinosaki *et al.* (1994) sur *Arabidopsis*. Deux gènes rd29A et rd29B, très similaires et proches sur le génome, répondaient différemment à la déshydratation, aux basses températures, à la salinité et à l'ABA. Les résultats laissaient supposer la présence de 2 éléments de régulation chez rd29A et un seul chez rd29B (Yamaguchi-Shinozaki and Shinozaki 1994). En étudiant les séquences promotrices et des plants transformés de tabac, les auteurs ont identifié un élément répondant à la déshydratation et contenant le motif TACCGACAT. Des protéines capables de se lier à ces éléments DRE sont connues chez *Arabidopsis*. Qiang *et al.* (1998) ont isolé les transcrits des gènes DREB1A et DREB2A dans une banque d'ADNc. L'expression constitutive de ces gènes chez des plants transgéniques de tabac a induit la transcription du gène rd29 ce qui confirme leur rôle de FT. Les auteurs ont également constaté que la tolérance au froid et à la dessiccation des plants transgéniques était améliorée. Par contre, si les deux gènes contrôlent l'expression de rd29, ils ne sont pas induits par les mêmes stress. DREB1A répond aux basses températures tandis que DREB2A est activé par la déshydratation (Qiang *et al.* 1998). Les FT n'induisent pas forcément la transcription d'un gène. Sur le maïs, Kizis *et al.* (2002) ont isolé des gènes codant pour les facteurs DBF1 et DBF2 du type AP2/EREBP. Contrairement à

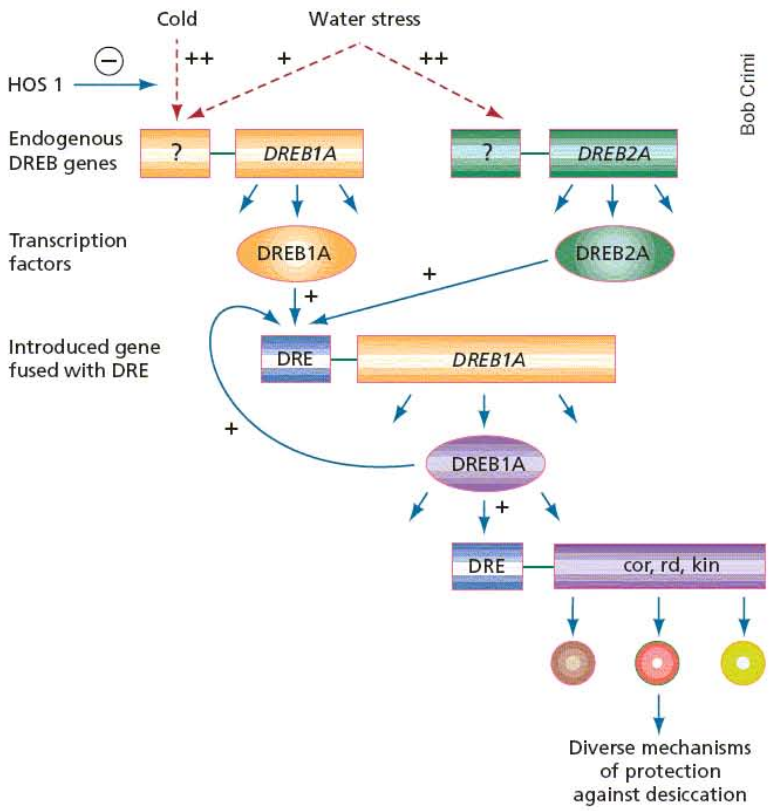


Figure B-38 : La surexpression d'un FT améliore la tolérance à la déshydratation, au froid et au sel chez Arabidopsis. La méthode présentée permet une surexpression d'un FT codant DREB1A en utilisant une construction où le promoteur du transgène est induit par le produit du transgène. (Smirnoff *et al.*, 1999)

DBF1, DBF2 n'est pas induit lors des stress couramment testés et son niveau de transcrits reste très faible. La surexpression de DBF2 a même inhibé l'activité du promoteur du gène rab17 porteur des éléments DRE2. Les auteurs avancent l'hypothèse d'une action de DBF2 sur une autre partie du promoteur (Kizis and Pages 2002). L'expression des FT dépend des conditions de stress appliquées à la plante mais aussi du tissu étudié. Ainsi lors de stress salins, DREB1D et DREB1F sont induit dans les racines, mais DREB2C, DREB2D et DREB2E ne sont induits que dans les feuilles. L'induction de l'expression de DREB2E n'est détectable que dans les racines de plants exposés à l'ABA (Sakuma *et al.* 2002).

En connaissant les FT et les séquences promotrices, il est possible de développer une nouvelle approche de construction de plants transgéniques. Si les mécanismes d'activation des gènes codant pour les FT ne sont pas connus, il est possible d'amplifier le pouvoir de régulation d'un facteur en construisant un transgène constitué du FT précédé d'un promoteur activé par le même facteur. Une telle boucle d'auto amplification a été proposée sur le modèle des DREB par Smirnoff *et al.* (1999) (Fig. B-38).

B.2.2.5.7.4 Protéines WRKY

Cette famille de gènes est l'un des groupes majeurs de régulateurs de transcription spécifiques des plantes. Elles se distinguent par la présence d'un domaine d'un 60^{aine} d'acides aminés hautement conservés parmi lesquels la séquence WRKYGQK suivie du motif « zinc binding » CCHH ou CCHC. Les éléments sur lesquels ces facteurs vont se fixer sont du type « W-box » dont la séquence est (T)TGCA(C/T) (Marè *et al.* 2004). Ces facteurs jouent un rôle important dans les réponses aux pathogènes, aux blessures, lors de la sénescence ou la germination, mais également dans d'autres stress (Seki *et al.* 2002). Dans le génome d'*Arabidopsis*, 72 gènes codant pour des protéines de la famille des **WRKY** ont été découverts par Dong *et al.* (Dong *et al.* 2003) et 49 d'entre eux sont régulés lors d'une attaque de pathogène. Les recherches sur ces facteurs n'en sont qu'à leur début. La fondation nationale pour la recherche (NSF) supporte un nouveau projet coordonné par Zhixiang Chen de l'Université de Purdue (http://www.btny.purdue.edu/WRKY/index.html). Ce projet a pour but de permettre : premièrement l'analyse de l'expression de ces gènes au cours du développement et de la réponse de la plante aux stress biotiques et abiotiques, deuxièmement de déterminer le rôle de ces gènes en observant le phénotype de mutants ne les exprimant pas ou les exprimant de façon constitutive, et troisièmement en analysant les interactions protéiques réalisées par ces WRKY.

B.2.2.5.7.5 Autres Protéines

D'autres protéines à fonction de FT impliquées dans les réponses aux stress ont été décrites dans la littérature.

Les facteurs **HD-ZIP** n'ont été identifiés que chez les plantes. Il s'agit de protéines avec un homéodomaine de fixation à l'ADN associé à un motif leucine zipper. Ils sont induits par la déshydratation mais réagissent différemment à l'ABA. Quelques représentants ont été identifiés sur *C. plantagineum* (CPHB-1 et -2) et d'autres sur *Arabidopsis* (ATHB-7, -6 et -12). Les gènes régulés par ce type de facteur ne sont pas encore identifiés.

Nous avons vu le rôle de protéines chaperonnes des Hsp. La séquence promotrice de ces protéines induites par des stress thermiques contient des éléments **HSE** sur lesquels peuvent se lier des facteurs d'activation en trans **HSF** (Heat Shock Factors). Les éléments HSE ont en commun une séquence consensus « nGAAnnTTCnnGAAn ». Chez *Arabidopsis*, Prändl *et al.* (1998) ont étudié le HSF3 et le HSF4 ayant chacun les caractéristiques de cette famille (régions DBD, HR-A et HR-B) mais dont le reste de la séquence ne présentait pas d'homologie. Le niveau de transcrit du HSF3 n'est pas affecté par les chocs thermiques tandis que le HSF4 est induit lors d'un stress de 2h à 37°C. Les auteurs ont testé si l'expression de ces HSF suffisait à induire des HSP de classe I. Seule la surexpression du HSF3 dans des plants transgéniques a permis d'induire ce type de HSP (Prandl *et al.* 1998). Il existe donc différents HSF répondant à des stress différents et activant des classes de HSP distinctes.

B.2.2.5.7.6 Histones et Méthylations

Jusqu'ici, nous avons parlé de régulations des gènes par des facteurs de transcription mais la régulation est possible en amont, par l'accessibilité au gène dans le noyau.

Le noyau est constitué par de la chromatine formée d'ADN et de protéines associées. Cette chromatine adopte plusieurs états selon le niveau de compaction de la molécule d'ADN. A ces états correspond un niveau d'accessibilité. L'ADN super enroulé d'une chromatine condensée ne pourra pas être transcrit. L'enroulement se fait en premier lieu autour du nucléosome qui est un octamère d'Histones nucléosomiques H4, H3, H2A et H2B ou « core Histones ». Une Histone H1 extranucléosomique permet la superspiralisation de l'enroulement initial (Maillet 1995). Il existe également des variants de ces histones (Huh *et al.* 1995) et d'autres protéines appelées « linker Histones » pouvant agir sur la structure de la chromatine et souvent induits lors de stress hydriques. L'expression d'un variant his1-s d'histone H1 a ainsi été étudié chez la tomate. Ce gène est induit précocement dans les feuilles lors d'un stress hydrique ou de l'application d'ABA. L'accumulation de la protéine de 31 kDa est également observée avec

une localisation nucléaire et dans la chromatine. L'intensité du stress appliqué n'a pas d'incidence sur le niveau d'expression enregistré (Scippa *et al.* 2000). Au vu des résultats et des connaissances sur les histones, les auteurs discutent d'un rôle structural possible de his1-s dans la protection de l'ADN lors du déficit hydrique ou un rôle de régulation des gènes. Ascenzi *et al.* (1997) ont isolé une « linked-Histone » His1-3 de 19 kDa qui diffère des histones connues His1-1 et His1-2 et présente des homologies avec une autre protéine H1-D. L'His1-3 est surexprimée spécifiquement lors du stress hydrique dans les feuilles et les parties aériennes d'*Arabidopsis* contrairement aux His-1,-2 et His4. Les auteurs n'ont par contre observé aucun changement du niveau de transcrits par les basses températures, les hautes températures ou les fortes concentrations en sel. Ce gène semble répondre spécifiquement au stress hydrique.

Les propriétés physico-chimiques des histones peuvent être changées par des modifications post traductionnelles. Les deux types de modifications les plus connues sont la méthylation et l'acétylation des queues N-terminales des H3 et H4, et N- et C-terminales des H2A et H2B. Les **HMT** (histones methyltransférases) et les **PRMT** (protein arginine methyl-transferases) sont des enzymes capables de méthyler les H3 et H4. L'acétylation est effectuée par des protéines de la famille des **HATs** (histones acetyl transferases), tandis que la réaction inverse est catalysée par les **HDAC** (histones déacétylases). Cet aspect de la régulation des histones par modification post-traductionnelle est reprise dans une revue de Peter Loidl (2004).

Les mécanismes de régulation par méthylation concernent également l'ADN. Chez *Arabidopsis*, l'action de la cytosine méthyltransférase **CMT3** et des autres cytosine méthylases joue un rôle essentiel dans l'inhibition de l'expression de régions chromosomiques (Finnegan *et al.* 1996; Finnegan and Kovac 2000).

B.2.2.5.7.7 Autres mécanismes de régulation

L'expression d'un gène et la production de la protéine sont déconnectés. Différentes isoformes peuvent être produites à partir d'un gène par épissage alternatif des exons. La quantité de protéines produites va dépendre du niveau de traduction des ARN messagers et de leur cycle de vie. L'activité de la protéine va ensuite dépendre des modifications post-traductionnelle des protéines qui peuvent être glycosylées ou phosphorylées.

Depuis peu de temps, les études sur les **microARN** suivent un engouement croissant. Les microARN (miRNA) sont de courtes séquences nucléotidiques capables comme les siRNA (Small interfering RNAs) de réguler négativement les ARNm cibles. Les miRNA sont codés par des gènes distincts du gène codant l'ARNm cible et sont obtenus à partir de précurseurs

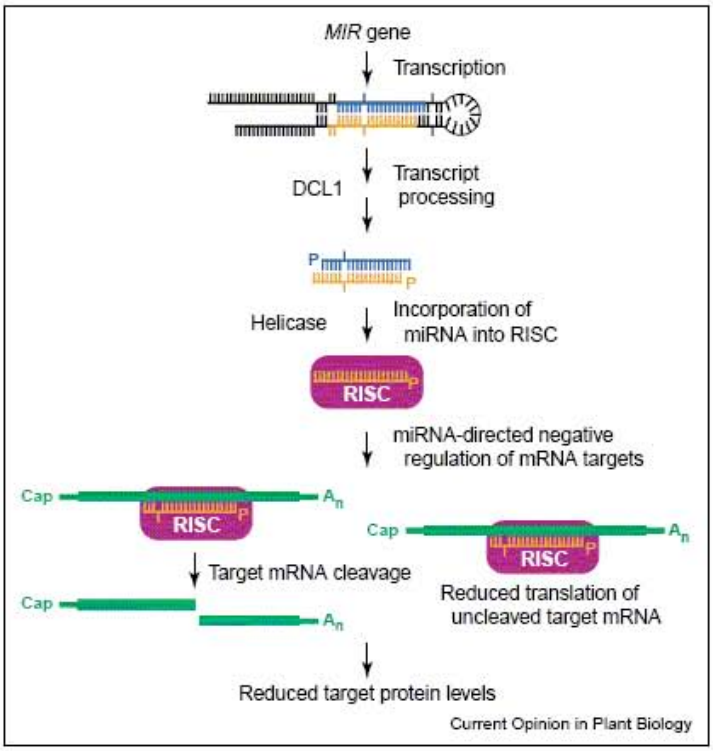


Figure B-39 : Représentation schématique de la formation des microARNs et leur fonction chez les plantes. (Dugas *et al.*, 2004)

dont la séquence forme une boucle en épingle à cheveux imparfaite (imperfectly paired hairpin precursors) (Fig. B-39). Sous l'action d'une Dicer ribonucléase, un fragment d'ARN double brin (dsRNA) est généré à partir du précurseur. Un des brins du miRNA mature est incorporé dans un complexe d'inhibition RISC (RNA-induced silencing complex). Lorsque ce complexe se fixe sur l'ARN messager cible, ce dernier est clivé, ce qui aboutit à une régulation négative de la traduction (Dugas and Bartel 2004; Kidner and Martienssen 2005). Dugas *et al.* (2004) rapportent une liste de miRNA identifiés ainsi que leur cible. Il apparaît que les cibles sont souvent des facteurs de transcription connus.

B.2.2.5.8 Perception et transduction du signal

Avant l'activation de gènes par des facteurs de transcription, les voies de transduction du signal commencent par la perception du signal suivi par la génération de seconds messagers comme l'Inositol phosphate et les ROS (reactive oxygen species). Classiquement, ces seconds messagers vont entraîner une variation de la teneur en Ca++ qui à son tour déclenche des cascades de phosphorylation jusqu'à l'activation finale de facteurs de transcription. Les gènes régulés par ces facteurs peuvent être impliqués dans la biosynthèse de molécules comme l'ABA qui agira comme signal sur de nouvelles voies de transduction. Plusieurs cascades de transduction du signal peuvent donc se succéder ou coexister en réponse à l'évènement initial –par exemple la baisse de disponibilité en eau-. Les études récentes ont montré qu'entre le stimulus initial et la réponse physiologique, la transduction du signal tient plus d'un réseau complexe d'activation et de répression que d'une cascade linéaire (Knight and Knight 2001). Un stress spécifique va donc pouvoir activer des cascades de transduction communes à différents stress et d'autres spécifiques, et beaucoup de gènes pourront être induit par plus d'un stimulus particulier. La revue de Xiong *et al.* (2002) reprend l'état des connaissances sur la perception et la transduction du signal au cours des stress froid, sécheresse et salin. Ces trois types de stress vont engendrer une déshydratation nécessitant des réponses communes mais ils vont aussi engendrer des perturbations plus spécifiques. On peut donc s'attendre à ce que la perception du stimulus initial se fasse par un réseau de senseurs.

B.2.2.5.8.1 Perception du signal

Les récepteurs primaires du stress restent difficiles à identifier. Les différents aspects du stress que la cellule serait susceptible de reconnaître par l'intermédiaire de capteurs regroupent la diminution du potentiel hydrique, la diminution de la pression de turgescence, la concentration de petites molécules, le changement du volume cellulaire ou des changements

de conformation de macromolécules (Bray 2002). Dans le cas des stress osmotiques, le gène **AtHK1** de *Arabidopsis* semble être le composant initial permettant de relier un changement de l'osmolarité à la chaîne de transduction du signal pour la régulation des gènes inductibles par la sécheresse (Urao *et al.* 1999). Ce gène est homologue au gène SLN1 de la levure codant pour une **histidine kinase** transmembranaire qui joue le rôle d'osmosenseur. La voie de perception et de transduction du signal pour l'osmolarité est connue pour cet organisme. L'osmosenseur permet d'activer la cascade des MAPK par l'intermédiaire de deux molécules phospho-relais YPD1 et SSK1 (Boudsocq and Lauriere 2005).

Des récepteurs aux hormones végétales ont également été mis en évidence. Nous avons déjà décrit la cascade de transduction suivant la fixation de l'éthylène sur son récepteur ETR1. Il existe également les récepteurs ERS1, ETR2, EIN4 et ERS2. Cette cascade de transduction a été décrite récemment par Chang *et al.* (2001). Inoue *et al.* (2001) ont découvert un récepteur histidine kinase CRE1 pour la cytokinine. Les senseurs primaires déclenchent souvent un influx massif de Ca^{++} dans la cellule. Les **canaux ioniques** membranaires pourraient aussi faire office de récepteur primaire. Il a été montré que l'ABA pouvait réguler l'activité de canaux potassiques et le potentiel de membrane de cellules racinaires (Roberts and Snowman 2000). La fermeture des stomates ferait intervenir la fixation de l'ABA sur des complexes **protéine-G** (GPCRs : G-protein coupling receptor) contenant des GTPases (Jones 2002).

B.2.2.5.8.2 Calcium et second messagers

A la suite de la perception initiale d'un des facteurs du stress par le récepteur, des messagers secondaires vont être les médiateurs de l'information. Ils vont transporter le signal du lieu de perception au lieu d'activation de la cascade de phosphorylation. Les principaux seconds messagers sont le Calcium, les diacylglycerols (DAG), l'inositol-1,4,5-trisphosphate (IP_3), les ROS et l'ABA (Xiong *et al.* 2002). L'inositol et les diacylglycerols sont les produits du clivage de la phosphatidylinositol bisphosphate par la phospholipase C. Au cours du stress hydrique, les plantes comme *Arabidopsis* peuvent augmenter la production d'IP_3 en induisant l'expression du gène PI5K qui code une phosphatidylinositol 4-phosphate 5-kinase catalysant la phosphorylation d'IPP (Mikami *et al.* 1998). Hirayama *et al.* (1995) ont montré que l'expression d'un gène codant une phospholipase-C était induite par la déshydratation et le stress salin. Les changements de concentration en Ca^{++} vont permettre l'intégration des différentes voies de transduction.

B.2.2.5.8.3 Cascade des MAPKs

Les cascades faisant intervenir les « mitogen-activated protein kinases » (MAPK)s sont probablement le mécanisme de transduction majeur chez les plantes. Une MAPK kinase kinase (**MAPKKK**) phosphoryle une MAPK kinase (**MAPKK**) qui à son tour phosphoryle une **MAPK**. Plusieurs gènes codant pour les MAPKKK (AtMEKK1-4), pour les MAPKK (AtMKK) et pour les MAPK (AtMPK1-6) (Mizoguchi *et al.* 1993) ont été isolés chez *Arabidopsis* (Ichimura *et al.* 1998). Asai *et al.* (2002) ont réussi à mettre en évidence une cascade complète de MAPKs intervenant dans la réponse de *Arabidopsis* aux pathogènes depuis les récepteurs LRR aux éliciteurs jusqu'à l'expression de facteurs de transcription. Les auteurs ont utilisé des mutants présentant des déficiences de cette voie de transduction. L'expression de ces MAPKs est régulée par les stress abiotiques (Mizoguchi *et al.* 1996).

B.2.2.5.8.4 Protéines régulées par le Calcium

Cette catégories regroupe la classe des calmodulines, des « calcium-dependent protein kinases (CDPKs) et des phosphatases régulées par le calcium. Il existe plusieurs isoformes pouvant avoir des spécificités pour un stress comme les AtCDPK1 et AtCDPK2 impliquées dans les stress salins et la sécheresse chez *Arabidopsis* (Urao *et al.* 1994). Chez le riz, les transcrits du gène OsCDPK7 qui code pour une CDPK sont accumulés dans les tissus vasculaires des racines lors du stress froid, salin et de la sécheresse. L'expression constitutive chez des plants transgéniques aboutit à une accumulation de transcrits dans les mêmes régions (Saijo *et al.* 2001).

B.3 Le stress hydrique chez les arbres

Les arbres, composante centrale de l'écosystème terrestre, sont des organismes ligneux longévifs pouvant atteindre plusieurs dizaines de mètres.

B.3.1 Une problématique spécifique

B.3.1.1 Particularités physiologiques

Ces organismes sont caractérisés par une croissance secondaire exacerbée au niveau du cambium où les cellules se différentient pour donner le phloème et le xylème secondaires. Ce dernier est lignifié et forme le bois de cœur et l'aubier. Du fait de leur grande taille et des temps de génération long, les arbres sont difficiles à étudier avec les approches modernes utilisées en génétique. Ils posent des questions de recherche spécifiques et leurs

caractéristiques différentes des autres plantes supérieures suggèrent qu'ils apportent des réponses physiologiques différentes. La hauteur de l'arbre est un des premiers facteurs influant sur les mécanismes physiologiques mis en œuvre. La sève doit parvenir à des hauteurs dépassant les 40 mètres et pouvant atteindre 110 mètres chez le Séquoia. A ces hauteurs, les organes foliaires subissent un stress hydrique et doivent répondre en abaissant leur potentiel osmotique et en fermant leurs stomates. Les valeurs de $\delta^{13}C$ et de potentiel atteignent chez cet arbre les valeurs les plus extrême mesurées chez les plantes (Koch *et al.* 2004). Nous avions expliqué (cf. B.2.1.1) que le potentiel hydrique Ψ_w dépendait du bilan de la pression osmotique Ψ_s et de la pression mécanique Ψ_p exercée par les parois cellulaires qui tend à faire sortir l'eau des cellules. D'un point de vue global de l'arbre par rapport au sol, la force de gravité s'exerçant sur une colonne d'eau de plusieurs dizaines de mètres ne peut plus être négligée et l'équation simplifiée du potentiel hydrique doit tenir compte de ce paramètre Ψ_g (Kramer 1983).

$$\psi_W = \psi_S + \psi_P + \psi_g$$

La taille importante du réseau vasculaire chez les arbres associé aux forces de tensions exercées sur la colonne d'eau et sur les parois des vaisseaux conducteurs souligne la nécessité d'une bonne régulation des pertes en eau et de réponses adéquates pour répondre au stress hydrique et éviter la cavitation (Cruiziat and Tyree 1990).

Une autre particularité est l'importance des mycorhizes chez les arbres qui forment une association symbiotique entre le mycélium d'un champignon et les racines de la plante. Elles vont jouer le rôle des poils absorbants et permettre à l'arbre d'accéder à l'eau présente dans le volume de sol occupé par le mycélium (Heller *et al.* 1993). Cette symbiose qui joue un rôle majeur dans les mécanismes de réponse de l'arbre à la sécheresse a été encore peu étudiée (Nouaim and Chaussod 1996; Garbaye 2000).

B.3.1.2 Ressources moléculaires

Le développement de recherches sur ces organismes est complémentaire des approches réalisées sur les plantes modèles et de grande culture. Le peuplier est venu rejoindre la liste des plantes modèles déjà citées. Les ressources génétiques disponibles se sont grandement enrichies avec le séquençage du génome du peuplier (Brunner *et al.* 2004) (http://genome.jgi-psf.org/Poptr1/Poptr1.home.html). Pour les résineux, si aucune espèce n'a le statut de plante modèle, les ressources génétiques se sont largement développées pour les espèces d'intérêt économique du genre Pinus comme le pin loblolly (*Pinus taeda*), le pin de Monterey / pin à encens (*Pinus radiata*) et le pin maritime (*Pinus pinaster*) ou l'épicéa (*Picea abies*).

Les projets de séquençage d'EST ont permis de construire une base de données conséquente de séquences de gènes exprimés. Les données publiées ont été regroupées par le TIGR qui donne accès aux séquences ainsi qu'à leur assemblage bioinformatique et leur annotation à l'adresse http://www.tigr.org/tdb/tgi/plant.shtml. Des bases dédiées sont également accessibles pour le peuplier http://poppel.fysbot.umu.se/, le pin *taeda* http://www.ccgb.umn.edu/biodata/ et le pin maritime http://cbi.labri.fr/outils/SAM2/COMPLETE/index.php.

La bibliographie sur l'étude de la sécheresse chez le peuplier est riche de références. La problématique a été abordée du point de vue physiologique (Chen *et al.* 2002; Johnson *et al.* 2002; Hukin *et al.* 2005; Marron *et al.* 2005; Monclus *et al.* 2005) comme moléculaire (Caruso *et al.* 2002; el-Khatib *et al.* 2004; Marjanovic *et al.* 2005). Cette espèce présente de nombreux avantages par rapport aux autres ligneux. Outre le fait que nous disposions de la séquence de son génome et de séquences d'EST, cette espèce a une croissance très rapide (certaines espèces exploitables à 15 ans) comparativement au pin (35-40 ans) ou au chêne (100-140 ans). L'obtention de plants transgéniques est également une technique bien maîtrisée sur le peuplier et les approches de biologie moléculaire adoptées chez *Arabidopsis* sont transférables à cet arbre (Strauss *et al.* 2004). Cette technique est aussi réalisable chez le pin maritime ou chez le pin *taeda* (Tang *et al.* 2005). Chez les conifères nous ne disposons pas de génome entièrement séquencé mais de banques d'EST et des cartes génétiques saturées permettant la comparaison des génomes de nombreuses espèces du genre Pinus (ex : (Chagne 2004)). Les grandes thématiques de recherche chez les résineux et les arbres en général se rapportent à la formation et la qualité du bois, la résistance aux pathogènes et la réponse au stress hydrique. Différents projets de séquençage d'EST se sont succédés sur le pin *taeda*. Un premier projet financé par la NSF avait pour objectif d'identifier des gènes impliqués dans la formation du bois (http://pinetree.ccgb.umn.edu/). Un nouveau projet en cours a pour but d'identifier des allèles pour des gènes candidats d'intérêt économique (http://dendrome.ucdavis.edu/adept/index.html). Il existe également un projet axé sur les réponses du transcriptome racinaire aux conditions environnementales (http://fungen.botany.uga.edu/Projects/Pine/PineInfo.htm) et un projet de la NSF sur les embryons somatiques et zygotiques (http://www.tigr.org/tdb/e2k1/pine/index.shtml).

Ces projets ont permis de démarrer des études du transcriptome en utilisant des microarrays sur la réponse au stress hydrique (Heath *et al.* 2002). Dans la bibliographie, les recherches sur la réponse au stress hydrique chez les conifères couvrent la plupart des paramètres physiologiques et catégories de gènes exposés précédemment. Sur le plan physiologique, on pourra citer par exemple une étude sur la comparaison de la résistance au stress osmotique de

huit provenances de *Pinus halepensis* (Calamassi *et al.* 2001) et la thèse de Froux (2002) sur les caractéristiques hydrauliques, la régulation stomatique et l'efficience d'utilisation de l'eau de quatre espèces.

Sur le plan moléculaire, Sathyan (2004) a combiné les techniques de PCR quantitative et d'hybridation par Northern blot pour identifier des gènes différentiellement exprimés lors de la sécheresse entre deux populations de pin d'Alep et entre deux populations de pin *taeda* (Sathyan 2004). L'auteur a notamment mis en évidence la régulation de gènes codant pour des facteurs MYB, des LEA ou la cyclophiline.

B.3.2 Le stress hydrique chez le pin maritime

Si le programme d'amélioration du pin maritime a débuté à l'INRA dans les années 60 sur des critères de rendement et de rectitude, l'étude de la réponse aux contraintes hydriques a débuté plus récemment. Différentes études, que nous avons résumées dans ce chapitre, ont permis d'aborder cette question sous différents aspects : physiologique/écophysiologique, génétique et moléculaire.

B.3.2.1 Les approches physiologiques

Les travaux précurseurs de Guyon (1980 et 1982) ont confirmé la présence d'une variabilité des individus au sein de cette espèce. Les données de mesures morphologiques sur 18 provenances ont permis de distinguer les groupes : Atlantique, Corse, Maroc-Espagne et Tunisie. Les mesures de pression de sève, de transpiration et de teneur en eau des provenances sur des stations différentes indiquent que la variabilité concerne également le comportement hydrique. Les variétés les plus opposées appartiennent aux groupes atlantiques, tunisiens et marocains qui diffèrent sur leurs stratégies de régulation de la transpiration et leur potentiel de base en condition de sécheresse (Guyon 1980; 1982). La sélection naturelle a favorisé les individus les plus adaptés aux conditions climatiques de leur région d'origine en privilégiant la limitation des pertes en eau au détriment de la croissance pour les régions les plus xériques et les meilleurs compétiteurs (les plus efficaces pour l'utilisation de l'eau) dans les zones humides (Alia *et al.* 1993).

La compréhension du comportement écophysiologique du pin maritime landais vis-à-vis de la disponibilité en eau a été améliorée par le travail de El Hadj Moussa (1989) dans le cadre de sa thèse. L'estimation des apports externes, de l'interception par le couvert, de la transpiration de la strate arborescente et celle du sous-bois, et l'évolution du stock hydrique du sol, ont permis d'établir un bilan hydrique à l'échelle de la parcelle et de l'arbre. Les résultats

montrent une régulation de la conductivité hydraulique et de la conductance stomatique avec une modulation de la transpiration suivant les disponibilités en eau. L'auteur indique que le comportement du pin des landes est du type « évitement ».

Les travaux se sont ensuite répartis entre l'aspect régulation stomatique et l'aspect régulation osmotique.

L'ouverture des stomates va dépendre de nombreux facteurs et aura un impact sur l'assimilation photosynthétique du carbone. La sensibilité stomatique à la sécheresse est très élevée chez le pin maritime comparée à d'autres espèces forestières comme le chêne (Guehl *et al.* 1994). En condition de déficit hydrique, la conductance stomatique chute et l'efficience d'utilisation de l'eau augmente (Picon *et al.* 1996). Cependant la conductance stomatique va également dépendre de la conductance du système conducteur des branches et des tiges qui sera fonction de la longueur de ces organes (Warren and Adams 2000). La conductance stomatique diminue avec l'accroissement de la taille ou de l'âge de l'arbre (Delzon *et al.* 2004). La mesure de la discrimination isotopique du ^{13}C ($\delta^{13}C$) dans les tissus reflète les conditions hydriques et les propriétés hydrauliques du continuum sol-plante-atmosphère lors de l'assimilation du carbone. Nguyen-Queyrens *et al.* (1998) ont mesuré le $\delta^{13}C$ de la cellulose du bois des cernes compris entre 1972 et 1991 sur des pins de 26 ans de provenance landaise et marocaine. Les valeurs de $\delta^{13}C$ enregistrées pour les cernes successives ont des variations suivant globalement les évolutions climatiques. Les périodes de sécheresses correspondaient à une augmentation du $\delta^{13}C$. La provenance marocaine avait des valeurs plus négatives de $\delta^{13}C$ quelque soit l'âge des cernes. Comme nous l'avons déjà signalé, ce ne sont donc pas forcément les provenances des régions les plus arides qui utilisent au mieux l'eau disponible. Les auteurs ont avancé les hypothèses que la provenance marocaine dont le système racinaire est plus développé serait plus évitante. Brendel *et al.* (2002) ont identifié des QTL (Quantitative trait Loci) pour le $\delta^{13}C$ et ont montré qu'il s'agissait d'un paramètre ayant une héritabilité significative. Il existe donc un déterminisme génétique de la discrimination isotopique du ^{13}C.

Le potentiel osmotique est régulé lors du stress hydrique et est variable entre génotypes issus du croisement des provenances corse et landaise (Nguyen-Queyrens *et al.* 2002). Un potentiel osmotique bas va permettre le maintien de la turgescence à des potentiels hydriques bas. L'application d'un stress osmotique par ajout de PEG a induit une diminution du potentiel osmotique des racines de plants marocains et landais. Le stress appliqué était de

courte durée (6 jours maximum) pour une intensité comprise entre -0.03MPa et -0.8MPa. Les provenances marocaines ont maintenu une croissance racinaire plus élevée et ont présenté un ajustement osmotique plus efficace que les provenances landaises (Nguyen and Lamant 1989b). Cette dernière caractéristique a été confirmée par Loustau *et al.* (1995) en exposant des plants d'origines landaises, espagnoles et marocaines à des conditions salines. L'impact du stress osmotique sur la croissance racinaire et le grandissement cellulaire a été étudié en détail par Triboulot (1996) dans le cadre de sa thèse.

Sur les arbres adultes des Landes de Gascogne, il semble que le potentiel hydrique foliaire ne dépasse jamais une valeur seuil de -2.0MPa indépendamment de la gravité et de l'intensité du stress hydrique (Delzon *et al.* 2004).

Nguyen et Lamant ont analysé des osmolytes pouvant être responsables de cet ajustement osmotique. Les auteurs ont mesuré la concentration en K^+ de plants soumis à un stress osmotique PEG. Les variations en concentration observées (+30mM) dans les racines de la provenance marocaine par rapport à la provenance landaise peuvent s'expliquer par la simple déshydratation des cellules. Le K^+ n'est pas responsable des variations de potentiel hydrique observées (Nguyen and Lamant 1989a). Les auteurs ont également conduit une expérience pour mesurer l'accumulation du D-pinitol et du myo-inositol entre ces deux provenances. Les auteurs ont montré tout d'abord que le niveau de pinitol était 2 à 3 fois plus élevé dans des conditions de bonne alimentation en eau pour la variété marocaine que pour la landaise, tant au niveau des racines qu'au niveau des parties aériennes. Le pinitol était accumulé de façon similaire entre les deux provenances lors du stress. Les concentrations de Myo-inositol semblaient augmenter uniquement dans les racines des plants marocains (Nguyen and Lamant 1988).

Une étude de Lamhamedi *et al.* (1992) sur la colonisation du système racinaire du pin maritime par des champignons ectomycorhiziens et l'effet sur le potentiel hydrique à l'interface sol-racine souligne l'impact de telles symbioses sur l'amélioration de la résistance de l'arbre à des conditions sèches.

Les travaux les plus récents de l'équipe d'écophysiologie INRA-EPHYSE, concernent la modélisation des effets des changements climatiques globaux sur la production des forêts françaises (Loustau *et al.* 2005). Si les modèles prévoient une augmentation de la production sur la période 2030-2050 suivie d'un plateau ou d'un déclin en 2070-2100, les auteurs soulignent aussi l'importance des variations régionales et des espèces cultivées. L'élévation de température et l'allongement de la saison de végétation qui en découlent, profiteront essentiellement aux espèces à feuilles caduques et non au pin maritime. Les espèces ont des

réponses différentes aux augmentations de la teneur en CO_2 et au déficit hydrique, le chêne ferme ses stomates si les concentrations augmentent ce qui n'est pas le cas du pin maritime. Finalement, les modèles étudiés prévoient que l'effet positif de l'accroissement en [CO_2] sur la production sera inhibé par la sécheresse dans le sud de la France.

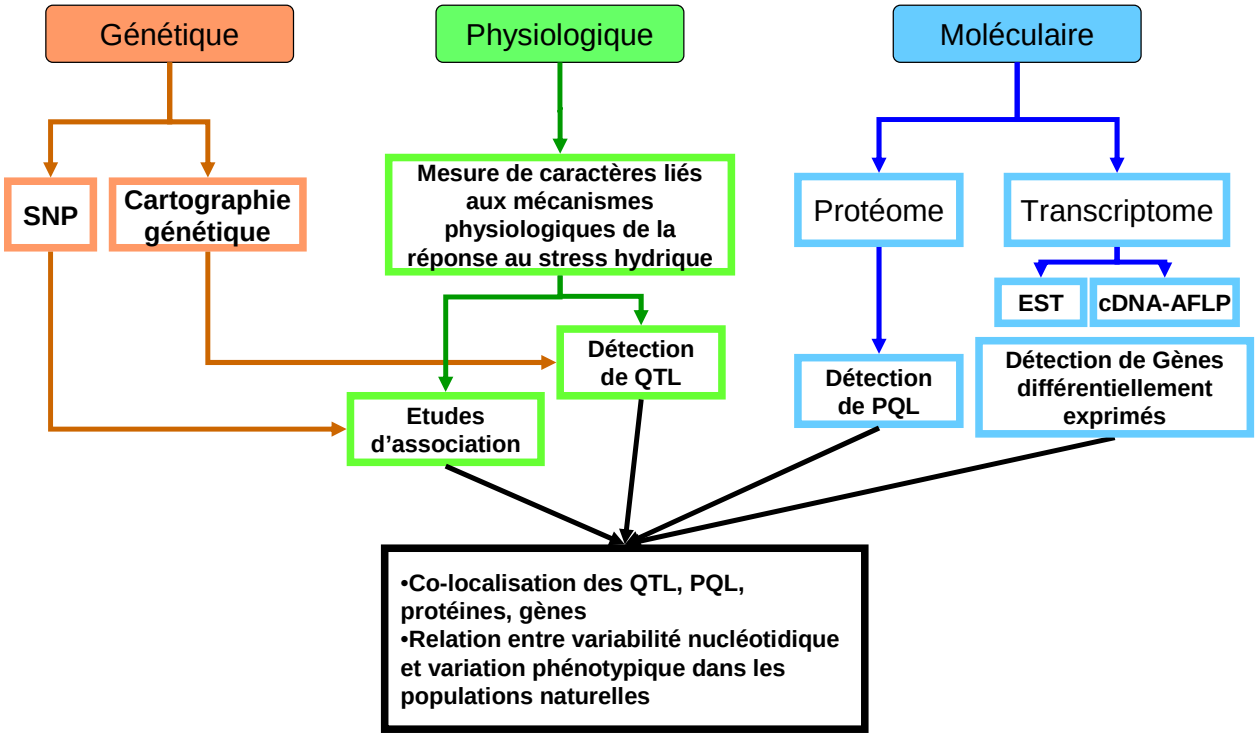
B.3.2.2 Les Approches génétiques et moléculaires

Les résultats des recherches citées ci-dessus ont montré qu'il existe une variabilité de réponse au stress hydrique entre les différentes provenances. De plus, il apparaît qu'une part de cette variabilité a un déterminisme génétique. L'UMR BioGEco a donc lancé des programmes de recherche dont le but est d'identifier les gènes impliqués dans la réponse à la sécheresse. Cela permettrait dans le futur de : 1) sélectionner des variétés sur leur capacité à résister à la sécheresse tout en conservant une bonne productivité et une bonne qualité du bois. 2) fournir des critères diagnostiques (diversité génétique dans ces gènes) pour évaluer la potentialité d'évolution des populations d'amélioration et des populations naturelles aux changements climatiques. La démarche qui a été adoptée pour la recherche du déterminisme moléculaire de la réponse au stress hydrique est schématisée à la figure B-40.

B.3.2.2.1 Cartographie et détection de QTL

Une carte génétique a été construite par Costa (1999) au cours de son travail de thèse et a servi de support pour la localisation de QTL. La descendance utilisée était constituée de 200 individus F2 issus de l'autofécondation d'un hybride entre 2 arbres élites Corse et Landais. Les plants âgés de 2 ans ont subi un stress hydrique progressif par simple arrêt d'arrosage pendant 6 semaines. Costa (1999) a recherché des QTL pour un ensemble de caractères décrivant les comportements hydriques (potentiel hydrique de base (Ψ_w), consommation en eau (C), potentiel osmotique (Ψ_s), teneur relative en eau (TRE)), les échanges gazeux (conductance stomatique (g_s), transpiration (E), photosynthèse maximale (A_{max}), $\delta^{13}C$) et la croissance (croissance en hauteur, biomasses aériennes et racinaires, analyses minérales (N, P, K)) avant et toutes les semaines pendant le stress. Les QTL détectés avaient des effets relativement faibles et se concentraient essentiellement sur les groupes de liaison 6, 8 et 12. Une des régions génomiques co-localisant des QTL pour des caractères mesurés au cours du stress (Ψ_w, TRE, g_s, A_{max} et E) semblait être particulièrement importante dans le contrôle génétique de la réponse au stress hydrique chez le pin maritime. Deux QTL pour le $\delta^{13}C$ (mesuré sur les aiguilles) ont été détectés. Le QTL majeur expliquait 12.3% de la variabilité du caractère et se positionnait sur le groupe 6.

Figure B-40 : Connexions entre les approches génétiques, physiologiques et moléculaires utilisées pour étudier la réponse à un stress hydrique chez le pin maritime.



Une étude complémentaire portant sur la discrimination isotopique du ^{13}C dans le bois et la largeur des cernes (croissance radiale) a été menée par Brendel *et al.* (2002). Un premier dispositif expérimental a été utilisé pour estimer la variabilité, l'héritabilité et les corrélations génétiques des caractères étudiés. Il s'agissait d'un demi-diallèle 12x12 de 564 arbres dont les parents avaient été sélectionnés pour leur croissance et leur rectitude. Les auteurs ont mis en évidence une héritabilité significative des deux caractères, ce qui laisse envisager des gains possible en croissance et en efficacité d'utilisation de l'eau (WUE) par croisement de génotypes favorables. Pour les auteurs, d'après le modèle de Farquhar, la corrélation positive constatée entre le $\delta^{13}C$ et la croissance suggère que la variation de WUE entre les arbres aurait pour origine une différence d'assimilation photosynthétique plutôt qu'une régulation stomatique différente. Une famille de plein-frères, constituée de 202 individus a permis d'étudier l'architecture génétique de ces caractère complexes. Quatre QTL ont été détectés pour le $\delta^{13}C$ expliquant près d'un tiers de la variation phénotypique observée. En revanche, aucune co-localisation entre les QTL de $\delta^{13}C$ et ceux de la largeur moyenne des cernes n'a pu être mise en évidence, ce qui suggère l'absence de contrôle génétique commun aux deux caractères.

B.3.2.2.2 Identification des protéines impliquées dans la réponse au stress hydrique

La première étude réalisée par Costa (1999) a porté sur les individus F2 utilisés pour la cartographie. Le protéome de 114 de ces plants a été analysé avant et après application du stress par électrophorèse bidimensionnelle des protéines extraites sur les aiguilles. L'objectif de cette expérience était d'identifier et de caractériser des protéines candidates pour la réponse au stress hydrique. Il a étudié 163 protéines (spots) et caractérisé 43 d'entre elles par microséquençage interne (Costa *et al.* 1998). Sur les 163 spots, 120 spots (dont 75 sous-exprimés et 45 surexprimés) ont montré une réponse au stress. La caractérisation des protéines d'aiguille du pin maritime qui répondent au stress hydrique, a montré que différents métabolismes cellulaires étaient impliqués : métabolisme du carbone, de l'azote, détoxication de la cellule, métabolisme de la paroi cellulaire, protéines chaperonnes, de choc thermique ou actine.

Parallèlement, le regroupement des spots en fonction de leur profil d'expression a permis d'obtenir un groupe de protéines accumulées et un autre de protéines dont la quantité diminue entre les deux conditions. Chaque groupe est lui-même divisé en sous-groupes de protéines dont l'évolution de la quantité évolue de façon similaire. Un sous-groupe intéressant

comprenait six variables physiologiques (Ψ_w, TRE, g_s, A_{max}, E, C) auxquelles se rattachaient des protéines de fonction inconnue et dont la quantité diminuait dans la condition stressée, montrant une implication possible de ces protéines dans la réponse au stress.
La disponibilité d'une carte génétique du pin maritime a rendu possible l'identification de PQL (Protein Quantity Loci) dont certains co-localisaient avec des QTL des caractères physiologiques tel que celui correspondant à la discrimination isotopique du carbone ($\delta^{13}C$).
Parmi les protéines identifiées et dont la quantité est corrélée négativement ou positivement aux mesures physiologiques on trouve des HSPs, des Cu/Zn SOD, la disulfide isomérase, les SAM synthases, l'isoflavone réductase, la glutathione peroxydase, la CCoA-OMT et des protéines de la photosynthèse. La conductance stomatique (g_s) était le paramètre présentant le plus de relations significatives avec l'intensité des spots dans la condition contrôle, tandis que dans la condition stressée, le paramètre interférant le plus était la Teneur Relative en Eau (TRE).

Une autre étude, réalisée par Dubos (2001), portait sur de jeunes plants cultivés en hydroponie et soumis à un stress PEG. Les protéines ont été extraites et des variations significatives de quantité ont été trouvées sur les aiguilles (34 spots (8%)), les tiges (51 spots (14%)) et les racines (53 spots (14%)). Dans chaque tissu, des spots présentaient des régulations positives (Aiguilles : 16, Tiges : 32, Racines : 32) lors du stress et d'autres une régulation négative. Certaines variations étaient d'ordre qualitatif avec apparition ou disparition lors du stress. Parmi ces protéines certaines étaient communes à celles identifiées par Costa comme la rubisco activase, une Hsp de faible poids moléculaire, des SAM synthase et une protéine chaperonne Hsp70.

B.3.2.2.3 Identification des transcrits impliqués dans la réponse au stress hydrique

A l'issu de l'approche protéomique, Dubos (2001) a démarré une thèse sur l'étude de la réponse au stress hydrique de jeunes plants de pin maritime en menant une investigation sur la régulation du transcriptome. Si l'arrêt d'arrosage de l'expérience précédente reproduisait assez bien la situation des plants lors d'une sécheresse estivale, il avait pour inconvénient de ne pas maîtriser de façon précise l'intensité du stress et les conditions de culture appliqués. De plus, l'accès au système racinaire n'est pas aisé et l'absence de mycorhizes n'est pas garantie. Dubos a donc opté pour une culture hydroponique et l'application d'un stress osmotique par adjonction de Polyéthylène Glycol. Dans sa thèse, Dubos a étudié un factoriel croisant intensité (-0.08 MPa ; -0.15 MPa ; -0.30 MPa ; -0.45 MPa ; -0.60 MPa ; Terre) et durée (0 ; 1 ;

10 ; 21 ; 45 jours) de stress hydrique sur des plantules d'origine landaise âgées de 3 semaines. Afin d'identifier les variations d'expression des transcrits dans les parties aériennes et les parties racinaires, une technique d'ADNc-AFLP a été adoptée. Elle a permis la détection de 81 gènes régulés au cours du stress. Des analyses d'expression par Northern et RT-PCR ont été réalisées sur 40% des gènes et ont confirmé leur expression différentielle. Les gènes de fonction connue impliqués dans la réponse au stress hydrique font partie du métabolisme des sucres, de la photosynthèse, de la synthèse des parois et des mécanismes de défense comme la PR-10 décrite comme régulée lors des stress biotiques et abiotiques (Dubos and Plomion 2001; Dubos *et al.* 2003). Au niveau des racines, des transcrits accumulés lors du stress sont homologues à l'aldéhyde déshydrogénase (métabolisme des lipides), la pLP3 (protéine induite lors du stress chez *P. teada*), une glycolate oxidase, une ATPase, une ACC oxydase (éthylène), une Xyloglucan Endotransglycosylase (Parois) ou une S-adenosyl-L-homocysteine hydrolase (SHH déjà décrite lors du stress)(Dubos and Plomion 2003). Dans les gènes réprimés lors du stress, on trouve des homologues à la COMT (caffeic acid ortho-methyl transferase)(synthèse lignine), une protéine putative RING zing finger (facteur de transcription), une pre-protéine translocase SecA, une protéine de la famille des LEA hydrophobes ou encore une histone H2B.

Au niveau des aiguilles, Dubos *et al.* (2003) ont détectés 25 transcrits présentant des différences significatives d'expression. Les amplitudes de régulation mesurées allaient de 1.18 à 2.14 pour les 11 transcrits induits, et de 1.49 à 16.6 pour les 14 transcrits dont l'expression diminuait lors du stress. Les fonctions identifiées avaient trait aux mécanismes de défense de la plante (flavonoid 3′,5′-hydroxylase (F3′H), dihydroflavol 4-reductase (DFR), pathogenesis-related protein, allene oxide synthase (cytochrome P450 protein), glucan endo-1,3-β-glucosidase) et étaient induits lors du stress. Tandis que les gènes réprimés étaient plutôt reliés aux gènes de la photosynthèse (preprotein translocase SecA, hlorophyll a/b-binding proteins type 4 et type 1, small subunit of the ribulose-biphosphate carboxylase).

B.4 Stratégies et techniques pour la détection des gènes impliqués dans l'adaptation du pin maritime à la sécheresse.

Pour la découverte des gènes dont le rôle est majeur dans la relation de la plante à son environnement, l'approche expérimentale se base sur la comparaison de génotypes contrastés placés dans des conditions où l'expression des gènes d'intérêts a un impacte sur le phénotype observé.

Lorsqu'ils sont disponibles, l'utilisation de mutants est une possibilité (Nadeau and Frank 2000). Nous avons cité plusieurs travaux adoptant cette démarche pour l'identification de gènes spécifiques. Cependant, cela ne permet la découverte que d'un petit nombre de gènes. Une autre option consiste à utiliser la variabilité des populations naturelles pour le caractère étudié. Dans le cas du pin maritime, nous disposons de variétés très contrastées vis-à-vis de leur réaction à un déficit en eau. L'étude de ces écotypes en situation de stress hydrique devrait permettre d'observer une plasticité moléculaire différente entre provenances. La notion de plasticité phénotypique se définit comme l'aptitude d'un génotype à donner différents phénotypes en fonction de l'environnement. Par analogie, nous définissons la plasticité moléculaire comme la propriété d'un génotype à modifier l'expression de son génome en réponse à des environnements distincts.

L'utilisation des QTL (Quantitative Trait Loci) pour la détection des régions du génome expliquant une partie de la variabilité du caractère étudié, est aussi une alternative séduisante (Borevitz and Chory 2004). Une des difficultés de ce type d'étude tient d'une part à la transférabilité des QTL détectés d'un dispositif expérimental à un autre, et d'autre part à l'identification d'un gène précis, étant donné l'intervalle de confiance autour de la position la plus vraisemblable d'un QTL. Les études d'associations permettent d'étudier de façon plus fine la corrélation entre un polymorphisme nucléotidique d'un gène candidat et la variation du caractère étudié (Cardon and Bell 2001). Ce type d'étude est actuellement en cours sur le pin maritime et le stress hydrique dans le cadre de la thèse d'Emmanuelle Eveno.

La découverte des gènes peut se limiter pour un tissu donné au génome exprimé. La transcriptomique regroupe un ensemble de techniques d'étude des transcrits. La comparaison des niveaux d'expression d'un gène pour les écotypes et dans les conditions testées permet de déterminer si une variabilité d'expression existe pour ce gène. L'analyse du profil d'expression permettra d'identifier des gènes candidats expressionnels. Certaines techniques comme la PCR quantitative vont permettre d'étudier précisément un gène mais ne sont pas adaptée à l'étude à de milliers de gènes exprimés dans un tissu. La réalisation de banques d'ADNc offre la possibilité d'insérer des ADNc dans des vecteurs puis de les cloner dans des bactéries. Une approche particulièrement intéressante est la création de banques suppressives soustractives par hybridation (SSH) (Diatchenko *et al.* 1996). Cette méthode permet d'éliminer la redondance et donc d'enrichir la banque obtenue en transcrits différentiellement exprimés. Un séquençage des clones permet d'identifier les gènes plus spécifiquement exprimés dans un écotype et/ou une condition donnée.

Les analyses d'expression par hybridation de cibles sur des sondes constituent une autre possibilité. Elles permettent l'étude simultanée de l'expression de milliers de gènes. Quelque soit le type de filtre (dépôt bactérien, ADNc ou oligomères), il faut néanmoins avoir accès à la séquence des sondes déposées. La génération d'EST s'est positionnée depuis une dizaine d'années à la fois comme une alternative et un complément au séquençage complet de génomes (Rudd 2003). Les projets de séquençage se sont développés avec l'utilisation des microarrays. La production massive de ce type de séquences courtes de basse qualité nécessite l'utilisation de moyens bioinformatiques de traitement des données. Parallèlement au développement de ces techniques, les outils bioinformatiques ont également fortement progressés pour répondre aux besoins croissant de traitement des données générées. Cela concerne le traitement des séquences, leur assemblage pour reconstituer la séquence complète des gènes, l'identification de leur fonction mais aussi l'analyse des données d'expression et la détection de gènes ou réseaux de gènes d'intérêt. Les bases de données telles que dbEST (http://www.ncbi.nlm.nih.gov/dbEST/), PlantGDB (http://www.plantgdb.org/) et TIGR PGI (http://www.tigr.org/tdb/tgi/plant.shtml) ont ouvert de nouvelles possibilités d'étude sur le génome exprimé. Mais les caractéristiques de ces séquences et leur traitement informatique soulève un certain nombre de questions que je développerai dans la partie résultat de cette thèse.

Nous disposons donc de nombreuses informations physiologiques et moléculaires sur la réponse du pin maritime au stress hydrique. Il nous reste cependant à identifier de façon plus exhaustive les gènes dont l'expression est régulée au cours du stress. Cette thèse a pour but d'utiliser de nouvelles approches pour pouvoir analyser l'expression et l'implication potentielle du plus grand nombre de gènes afin d'identifier les gènes candidats pour les mécanismes de résistance au stress hydrique.

Matériel & Méthodes

C MATERIEL ET METHODES

C.1 Matériel végétal et conditions de culture

Deux écotypes (provenances) de pin maritime (*Pinus pinaster* Ait.) –Aquitaine (France) et Tamjout (Maroc) – ont été utilisés pour réaliser cette étude.

Les graines d'origine Aquitaine ont été récoltées lors d'une campagne de ramassage réalisée par le CEMAGREF (Centre de Machinisme Agricole et Génie Rural des Eaux et Forêts) dans des peuplements classés du Médoc. Ce lot de graines avait également été utilisé dans le cadre de la thèse de C. Dubos (http://www.pierroton.inra.fr/biogeco/genetique/theses.html) (Dubos 2001).

La provenance marocaine est issue du peuplement de Tamjout (Fig. C-1), massif de 500 ha caractérisé par une proportion importante de carbonate et de calcaire actif dans le sol à partir d'un mètre de profondeur. Les pins y sont très droits, à branches fines régulièrement réparties le long du tronc.

Les graines ont été conservées en chambre froide à 4°C et maintiennent une très bonne capacité germinative.

C.1.1 Germination

Les conditions décrites par Dubos (2001) exigeaient une stérilisation des graines par de l'H_2O_2 30% puis une germination sur bande de papier Wattman (3M) stériles humides traitées par 3 antifongiques dans des germoirs placés à 18°C. Les graines ainsi traitées se sont révélées sensibilisées à la présence de pathogènes.

Une technique plus classique a donc été adoptée. Des germoirs désinfectés à la Javel sont remplis d'une couche de Vermiculite stérilisée à l'autoclave (121°C, 20min.) (Fig. C-2). Les graines sont ensuite réparties dans les germoirs puis recouvertes d'une seconde couche de Vermiculite. Le substrat est largement humidifié et la germination s'effectue à température ambiante sous éclairage fluorescent. La vermiculite offre l'avantage par rapport au papier humide de ne pas être un substrat aussi favorable au développement bactérien ou fongique. Les plantules doivent rester 1 semaine maximum dans les germoirs et les mégagamétophytes doivent être retirés des cotylédons pour éviter une « fonte du semis ». Aucun traitement sanitaire n'a été nécessaire.

Afin d'obtenir un nombre suffisant de plantules levant au même moment, la quantité de graines a été multipliée par 4 par rapport aux besoins expérimentaux.

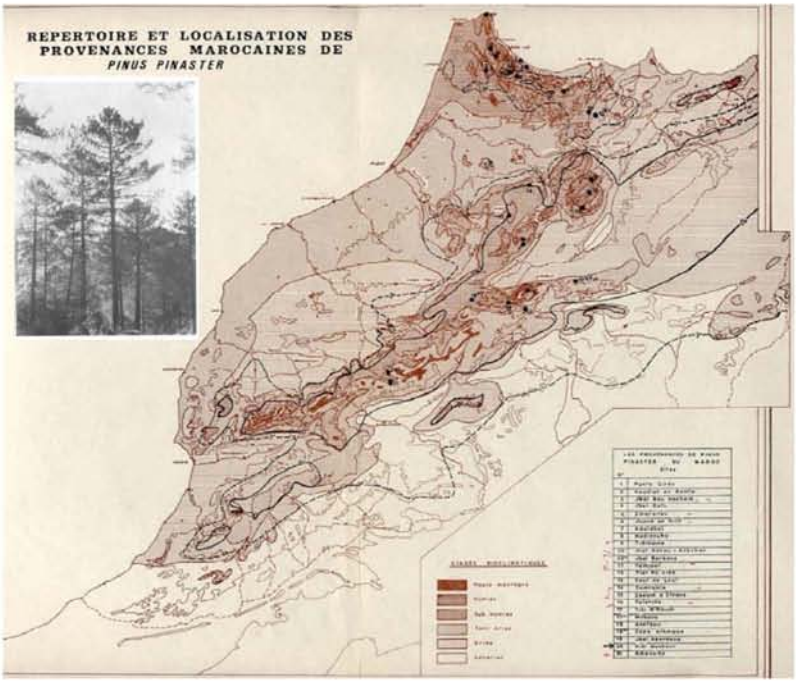


Figure C-1 : Situation géographique et photographie de l'écotype marocain « Tamjout ».

Figure C-2 : Dispositif utilisé pour la germination des graines.

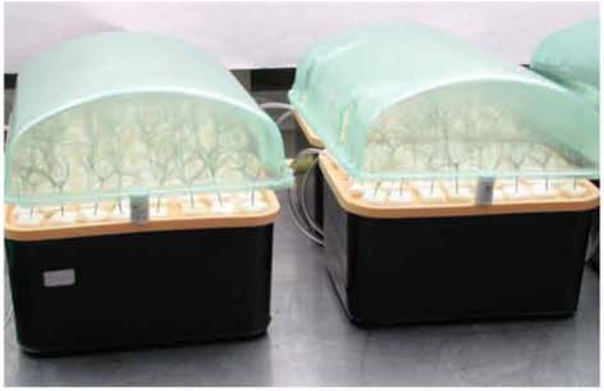


Figure C-3 : Etape de transfert des plantules depuis les germoires vers les conditions hydroponiques.

Figure C-4 : Détail du système de maintient des plants par des carrés de mousse et opacification des bacs par un film plastique.

Figure C-5 : Système d'oxygénation de la solution nutritive par injection d'air à travers des diffuseurs pour aquarium.

Tableau C-1 : Composition du milieu de Seillac (1960)

Solution 1X	[]° milieu final (µM)
$MgSO_4$ $7H_2O$	500
$CaCl_2$ $2H_2O$	500
KH_2PO_4	1000
NH_4NO_3	4000
H_3BO_3	16,7
$MnCl_2$ $4H_2O$	3
$ZnCl_2$	0,3
$CuCl_2$ $2H_2O$	0,3
Na_2MoO_4	0,07
EDTA	100
$FeSO_4$ $7H_20$	100

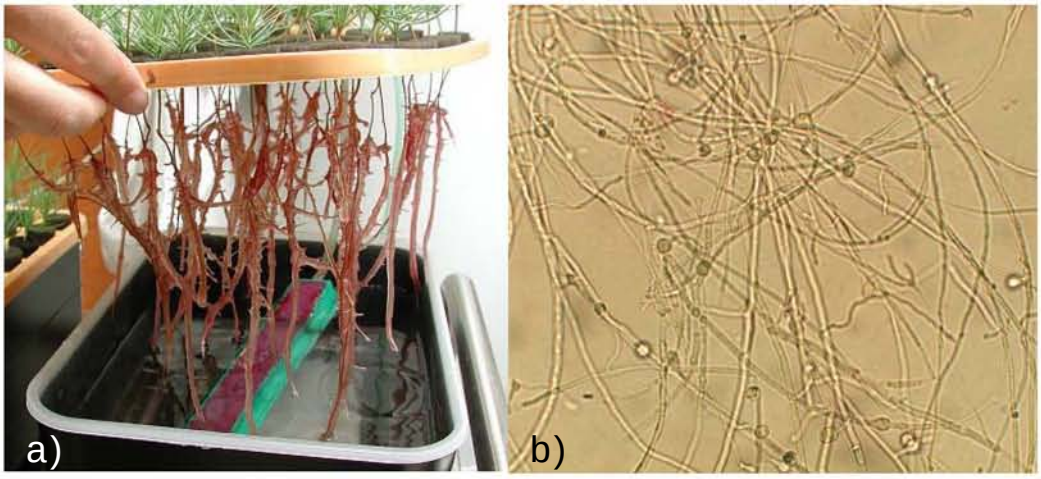


Figure C-6 : a) Mucilage de couleur rose provoqué par l'attaque d'un pathogène.
b) Vue microscopique d'un prélèvement du mucilage.

C.1.2 Culture Hydroponique

Ce mode de culture offre l'avantage de faciliter l'accès aux parties racinaires (objet de notre étude), d'éviter les mycorhisations qui perturberaient l'analyse de la réponse propre des plants à la sécheresse et de contrôler et d'homogénéiser l'apport en éléments nutritifs. Cependant, l'application d'un stress hydrique ne peut se faire que par un stress osmotique et le dispositif ne permet l'étude des plants qu'à un stade de développement précoce.

Après l'apparition des euphylles, les plantules ont été débarrassées des téguments pouvant rester au niveau du collet (source de développement de pathogènes), puis transférés en milieu hydroponique. Pour éviter un desséchement trop violent dû au transfert des bacs de germination (humidité 100%) vers les conditions hydroponiques, des couvercles de germoirs ont été placés sur les bacs de culture pendant 48h (Fig. C-3). Les bacs ont été remplis d'eau le temps de transférer la totalité des plants et de mettre en route le dispositif en présence de solution nutritive.

Les bacs de culture d'une contenance de 10L ont été opacifiés à l'aide de film plastique noir opaque. Leurs couvercles ont été percés et les plantules sont maintenues par des carrés de mousse fendue permettant la croissance de la tige (Fig. C-4). La solution nutritive, basée sur celle décrite par Seillac (1960) (Tab. C-1) est autoclavée et renouvelée une fois par semaine pour éviter un appauvrissement en éléments nutritifs, une variation du pH ou le développement de pathogènes. L'utilisation d'H_2O miliQ permet de s'affranchir d'une variation dans les teneurs en micro-éléments et de l'apport de pathogènes. A chaque renouvellement, les bacs sont lavés et javellisés. L'oxygénation du milieu est assurée par un compresseur relié à des diffuseurs d'aquarium de 25 cm. Le débit est réglé individuellement pour chaque bac via un robinet pour obtenir un bullage optimal tout en évitant des turbulences trop prononcées au niveau des racines (Fig. C-5).

C.1.3 Contrôle des pathogènes

Deux séries de cultures successives ont été mises en œuvre en utilisant du milieu de Seillac 1x puis 0.2x et ont subi une attaque de pathogène dans les jours suivant l'application du stress. L'agent pathogène a un développement fulgurant sur 48h. Les racines se recouvrent d'un mucilage rose puis se nécrosent très rapidement (Fig. C-6a). Une préparation microscopique a permis d'écarter l'hypothèse d'une attaque bactérienne et mettre en évidence une origine fongique (Fig. C-6b). Des isolements réalisés par l'unité de santé végétale de l'INRA Bordeaux ont permis d'identifier un *Fusarium oxysporum*. Ce champignon est couramment rencontré sur les cultures de tomate hors sol. Bien qu'un traitement

phytosanitaire ne soit pas souhaitable, il apparaît comme incontournable. Le choix s'est porté sur le Tachigaren 360 qui est un fongicide systémique utilisé pour lutter contre la nécrose du collet et la pourriture des racines : *Fusarium oxysporum f. sp. radicis-lycopersici* (F.O.R.L.) en cultures de tomates. Sous forme liquide, il est particulièrement adapté à une utilisation dans le système d'irrigation. Un traitement préventif à 5ppm une semaine avant l'application et pendant toute la durée du stress a été retenu.

C.1.4 Eclairage

Le pin maritime est une espèce héliophile, la quantité et la qualité de la lumière apportée doivent donc être suffisantes et non limitantes pour la croissance.

Les expériences déjà mises en œuvre dans l'Unité avaient été réalisées dans une chambre de culture climatisée (24/18°C) et équipée de lampes MAZDA aux halogénures métalliques de 250W /4200K à raison d'environ 200W/bac apportant 260µmol.m^{-2}.s^{-1} (16/8h). Cette salle n'étant plus disponible, deux solutions se présentaient :

- la culture en serre où la quantité de lumière n'est pas limitante mais non contrôlée et où la régulation de la température devient un facteur critique essentiellement l'été où la température atteint 35°C sous cooling et peut atteindre des températures létales pour les plants en cas de panne. Cette solution, quoique envisageable pour l'automne est à éviter durant la saison estivale.
- La réalisation d'une seconde « chambre de culture » sans climatisation possible. Le maintient d'une température homogène dépend alors étroitement de la chaleur dégagée par le système d'éclairage. Cela implique l'utilisation d'un éclairage par tubes fluorescents avec déport des ballasts, et extraction de la chaleur qu'ils émettent.

C'est cette seconde solution qui a été retenue. Deux tables métalliques sur roulettes de 2m*80cm ont ainsi été construites et équipées de rampes d'éclairage (Fig. C-7).

C.1.4.1 Photosynthèse

La quantité d'énergie lumineuse est un paramètre important. L'éclairement solaire direct maximum est de 95000 lux (unité se rapportant à la vision humaine) soit 950W/m² dont 4% dans l'UV, 42% dans l'IR et 54% dans le visible =515W/m² dont 410W/m² dans la PAR (photosynthetically active radiation : 400-700) ce qui équivaut à un flux de photons d'environ 1800µmol/m²s (Heller *et al.* 1993).

Figure C-7 : Vue générale des tables métalliques équipées de leur système d'éclairage à hauteur réglable et avec déport des ballasts dans des caissons ventilés.

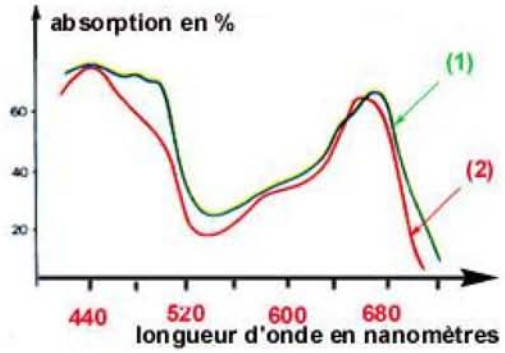


Figure C-8 : Spectres d'absorption (1) et d'activité photosynthétique (2) de l'algue verte Ulva taeniata.

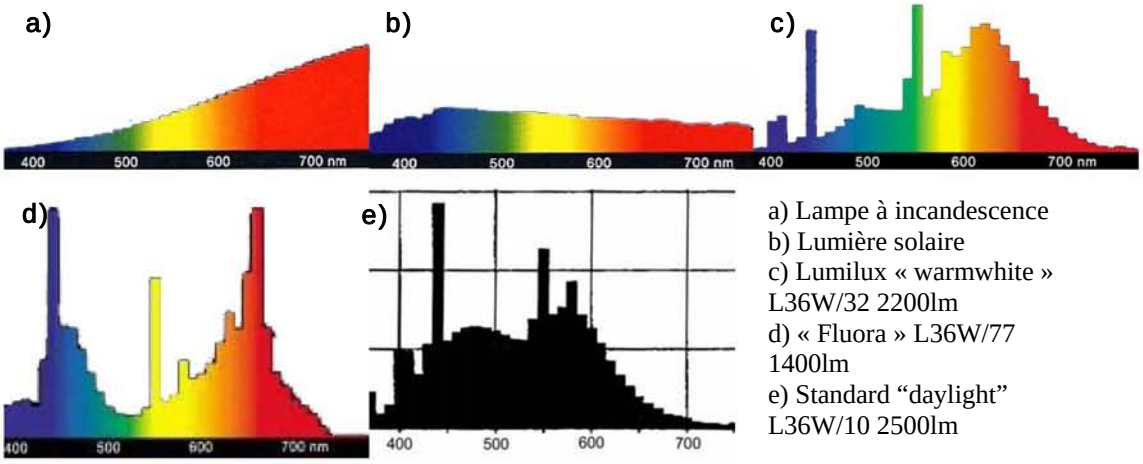


Figure C-9 : Spectres d'émission entre 400 et 700 nm de différentes sources lumineuses. (http://www.osram.com)

Comme on peut le constater chez *Ulva taeniata*, les pics d'efficacité photosynthétique et d'absorption de la lumière par la chlorophylle coïncident (Fig. C-8). Les photons utilisable par la plante pour la photosynthèse seront compris dans une gamme de longueur d'onde de 400 à 700nm (PAR). Le risque avec un éclairage fluorescent est d'avoir un éclairage insuffisant, trop proche du point de compensation (point critique où le carbone fixé lors de la photosynthèse équilibre celui rejeté par la respiration). Le rayonnement variant en fonction de l'inverse du carré de la distance à la source (Gosse, 1984 doc INRA), il est intéressant de réduire la distance entre les plantules et les tubes dont le dégagement calorifique est faible. Dans notre cas, la distance à été réduite à environ 15 cm.

C.1.4.2 Action Phytochromique

Si la croissance de la plante dépend directement de la quantité de lumière disponible, elle est également sensible à l'équilibre des couleurs émises. La morphologie de la plante est sous le contrôle du phytochrome. Ce pigment existe sous deux formes différentes, convertibles l'une en l'autre. La forme P660 absorbe la lumière dans le rouge clair (RC), l'autre, P730, dans le rouge lointain (RL). Sous l'effet de radiations RC la forme P660 se convertit en sa forme active P730 et inversement sous des radiations RL. Le phytochrome contrôle entre autres la longueur de l'entrenœud, la floraison et la germination.

Le facteur zeta qui est le rapport des photons 660nm/730nm détermine un équilibre « phytochrome actif » (phi) / « phytochrome total ». Une valeur de phi faible (RC/RL faible) favorise l'élongation des entrenœuds, une valeur de phi élevée (RC/RL important) produit des plantes très ramassées, à entrenœuds courts. En conditions naturelles, phi varie entre 0.5 (soleil) et 0.20 (sous couvert végétal), des tubes « cool white » donnent un phi de 0.7 et des tubes « Warm white » 0.65 (Gaudillere and Chasles 1986). Les sources incandescentes donneront donc des plantes étiolées à petites feuilles tandis que les sources fluorescentes classiques impliqueront des plantes trapues et ramifiées.

Avant expérimentation, il est cependant difficile de déterminer l'équilibre RC/RL obtenu.

C.1.4.3 Sources

Les courbes de distribution spectrale de l'énergie sont très variables selon la source lumineuse artificielle et diffèrent beaucoup de celles de la lumière solaire (Fig. C-9b). Les sources incandescentes ont des spectres très riches dans le rouge et l'infra- rouge (Fig. C-9a). Les tubes fluorescents classiques de type « daylight » ont une distribution assez homogène sur la gamme 400 à 700nm (Fig. C-9e) tandis que les tubes "Warmwhite" ont une distribution

centrée sur 620nm (Fig. C-9c). Les tubes spécifiques comme le « Fluora » (OSRAM) ont été développés spécialement pour un rendement élevé autour des pics d'absorption des chlorophylles a et b dans le bleu 440 / 480 nm et dans le rouge 650 / 680 nm (Fig. C-9d).
Le choix a donc porté sur l'utilisation de tubes standard « daylight », d'un tube « Fluora » à haut rendement photosynthétique centré au dessus de chaque rangée de bacs et d'un tube « warmwhite » pour rééquilibré le spectre vers le rouge.
Les mesures de PAR réalisée avec un capteur Li-COR Quantum (Fig. C-10) ont données des valeurs comprises entre 360µmol/m²s à 15 cm et 380µmol/m²s à 10cm. Les tubes sont espacés de 5 cm soit une puissance de 400W/m^2.
Une photopériode classique de 16h jour et 8h nuit a été appliquée.

C.1.5 Application du stress

Le stress hydrique est obtenu par ajout de Polyéthylène glycol (PEG) à la solution nutritive afin d'en augmenter son osmolarité. Le PEG utilisé (Sigma, réf. P3640) ne nécessite pas de purification avant utilisation pour éliminer les résidus d'aluminium ou de magnésium qui à forte concentration présentent une toxicité pour les plantules (Triboulot, 1996). La masse molaire de 3350g/mol est suffisamment élevée pour limiter l'absorption racinaire tout en préservant la fluidité de la solution nutritive. Les courbes de correspondance du potentiel osmotique de la solution de Seillac en fonction de la concentration en PEG (Dubos 2000) nous donnent –0.45MPa pour 180g/L *(milieu M3)* et –0.60Mpa pour 210g/L *(milieu M4).*
L'apparition des premières aiguilles secondaires ou pseudophylles (Fig. C-11) a été choisie comme témoin du stade de développement des plantules pour l'application du stress (Fig. C-12).

C.1.5.1 Stress court

Pour l'étude de la réponse des plantules à court terme, la solution nutritive de 4 bacs est remplacée par une solution à -0.60MPa. Un bac est également récolté pour servir de témoin à l'expérience « Stress court ». Les autres bacs sont échantillonnés à T0+2h, T0+6h, T0+24h et T0+48h.

C.1.5.2 Stress long

Pour chaque provenance, un bac est placé dans une solution à -0.45MPa pour une durée de 3 semaines et un bac est conservé dans une solution nutritive de contrôle. A l'issue de la période de « Stress Long », les échantillons témoins et stressés sont récoltés.

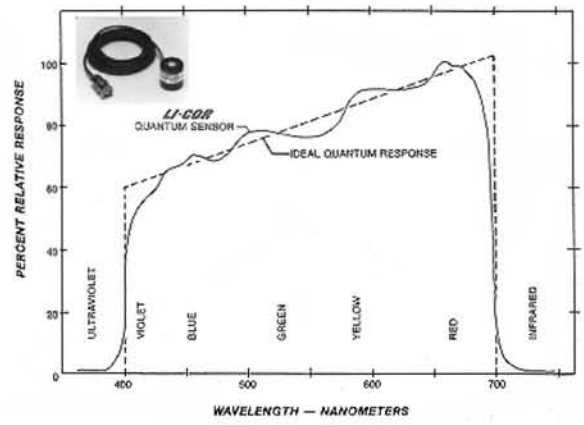


Figure C-10 : Réponse spectrale type des détecteurs « Li-COR Quantum » en fonction de la longueur d'onde et la réponse quantique idéale (réponse identique à tous les photons dans la gamma 400-700nm).

Figure C-11 : Développement des premières aiguilles secondaires ou pseudophylles.

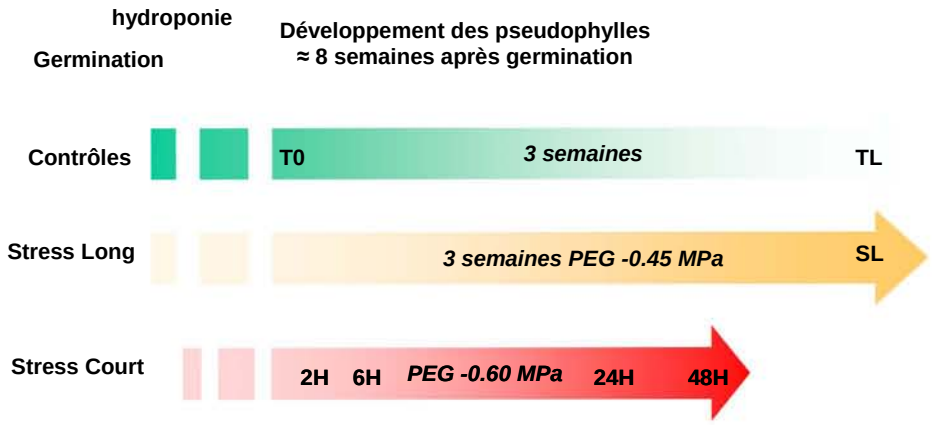


Figure C-12 : Schéma chronologique situant les prélèvements effectués à partir du développement des premières aiguilles secondaires. 7 bacs de cultures par provenance. T0 = Témoin du stress court, TL = Témoin du stress long, SL = Stress Long, 2H/6H/24H/48H = échantillons prélevés 2/6/24/48 heures après application du stress.

Figure C-13 : Enchevêtrement des racines rendant l'individualisation des plants impossible. Système racinaire des plants après 7 semaines de croissances dans la solution nutritive.

a) **b)** **c)** **d)**

Figure C-14 : Mesure du potentiel de base sur une plantule en utilisant la chambre de Schölander. De gauche à droite, dispositif avec chambre et loupe binoculaire (a), couvercle de la chambre et maintient de la plantule au niveau du collet (b), vue de l'intérieur de la chambre avec les arrivées et sortie d'azote comprimé (c), vue de la tige avec début d'exsudation de la sève sous l'effet de la pression (d).

Dans la suite de ce mémoire, une combinaison Provenance x Condition sera nommée un « échantillon ».

Etant donné que nous avons 5 conditions pour le stress court (T0, T0+2h, T0+6h, T0+24h et T0+48h), 2 conditions pour le stress long (TL, SL), et 2 provenances (France, Maroc), nous parlerons de 10 échantillons pour l'étude stress court et 4 échantillons pour l'étude stress long. Les échantillons T0+2h et T0+24h, n'ont pas été utilisés dans cette thèse pour la réalisation des études d'expression.

C.1.6 Récolte du matériel végétal

Les acides nucléiques et en particulier les ARN sont particulièrement sensibles à la dégradation par les enzymes du type ARNase. Certains gènes peuvent également répondre rapidement à un changement de l'environnement. Les parties racinaires formant un enchevêtrement inextricable (Fig. C-13), les plantules ont donc été sectionnées au niveau du collet, puis la masse racinaire coupée par lots d'une dizaine d'individus. Après un égouttage rapide, les tissus sont plongés immédiatement dans un bain d'azote liquide, emballés puis stockés à -80°C. L'option d'un rinçage des racines pour éliminer le PEG a été écartée pour préserver les pousses secondaires fragiles. Les parties aériennes (aiguilles + tige) sont traitées de la même manière et stockées à -80°C. Elles n'ont pas été analysées dans cette étude.

C.2 Mesure du niveau de stress et paramètres physiologiques

C.2.1 Mesure du potentiel de base

Lors de l'application d'un stress de type osmotique par ajout d'un osmoticum à la solution, il est important d'évaluer le niveau de stress réellement provoqué chez la plante.
Un paramètre couramment utilisé en écophysiologie est la mesure du potentiel hydrique de base, estimée par la mesure du potentiel hydrique foliaire à l'aube et exprimée en unité de pression. Nous avons mesuré ce potentiel avant la remise en route de l'éclairage. La pièce étant totalement dans l'obscurité pendant toute la durée des mesures. Dans notre cas, la taille de nos plants a permis de mesurer le potentiel directement sur la tige en plaçant l'ensemble des parties aériennes dans une chambre de Schölander (Fig. C-14). Des mesures ont été réalisées sur les stades T0, T0+48h, TL et SL.

C.2.2 Dosage des terpènes

La teneur de ces composés volatiles dans les parties aériennes est connue pour varier au cours de stress abiotiques (Penuelas and Llusia 1999) (Llusia and Penuelas 2000) et biotiques (Faldt *et al.* 2003) (Tiberi *et al.* 1999). Elle permet également de différentier les provenances Aquitaines et Marocaines. Au sein du laboratoire, des études d'association entre caractères phénotypiques dont la teneur en terpènes et variants nucléotidiques de gènes candidats sont actuellement développées dans le cadre d'une thèse. L'étude porte sur les variations de la résistance au stress hydrique de 20 populations de pin maritime distribuées selon un cline de précipitation. Différentes familles de France, Espagne, Corse, Tunisie et Maroc ont été échantillonnées et plantée en Espagne et en France dans un dispositif expérimental comprenant 9000 plants. Parmi les caractères mesurés, un dosage des terpènes doit permettre de caractériser des différences de comportement entre ses familles. Nous avons donc fait le choix de mesurer également les terpènes dans le cadre de cette thèse.

Des aiguilles ont été prélevées sur plusieurs plantules des échantillons de l'expérience « stress long ». Chaque prélèvement a été pesé, découpé et conservé dans des piluliers remplis de pentane. Les échantillons ont été analysés au laboratoire Chimie Analyse de l'Institut du Pin (Patrick PARDON). La teneur en terpène est mesurée par chromatographie en phase gazeuse (Trace GC, Thermo Quest ; injecteur AS 2000, Thermo Finningan ; colonne, SGE 25 QC2/BP21 0.25) et la détection se fait par un détecteur à ionisation de flamme (FID) intégré à l'appareil.

C.3 Analyse du transcriptome et caractérisation des EST

C.3.1 Extraction des ARN

Les échantillons racinaires sont broyés manuellement sous azote liquide. Cette étape est déterminante et le matériel doit être broyé très finement.

Les ARN totaux sont obtenus selon le protocole de Chang et al. (Chang *et al.* 1993) modifié pour améliorer les rendements et la qualité des extractions sur notre matériel d'étude :

- Le SDS a été retiré du tampon d'extraction et le volume a été ramené à 15mL pour 1g de matière fraîche.
- Une extraction Chloroforme/Isoamyl alcool (24 :1) avec rinçage de la phase aqueuse a été ajoutée au protocole. Cette étape permet d'augmenter le rendement jusqu'à 30%.

Les acides nucléiques sont précipités en présence d'Acétate de sodium à 0.3M et 2 volumes d'éthanol à -20°C pour une durée de 2 heures. Le culot obtenu après centrifugation est repris dans de l'eau exempte de nucléases.
Un traitement à la DNAse suivit d'une nouvelle extraction Chloroforme/Isoamyl alcool (24 :1), précipitation puis solubilisation et purification est effectuée pour éliminer la DNAse, le tampon et les produits de dégradation.

La qualité des ARN est ensuite contrôlée par électrophorèse sur gel d'agarose permettant de détecter les dégradations éventuelles et la présence d'ADN. Un spectre d'absorption entre 220 nm et 300nm nous permet d'estimer la pureté de l'extrait en calculant les rapports de D.O. (DO_{260nm}/DO_{280nm} et DO_{260nm}/DO_{230nm}) et de déterminer la quantité d'ARN (DO_{260nm}).
Le rendement varie entre 70 et 160 µg ARNt/g MF suivant la nature de l'échantillon. Le rapport DO_{260nm}/DO_{280nm} était voisin de 2.

Lors de chaque série d'extraction, 5-6 g de matière fraiche sont traitées dans 6 tubes indépendants.
Pour chaque traitement, un minimum de 3 séries d'extractions d'ARNs distinctes ont été réalisées jusqu'à obtention d'une quantité suffisante pour les études d'expression.
Les extraits de bonne qualité ont été « poolés » - ce qui permet d'éviter les biais possibles d'une seule extraction - et dosés. Ils ont alors été précipités et repris dans de l'eau DEPC afin d'obtenir une solution stock de concentration élevée comprise entre 6 et 10 µg/µL et de rapport DO_{260nm}/DO_{280nm} proche de 3, indispensable à la synthèse des cibles à hybrider sur les microarrays.

C.3.2 Banques d'ADNc

Les études d'expression à haut débit telles que les puces à ADNc nécessitent le développement préalable de ressources génomiques et en particulier de banques d'ADNc. Dans le cas présent, deux banques distinctes ont été construites : l'une à partir de xylème (tissus conducteur) en différentiation et l'autre à partir de tissus racinaires.

C.3.2.1 Banque de xylème

Dans le cadre du projet européen GEMINI (http://www.pierroton.inra.fr/Gemini/), l'équipe du Professeur Canovas de l'Université de Malaga a produit une banque d'ADNc en utilisant

du xylème en différentiation associé à 4 types de bois (Xylème associé à la face comprimée, à la face opposée, à du bois initial et à du bois final) fournit par notre unité.
Les ARN totaux des différents types de xylème ont été obtenus selon la méthode de Chang et al. (1993) et mélangés en quantité égale. Les ADNc et les banques ont été synthétisées en utilisant le ZAP-cDNA synthesis kit (Stratagene, La Jolla, CA, USA). Environ 10000 clones ont été repiqués et expédiés à la société Genome Express pour le séquençage d'EST en 5'. Une analyse détaillée et l'exploitation de cette ressource a été menée dans le cadre de la thèse de Jorge Paiva (2006).

C.3.2.2 Banques de racines et aiguilles

Ces banques ont été construites par J.M. Frigério (INRA de Pierroton) sur du matériel végétal issu d'une expérience de culture hydroponique de plantules de pin maritime soumis à un stress osmotique par ajout de PEG. Les plantules de 2 provenances (Landes et Maroc) âgées de 3 semaines sont soumises à un stress de -0.45MPa sur une durée de 3 semaines. Les ARN totaux des parties aériennes (tige et appareil foliaire) et des parties racinaires ont été extraits séparément selon la méthode de Chang et al. (1993). Les ADNc et les banques ont été synthétisés en utilisant le ZAP-cDNA synthesis kit (Stratagene, La Jolla, CA, USA) et nommés RN (Root Normal), RS (Root Stress), AN (Aerial part Normal) et AS (Aerial part Stress).
Dans le cadre de cette thèse, nous avons étalé des aliquotes sur milieu Luria Bertani (LB)-agar à 50 µg/ml d'ampicilline et repiqué en milieu liquide LB-ampicilline au format 96 puits un total d'environ 15000 clones. Après une phase de croissance « overnight » sous agitation à 37°C, les plaques sont stockées à -80°C en présence de glycérol à une concentration de 25%.

La duplication des plaques des différentes banques a toujours été réalisée selon les conditions décrites ci-dessus.

C.3.3 Obtention des séquences

Au début de la thèse (en 2002), nous avons mis en place une chaîne semi-automatique de séquençage dans le cadre du pôle Génotypage-Séquençage de la plateforme Génomique fonctionnelle de Bordeaux. Les étapes de séquençage ont été possibles grâce aux outils nouvellement disponibles. Les PCR sur colonies ont été conduites en plaque 384 sur un robocycleur GeneAmp PCR system 9700 (Perkin Elmer, Foster City, CA, USA). L'amplification de l'insert se fait dans 20 µL contenant du tampon 1X, 0.2mM dNTPs, 0.2

µM d'amorce M13F et M13R, 2 mM $MgCl_2$ et 0.3 U *Taq* DNA Polymerase (Invitrogen, Carlsbad, CA, USA). La matrice est apportée en piquant des cultures fraîches. Les paramètres sont : 95°C, 10 min, 95°C, 1 min, 55°C, 1 min, 72°C for 1 min pour 40 cycles et une élongation finale à 72°C pour 10 min. Le produit PCR est purifié avec le système « Montage PCR_{96} Cleanup Kit (Millipore, Billerica, MA, USA) » selon le protocole du fabriquant. Un robot pipeteur Genesis RSP 100 (Tecan, Maennedorf, Switzerland) prend en charge le transfert de liquide des plaques 384 puits vers 96 puits. Les produits PCR purifiés servent de matrice (2.5µL) pour la réaction de séquence utilisant 2.5µL de mixe inclus dans le « « DYEnamic™ ET Dye Terminator Kit (MegaBACE™) » (Amersham Biosciences, Buckinghamshire, UK).

Nous avons choisi un séquençage en simple passe avec une amorce T3 pour séquencer l'extrémité 5' des gènes. L'avantage est d'augmenter la probabilité de trouver des homologies avec des séquences connues et donc d'attribuer une fonction au gène. Le robocycleur GeneAmp system 9700 384 puits est réglé comme suit : 94°C, 20 sec, 50°C, 15 sec, 60°C, 1min et 35 cycles. La totalité du produit de la réaction de séquence est purifiée par centrifugation (1000xg, 5min) sur plaques (Millipore Multiscreen MAHVN45) contenant du Sephadex G-50 (Amersham Biosciences).

Les données brutes de séquences sont détectées sur un séquenceur 96 capillaires MEGABACE™ 1000 (Amersham Biosciences).

C.4 Analyse d'expression

C.4.1 Synthèse des ADNc

Pour la fabrication des sondes, la rétrotranscription en ADNc à partir des ARN totaux « microarrays » a été réalisée selon le protocole du kit de marquage CyScribe Post-Labeling kit (Amersham) en présence d'amino allyl-dUTP. Le kit étant prévu pour 500 ng d'ARN messager et considérant un rapport ARNm/ARNt de 0.5 à 1%, il faut une quantité de l'ordre de 50 à 100 µg d'ARN totaux. Le volume maximum d'ARN totaux concentrés utilisable dans le kit, soit 8 µL, a donc été retenu. Le temps de rétrotranscription de 1.5 heure prévu a été modifié. Après 1.5 h d'incubation à 42°C, nous avons ajouté 1µL de reverse transcriptase, et la réaction d'extension a été prolongée de 1.5 h pour s'assurer d'une rétrotranscription optimale. Les ADNc de l'expérience « stress court » ont été obtenus en utilisant l'enzyme CyScript reverse transcriptase fournie tandis que les ADNc de l'expérience « stress long » ont été synthétisés avec de la Superscript II kit (Invitrogen Life Technologies, Grand Island, NY).

Pour les expériences de PCR quantitative, les ADNc ont été synthétisés avec de l'enzyme Promega ImProm II (Promega, Madison, WI) à ARN totaux constant. Les solutions d'ARN totaux sont ramenées à une concentration d'environ 300 ng/µL puis dosées précisément par spectrophotométrie. Le volume exact nécessaire pour obtenir 1 µg d'ARN totaux est prélevé pour conduire la rétrotranscription selon le protocole du fabricant.

C.4.2 Microréseaux à ADNc

C.4.2.1 Réalisation de l'Unigène

L'Unigène représente l'ensemble non-redondant des gènes contenus dans les banques d'ADNc. Chaque membre peut se définir comme la sélection d'une (ou plusieurs) séquence représentative d'un gène unique. Les possibilités techniques des microarrays utilisées sont de l'ordre de 20000 dépôts par lame. L'objectif était de réduire de façon conséquente le coût et le nombre de sondes à déposer tout en conservant le maximum d'information biologique.
Nous avons exploité les données bioinformatiques disponibles dans nos bases de séquences de pin maritime à la date du 22 septembre 2003. Les résultats d'assemblage (Fig. C-18) semblaient montrer que les séquences d'un même gène appartenaient au même consensus. Nous avons donc considéré que chaque consensus représentait un gène différent et choisi un représentant unique par consensus (cf. C.6.1).

Dans le cas des singletons, chaque séquence doit, par définition, correspondre à une séquence unique. Cependant, certains singletons peuvent appartenir au même gène, mais n'ont pu être regroupés. Nous avons fait le choix d'écarter ces "doublons" de l'Unigène. Pour cela, les singletons ont été comparés sur la base du résultat de la recherche d'homologie obtenue par la stratégie multiBlast du pipeline (cf chapitres suivants) avec un seuil de significativité placé à 10^{-7}. Les singletons ne comportant aucune homologie avec un score inférieur à cette valeur seuil sont intégrés directement à l'Unigène. Dans le cas contraire, les singletons ayant au moins un résultat commun entre eux parmi les 5 premiers résultats (Blast) inférieurs au seuil, sont considérés comme similaires et un seul représentant du groupe est retenu.

Pour choisir un représentant de chaque consensus, nous avons essayé d'homogénéiser la taille des fragments déposés. En effet, à nombre de copies et teneur en GC identique, la fluorescence produite par la cible dépend de la longueur du fragment. Un fragment long incorpore plus de dNTP fluorescents qu'un fragment court. Cependant, nous ne disposions pas de la longueur réelle des inserts correspondant aux EST. Nous avons donc calculé une

longueur théorique à partir des données de bioinformatique. Les séquences ont été obtenues par séquençage d'EST en 5'. Les EST sont des ADNc issus d'ARNm, or les EST correspondent rarement à l'ARNm pleine longueur (Fig. C-15). En effet, la transcription inverse produit un fragment d'ADNc débutant en région 3' de l'ARNm (queue polyA) qui se termine souvent avant la fin de la région 5'. De plus, le fragment d'ADNc peut se casser avant ou pendant son insertion dans le vecteur de clonage. Il résulte de ces événements un fragment partiel de l'ARNm de départ.

Ainsi, après l'assemblage des EST en contigs, permettant d'en déduire des séquences consensus (CN), on peut calculer une longueur théorique pour chaque insert (Fig. C-15):

Longueur Théorique de l'insert = Longueur du Consensus – nb bases entre l'extrémité 5' de l'EST et l'extrémité 5'du Consensus

Dans le cas des contigs, nous avons adopté la stratégie suivante pour choisir le clone représentatif du contig (Fig. C-16) :

- Choix des inserts situés au plus proche de la région 3' du gène, afin de garantir une spécificité accrue. En effet, la différenciation au niveau des membres d'une famille multigénique se fait généralement dans la partie 3' de la séquence. C'est la zone la plus variable.
- Choix des inserts les plus proches d'une longueur théorique définie (borne), garantissant une hybridation correcte des cibles sur les sondes, en termes de spécificité et d'homogénéité de quantité de cibles fixées. La longueur théorique doit également être comprise entre un seuil minimum et un seuil maximum. Nous avons opté pour une longueur théorique d'environ 500 pb (paires de bases), un seuil minimum de 200 pb et un seuil maximum de 800 pb.

Ces deux critères de choix engendrent plusieurs cas possibles (Figure C-16):

1) la longueur théorique de l'insert est comprise entre 500 pb et 650 pb.

Il s'agit du cas le plus intéressant pour notre stratégie (le meilleur intervalle de longueur théorique), c'est pourquoi seulement dans ces cas nous avons choisi l'insert dont l'EST était le plus proche de la queue polyA.

2) la longueur théorique de l'insert est comprise entre 450 pb et 500 pb.

Choix de l'EST avec le plus faible écart (Δ) entre la longueur théorique de l'insert et la valeur 500 pb.

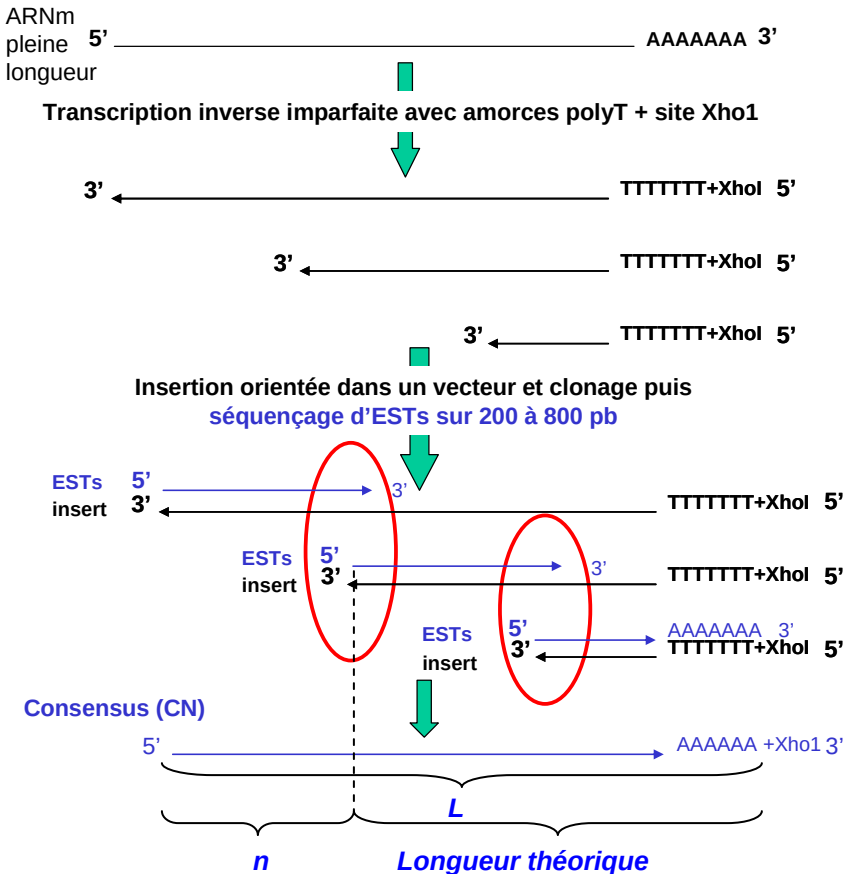


Figure C-15 : Conséquences de la rétrotranscription imparfaite et du clonage sur les ESTs générés en 5'. Les ESTs obtenus (en bleu) ne commencent pas tous à l'extrémité 5' correspondant à l'ARNm pleine longueur. Les zones de recouvrement entre ESTs (entourées) permettent par assemblage bioinformatique de retrouver l'intégralité de la séquence du transcrit (consensus). Nous avons défini la longueur théorique d'un EST comme étant la longueur du consensus (L) moins le nombre de paires de bases (n) entre l'extrémité 5' du consensus et 5' de l'EST.

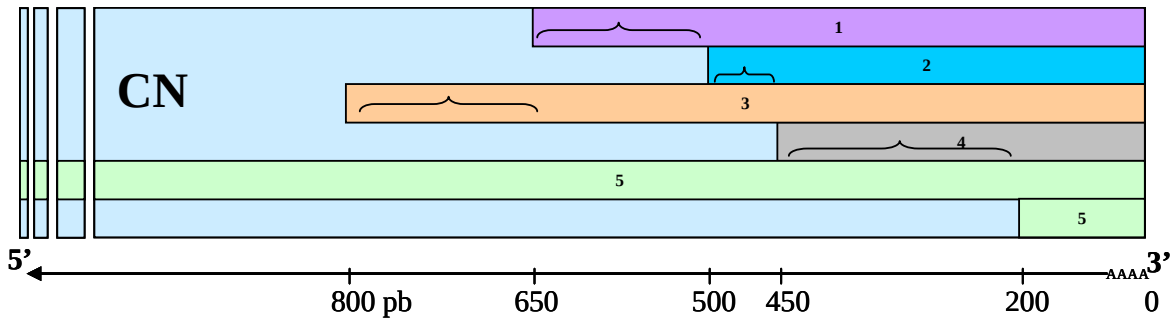


Figure C-16 : Sélection des clones pour l'Unigène. Stratégie adoptée pour le choix du clone représentatif d'un contig sur la base de la longueur théorique de l'insert.
Dans l'ordre de priorité :

1. S'il existe un insert de longueur théorique comprise entre 500 et 650 pb, choix du clone pour lequel l'EST est le plus proche de l'extrémité 3'
2. Insert compris entre 450 et 500 pb, choix du clone avec l'insert le plus proche de 500 pb
3. Insert entre 650 et 800 pb, choix du clone avec l'insert le plus proche de 650 pb
4. Insert entre 200 et 450 pb, choix du clone avec l'insert le plus proche de 450 pb
5. Choix du clone dont l'insert est le plus proche de 500 pb

3) la longueur théorique de l'insert est comprise entre 650 pb et 800 pb.

Choix de l'EST avec le plus faible Δ.

4) la longueur théorique de l'insert est comprise entre 200 pb et 450 pb.

Choix de l'EST avec le plus faible Δ.

5) la longueur théorique de l'insert est inférieur à 200 pb ou supérieur à 800 pb.

Choix de l'EST avec le plus faible Δ.

Nous avons également ajouté à cet Unigène, une sélection de clones issus des banques américaines de *Pinus taeda* dont les inserts codaient des enzymes impliqués dans le métabolisme des sucres et de l'azote et pour lesquels nous ne disposions pas de séquences de pin maritime. L'effet du PEG sur l'expression de ces gènes avait été étudié lors du développement d'embryons somatiques de *Picea glauca* (Stasolla *et al.* 2003). Des contrôles positifs et négatifs issus des études d'expressions sur la formation du bois ont aussi été inclus. Nous avons déposé des séquences de : 4CL (4-coumarate-CoA ligase (E.C. 2.1.1.14)(AL751056)), 5P, Cytochrome P450 (BX255398), Desmine, GP (glycine rich protein (BX000600)), G3PDH (Glycéraldéhyde 3-phosphate déshydrogénase), HSP (Class I LMW heat shock protein (BX000656)), Luciférase (1,3,5,12), Ras, Thioredoxin H (BX000658), Tubuline alpha chain (BX000608), Ai6V(gène d'aiguille différentiellement exprimé lors d'une expérience d'exposition au vent), X42 et X46 (gènes différentiellement exprimés lors d'études sur la formation du bois en ADNc-AFLP).

C.4.2.2 Dépôt des sondes sur lame de verre

Cette étape a été réalisée en partenariat avec la société autrichienne PICME (http://www.picme.at).

La première étape de réarrangement des clones sélectionnés pour l'Unigène a été effectuée au laboratoire en utilisant un robot pipeteur Tecan Genesis 100 adapté pour cette utilisation spécifique. Un double de l'ensemble des clones de pin maritime ainsi que les plaques de l'Unigène ont été envoyé à cet organisme qui est appelé à jouer le rôle de centre de ressource génomique pour la communauté scientifique forestière européenne dans le cadre du réseau d'excellence EVOLTREE. Après extraction des plasmides et amplification par PCR, les produits ont été déposés dans du tampon SSC additionné de Bétaïne sur des lames de verre de marque Corning, modèle GAP.

C.4.2.3 Plan d'expérience

Pour les études d'expression, nous avons décidé de n'étudier que les stades Contrôle, Stress 6h et Stress 48h pour l'expérience « stress court » ; Contrôle long et stress long pour l'expérience « stress long ».

Comme il le sera discuté plus loin, le choix des cibles à co-hybrider sur une même lame détermine la puissance de l'analyse statistique et le nombre de lames à hybrider.

Dans notre cas, pour réduire les variabilités expérimentales et le coût, nous avons choisi de conduire séparément les expériences de stress court et de stress long. Le plan d'expérience est basé sur un dessin en boucle (loop design). L'expérience « stress long » consiste initialement en une boucle simple de 8 lames tandis que l'expérience « stress court » présente 3 boucles imbriquées pour un total de 18 lames (Fig. C-17). Au total, 26 lames sont donc nécessaires pour étudier l'expression de 10 échantillons (2 provenances dans 5 conditions).

C.4.2.4 Synthèse des cibles

Il existe deux méthodes principales pour le marquage des cibles avec des fluorochromes (Cy3 et Cy5 dans le cas des puces à ADNc).

La première, dite marquage direct, utilise l'incorporation de nucléotides liés à un fluorochrome lors de la rétrotranscription. La rétrotranscription et le marquage se font donc en une seule et même étape. L'inconvénient de cette méthode vient de l'encombrement de la molécule fluorescente qui peut gêner l'incorporation du nucléotide lors de la transcription et faire chuter le rendement de la réaction.

La seconde méthode utilise un marquage indirect. Le brin d'ADNc est synthétisé en présence d'un amino allyl-dNTP dont l'encombrement diffère peu de celui des dNTP et par conséquent ne modifie pas le rendement de la transcription. Le marquage est obtenu après dégradation des ARN et purification des ADNc par réaction chimique entre l'amino allyl-dNTP et un NHS-ester du fluorochrome. C'est cette seconde approche que nous avons utilisé en suivant le protocole du kit CyScribe Post-Labeling (cat: RPN5660, Amersham) comprenant les étapes de synthèse des ADNc, dégradation par lyse alcaline des ARN, purification sur colonne, marquage indirect (Allyl-dUTP) et purification finale sur colonne.

Les lames sont hybridées avec une quantité d'ADNc constante. Les cibles ont donc été dosées par spectrophotométrie à 260 nm et pour chaque série d'hybridation, une quantité identique d'ADNc a été pipetée en s'alignant sur l'échantillon contenant la quantité d'ADNc la plus faible. Le volume des cibles prélevées est ramené à une valeur inférieure à 5 µL en utilisant des colonnes Microcon YM-30 (cat : 42410, Millipore).

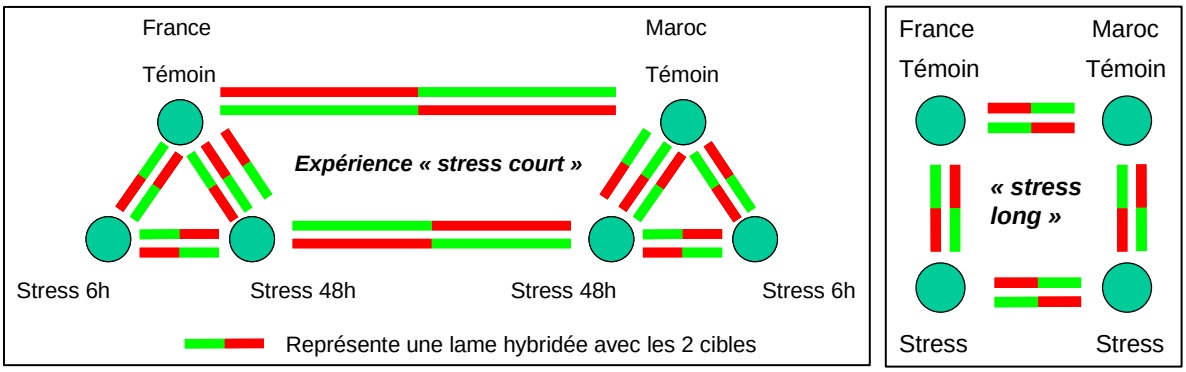


Figure C-17 : Plan d'expérience en « loop design » choisi pour les hybridations des lames. A gauche, l'expérience « stress court » nécessite 18 lames. A droite, 8 lames sont utilisées pour l'expérience « stress long ».

Les deux cibles marquées de fluorochromes différents et devant être hybridées sur un même lame sont mélangées. Le volume de la cible composite est ramené à un volume de 100 µL avec de la solution d'hybridation composée de 58 µL de Formamide, 12 µL de Denhardt 50x, 29 µL de SSC 20x, 2 µL H_20, 2 µL SDS 10% et 1 µL d'ADN de sperme de saumon 10 mg/ml dénaturé 5 minutes à 100°C avant utilisation. La solution cible composite est alors dénaturée 1 minute à 100°C puis placée sur glace 1 minute et préchauffer à 42°C avant hybridation.

C.4.2.5 Préparation, hybridation et lecture des lames

L'ensemble de ces étapes a été réalisé avec les équipements du pôle Transcriptome de la Plateforme Génomique Fonctionnelle de Bordeaux installée pour partie au sein de l'Institut de Biologie Végétale Intégrative IFR 103 (IBVI) sur le site de l'INRA de Villenave d'Ornon.

Pour une meilleure conservation, les lames sont stockées à température ambiante sans autre traitement après dépôt que l'exposition aux UV pour fixer les sondes sur le support. Pour de meilleurs résultats, il est préférable d'attendre au minimum 2 semaines (3 mois dans notre cas) avant d'effectuer un traitement de préparation (Post-processing) des lames qui doivent dans ce cas être hybridées rapidement (http://microarrays.nki.nl/download/protocols/Post-Processing%20cDNA%20arrays.pdf).

Les lames sont préparées manuellement par trempage dans différent bains de solutions filtrées à 0.45µm et à température ambiante : 2 lavages successifs dans des solutions de SDS à 0.2% pendant 2 minutes, 2 lavages successifs dans de l'eau miliQ pendant 2 minutes, 2 minutes dans de l'eau miliQ bouillante, 5 minutes de séchage en agitant manuellement à température ambiante, 3 lavages successifs dans des solutions de SDS à 0.2% pendant 1 minute, 1 minute dans un bain d'eau miliQ, 10 secondes dans de l'eau miliQ bouillante, séchage pendant 5 minutes et élimination éventuelle de l'humidité résiduelle à l'azote.

Les lames ainsi préparées sont placées dans une station d'hybridation automatique Tecan HS 4800, ce qui permet de réduire les variations entre les différentes expériences d'hybridations. L'ensemble des solutions utilisées ont été filtrées à 0.45µm. La station gère les températures, l'agitation des cibles complexes fluorescentes, la fluidique et le séchage des lames. A la sortie de la station, les lames sont prêtes à être scannées.

Le programme adopté commence par un lavage à 42°C avec une solution de SSC 1X et SDS 0.1X sous agitation pendant 30s. Les solutions cibles préchauffées à 42°C sont introduites dans les chambres d'hybridation dont le volume est de 100µL. L'étape d'hybridation se

déroule à 42°C sous agitation moyenne pendant 16h. Les lavages se font ensuite à 30°C et commencent par une solution SSC 1X et SDS 0.1X pendant 1 min et 30s d'agitation, suivi d'une solution SSC 0.1X et SDS 0.1X pendant 1 min et 30s d'agitation, puis un lavage SSC 0.1X pendant 30s et 30s d'agitation. Un séchage final à l'azote est effectué à 30°C pendant 30s.

Les lames sont conservées à l'abri de la lumière et scannées immédiatement après hybridation sur un scanneur Axon Instruments Genepix 4000 B. La lecture a été réalisée avec un pas de 5µm. La détection des spots s'est faite à l'aide du logiciel Genepix Pro qui permet d'aligner les grilles de « spottage », d'ajuster automatiquement et manuellement la taille des zones de détection, de labelliser les spots de mauvaise qualité et d'extraire les données numériques de fluorescence en générant des fichiers exportables vers un tableur.

C.4.2.6 Traitement des données

La lecture des lames génère des fichiers de taille conséquente, difficiles à exploiter directement sous des tableurs classiques. Pour effectuer l'ensemble des étapes de traitement et manipulation des données il m'a fallu générer un ensemble de scripts sous langage Perl (ActivePerl, http://www.activesate.com). Les analyses statistiques ont été conduites sous l'environnement R de calculs statistiques (http://www.r-project.org/).

Les données brutes de fluorescence sont filtrées, nettoyées et formatées pour permettre l'application de la fonction « rlowess » de la bibliothèque « MAANOVA » (Wu 2003).

Cette fonction va normaliser les données lame par lame par une transformation « LOWESS ou LOESS » (Draghici 2003) en tenant compte de l'intensité et de la position des spots. Elle a pour effet de redresser le nuage de point d'un « scatter plot » représentant les ratios du signal Rouge/Vert en fonction de l'intensité Rouge*Vert.

Les données normalisées sont récupérées et les gènes filtrés sur le nombre de données manquantes. Considérant le nombre de données manquantes pour chaque facteur : condition, provenance et fluorochrome, seuls les gènes ayant moins de 1/3 de données manquantes quelque soit le facteur, sont retenus. Les données sont alors réorganisées et transformées en log de base 2 pour l'analyse statistique.

Nous adoptons une approche similaire à celle développée par le groupe de Gary Churchill (http://www.jax.org/staff/churchill/labsite/software/anova/) et présenté dans la bibliothèque MAANOVA (Wu 2003). Deux analyses de variance consécutives sont appliquées.

Le premier modèle effectue une normalisation de l'ensemble des données. Dans ce cas, les gènes sont considérés comme des répétitions :

Modèle C-1 : $\mathrm{Y}_{ijk\lg} = \mu + \mathrm{A}_i + D_j + \mathrm{A}D_{ij} + \mathrm{P}t_k + Ct_l + \mathrm{P}tCt_{kl} + \varepsilon_{ijk\lg}$

Ce modèle permet de fixer les effets globaux techniques de lame (A), du fluorochrome (D) et d'interaction (AD), mais aussi des effets techniques confondus à un effet global Provenance (Pt), Condition (Ct) et interaction (PtCt). On retire donc l'influence de la Provenance et de la Condition sur l'ensemble des gènes. Cette approche permet d'éliminer un effet global de ces facteurs sur l'ensemble des gènes afin de générer des données corrigées en récupérant l'erreur résiduelle (ε) et la moyenne (μ). Ces données corrigées vont nous permettre d'identifier les effets significatifs dans un second modèle.

Le second modèle consiste cette fois en une analyse gène par gène (g) en appliquant le modèle :

Modèle C-2 : $\mu + \varepsilon_{ijk\lg} = \mathrm{A}_i + D_j + \mathrm{A}D_{ij} + R_r + \mathrm{P}_k + C_l + \mathrm{P}C_{kl} + e_{ijklr}$

Dans ce modèle, R symbolise les répétitions de dépôts sur la lame (au nombre de 2).
Un des problèmes majeur de ce type d'analyse de données d'expression provient de la multiplicité des tests. Le test d'analyse de variance est en effet répété autant de fois qu'il y a de gènes (dépôts d'ADNc distincts). Le risque α fixé pour un test unique peut amener à un nombre de faux positifs important lorsque le test est répéter des milliers de fois. Les valeurs des tests générées par les ANOVA ont donc été corrigées par la méthode « FDR » (False Discovery Rate) de Benjamini et Hochberg incluse dans la fonction « p.adjust » de R (Benjamini and Hochberg 1995). Le FDR se définit comme la proportion attendue d'hypothèses nulles – ici H0 « il n'y a pas de différences entre les niveaux des facteurs » - rejetées à tort (faux positifs) parmi les hypothèses rejetées. Cette méthode contrôle le taux de faux positifs tout en conservant la proportion attendue de faux positifs (carence de type I) selon l'hypothèse nulle.
Les résultats sont exportés vers un tableur Excel pour réaliser des tris sur les significativités des tests et la part de variation expliquée par le modèle afin de déterminer les gènes régulés par les facteurs Provenance, Condition et Provenance x Condition.

C.4.2.7 Identification des gènes différentiellement exprimés

Les données de l'expérience « stress court » et celles de l'expérience « stress long » sont traitées séparément. La détection des gènes différentiellement exprimés est réalisée en tenant compte des valeurs des probabilités associées au test et de la part de variation expliquée par le modèle et par les facteurs provenance (P), condition (C) et interaction (I). La première

étape de filtrage consiste à retenir uniquement les gènes pour lesquels un effet significatif d'au moins un des trois facteurs est détecté. Cela est obtenu classiquement en fixant arbitrairement un seuil de significativité en dessous duquel l'effet est considéré significatif. Le choix de ce seuil, fixé à 0.01, sera discuté dans la partie « Résultats et Discussion ». Le second filtrage porte sur la part de variabilité expliquée représentée par le coefficient de détermination « r^2 ». La part de variabilité expliquée par les trois facteurs cités plus haut doit être au minimum de 40 % ou au moins un des facteurs explique plus de 33 % de la variabilité. Cela permet de s'assurer que les gènes détectés présentent une part non négligeable de la variabilité expliquée par les facteurs d'intérêt. Comme il le sera discuté lors de la présentation des résultats, ces valeurs ont été fixées en fonction du jeu de données analysé. Aucun filtrage n'a été appliqué concernant un niveau de régulation minimum étant donné les faibles amplitudes de variations observées sur les données normalisées.

C.4.2.8 Agrégation des données (Clustering)

Les données d'expression d'un gène (g) pour une série d'échantillons (n) peuvent être représentées par des vecteurs à n dimensions : $g_g = (e_1, e_2, \ldots, e_n)$

Le clustering se définit comme un processus de regroupement d'entités similaires reposant sur la mesure de distances entre vecteurs à n dimension. Un algorithme regroupe en différentes classes de similarité les entités sur la base de leurs distances respectives. Cet algorithme est donc très sensible à la distance utilisée. Les classes peuvent alors être organisées sous la forme de dendrogramme suivant différentes méthodes comme la classification hiérarchique qui est agglomérative, les k-means qui vont regrouper les données selon un nombre de classes défini par l'utilisateur et le « Self-Organizing Map » (SOM) qui est une méthode non supervisée, basée sur les réseaux neuronaux.

De nombreuses mesures peuvent être utilisées pour évaluer cette similarité. S. Draghci (2003) décrit et compare de nombreuses distances métrique dont les plus courantes sont la distance euclidienne et le coefficient de corrélation de Pearson.

- La mesure de distance euclidienne est la distance classique entre 2 vecteurs à n dimensions basée sur le théorème de Pythagore.
- La distance de corrélation de Pearson mesure la corrélation entre 2 vecteurs à n dimension. Si les profils des gènes sont clustérisés, ceux ayant des niveaux d'expression entre échantillons évoluant dans le même sens seront similaires mais l'influence du niveau d'intensité ne sera pas prise en compte. Un gène fortement exprimé quelque soit les conditions sera similaire à un gène faiblement exprimé tant

que les variations entre échantillon restent semblables. Cependant la méthode est sensible aux entités sortant de la distribution moyenne (« outliers »).

Le choix de la méthode de calcul de la distance permet donc d'orienter le regroupement des données en fonction des caractéristiques que l'on veut mettre en évidence. Pour l'analyse de nos données, nous avons choisi d'utiliser la distance euclidienne.

Le clustering peut être réalisé sur l'ensemble des données. Dans notre cas, nous nous intéresserons uniquement au regroupement des gènes dont les variations d'expression auront été détectées significatives par l'ANOVA.

Nous avons pris comme valeur pour chaque échantillon - c'est-à-dire pour une provenance et une condition donnée -, la moyenne des données corrigées issues du modèle d'ANOVA pour cet échantillon. Pour chaque gène, les valeurs des échantillons sont normalisées en les divisant par la moyenne des données pour ce gène. Cela permet d'éviter un regroupement sur le niveau d'expression global d'un gène tout en conservant la possibilité de regroupement sur l'amplitude de la régulation et sur la norme de réaction. Le calcul d'un ratio d'expression par rapport à un échantillon contrôle normalisera également le niveau d'expression mais sera plus sensible à la valeur du contrôle, ce qui risque de regrouper les gènes en fonction du niveau d'expression du contrôle.

Pour réaliser le clustering, nous avons utilisé le logiciel en ligne EPCLUST (http://ep.ebi.ac.uk/EP/EPCLUST/). Les valeurs centrées sur la moyenne du gène sont passées en log 2 pour obtenir une représentation graphique avec des valeurs autour de 0. Les données sont alors regroupées par similarité des gènes ou des échantillons en utilisant une distance euclidienne et la méthode de classification hiérarchique UPGMA de calcul de la distance moyenne. Nous avons également utilisé le module intégré SOTArray disponible au sein de la suite d'analyse GEPAS (http://gepas.bioinfo.cnio.es/cgi-bin/sotarray) toujours en distance euclidienne et en conservant le seuil de variabilité proposé par défaut pour générer des classes de similarité plus synthétiques.

Une approche complémentaire a été utilisée pour regrouper les données par similarité des profils d'expression en utilisant le module Click de l'outil en ligne Expander de l'université de Tel-Aviv (http://www.cs.tau.ac.il/~rshamir/expander/expander.html). L'algorithme utilise des techniques statistiques et graphiques sans a priori sur la structure ou le nombre de groupes

attendus (Sharan *et al.* 2003). Un paramètre d'homogénéité intra-classe est modulé pour affiner la composition des groupes.

C.4.3 PCR quantitative en temps réel

Le principe de base de cette technique réside en la détection et l'enregistrement de la fluorescence au cours de la réaction PCR ; ce qui permet de calculer la quantité initiale de matrice dans différents échantillons. Cela est possible en comptant le nombre de cycles PCR nécessaires pour atteindre un seuil donné de fluorescence. Dans notre cas, la fluorescence, produite par du SybrGreen™, est proportionnelle à la quantité d'ADN double brin. Le nombre de cycles (Ct) nécessaires pour atteindre le seuil sera d'autant plus faible que le nombre de copies du transcrit étudié sera élevé dans l'échantillon.

L'utilisation de cette technique pour l'estimation de la quantité de transcrits présents dans les échantillons englobe deux approches principales, la quantification absolue et la quantification relative. La première permet d'estimer le nombre de copies d'un transcrit dans un échantillon en se basant sur des courbes de calibration obtenues à partir d'échantillons contenant un nombre de copies connu du transcrit ou ADNc étudié. La seconde, utilisée dans notre cas, permet de déterminer des changements du niveau de transcrits d'un gène entre de multiples échantillons et de l'exprimer par rapport au transcrit d'un gène de référence. En effet, l'utilisation d'un contrôle endogène permet de corriger les variations inter échantillons dues entre autre à des erreurs de quantification des ARNt. Le choix de ce gène de référence est délicat, toute variation de son niveau d'expression atténuera voire masquera les changements réels du transcrit étudié et génèrera des artéfacts d'expression (Bustin 2000).

C.4.3.1 Experimentation

Pour prendre en compte les variations expérimentales lors de la synthèse d'ADNc et réduire les variations intra échantillon, 3 synthèses indépendantes ont été réalisées pour chaque échantillon La méthode adoptée est similaire à celle décrite en C.4.1 avec utilisation d'oligo dT. Les ADNc sont dilués au 10ème avant utilisation. Les réactions PCR sont conduites sur un robocycleur Chromo4 (MJ Geneworks, Inc) avec les paramètres suivant : 95°C pendant 3 min, puis 40 cycles à 95°C pendant 15s et 60°C pendant 45s et une courbe de fusion finale de 65°C à 95°C par 0.5°C d'intervalle. Le volume réactionnel de 20 µL contenait du « iQ™ SYBR Green Supermix » (Bio-Rad Laboratories, Inc, CA) à concentration 1X, les amorces du gène d'intérêt à 150nM et 3µL d'ADNc. La fluorescence est enregistrée à la fin de chaque

cycle par le logiciel Opticon Monitor 3.1.32 (Bio-Rad Laboratories, Inc, CA) qui permet également de déterminer la valeur seuil de Ct et l'exportation des données de fluorescence.

L'estimation de l'efficacité de la réaction est faite selon la méthode classique des courbes de dilution et du calcul de la pente (http://www.gene-quantification.de/). Pour chaque rétrotranscription, un « pool » des échantillons est utilisé dilué au 1/10^{éme}, 1/100^{éme}, 1/1000^{éme} et 1/10000^{éme}. La représentation graphique du nombre de cycles (Ct) en fonction du logarithme de base 10 de la quantité initiale permet de déterminer une droite de régression dont la pente nous informe sur l'efficacité de la réaction en pourcentage selon la formule : $10^{-((1/pente)-1)}$.

Les données brutes sont corrigées en choisissant de leur soustraire la valeur de la ligne de base déterminée par la moyenne de la fluorescence comprise entre le 3ème cycle et le 14-16ème selon le gène étudié. La valeur du seuil de fluorescence est déterminée manuellement en se plaçant dans la partie linéaire de la courbe du logarithme de base 10 de fluorescence en fonction du nombre de cycles.

C.4.3.2 Choix des amorces

Le choix des amorces est particulièrement important en PCR quantitative. Dans notre cas, nous ne connaissons pas a priori la position des introns et exons. Il est donc indispensable de travailler sur des ARN traités à la DNAse pour éviter toute amplification sur l'ADN génomique. Il est préférable de rechercher à dessiner les amorces dans la région 3' des gènes et de lancer une recherche d'homologie sur le fragment étudié pour déterminer les régions les plus spécifiques. Le risque est de co-amplifier différents membres de familles multigéniques régulées de manière différentes.

Pour chaque gène étudié, les amorces ont été définies à l'aide du logiciel en ligne Primer3 http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi.

Les principales conditions à respecter sont :

- Amplification de fragments courts, taille recommandée 100-150pb
- Teneur en G/C basse préférable, tolérée entre 20% et 80% mais 50-60% recommandée
- Température de fusion (Tm) des paires d'amorces proche de 60°C
- Différence entre Tm ≤ 0.5°C si possible
- Pas de formation de dimères
- Pas plus de 4 nucléotides identiques consécutifs, spécialement les G
- Pas plus de 2 G ou C dans les 5 derniers nucléotides (extrémité 3')

Dans cette étude, nous avions choisi les amorces afin d'amplifier des fragments d'une longueur comprise entre 250 et 350 pb ce qui nous laissait éventuellement la possibilité d'un marquage radioactif et leur utilisation pour des études d'hybridation sur filtres.

C.4.3.3 Choix des gènes de référence

Parmi les gènes les plus utilisés pour normaliser les données d'expression que l'on rencontre dans la littérature, on peut citer l'isoforme beta de l'actine, la Glyceraldehyde-3-phosphate déshydrogénase, l'Ubiquitine C ou certaines protéines ribosomales comme la 60S P0, L13a, L10 ou 28S et 18S (Vandesompele *et al.* 2002). Le gène idéal serait exprimé à un niveau constant dans tous les tissus d'un organisme, à tous les stades de développement, et ne serait pas affecté par les traitements appliqués. En pratique, un gène adapté dans un cas n'est pas forcément transférable à une autre expérimentation.

Pour notre étude, nous avons recherché dans les données d'expression des microarrays un gène dont l'expression dans les racines ne variait statistiquement pas quelque soient les conditions et la provenance. Après sélection de 3 gènes, un couple d'amorce a permis de détecter par PCR quantitative des niveaux d'expression parfaitement stables dans toutes les conditions. Il s'agit d'un ARNm codant pour une protéine ribosomale 40S S27.

C.4.3.4 Traitement des données de fluorescence

Les valeurs sont exportées vers une macro Excel nommée GENEX – Gene Expression Macro (v1.10–2004, Bio-Rad) accessible à l'adresse http://www.gene-quantification.de/download.html#genex.

La méthode utilisée dans cette macro est dérivée de celle décrite par Vandesompele (Vandesompele *et al.* 2002) et très proche de celle proposée par Pfaffl (Pfaffl 2001). Elle permet de prendre en compte les données d'un gène de référence ainsi que l'efficacité de la réaction contrairement à la méthode classique du ΔΔCt où :

$$\Delta\Delta Ct = \Delta Ct_{\text{échantillon 1}} - \Delta Ct_{\text{échantillon 2}}$$

avec $\Delta Ct = Ct_{\text{gène}} - Ct_{\text{gène référence}}$ pour un échantillon donné.

Le gène de référence, d'expression supposée constante dans l'ensemble des échantillons étudiés, a pour but de corriger les variations expérimentales reliées à l'échantillon.

Dans le cas de la macro, une quantité relative (QR) pour un gène (j) est calculée à partir des Ct moyens de chaque échantillon (i) et d'un échantillon contrôle pour tenir compte de l'efficacité :

$$QR_{ij} = \text{Efficacité}_{j}^{\left(Ct_j \text{ de l'échantillon contrôle} - Ct_{ij} \text{ de l'échantillon X}\right)}$$

L'échantillon contrôle peut être choisi arbitrairement ou l'échantillon avec le Ct moyen le plus faible.

La normalisation par un ou plusieurs gènes de référence (k) est effectuée en calculant un facteur de normalisation pour chaque échantillon (i) :

$$\text{Facteur de Normalisation}_i = \left(QR_{i1} \times QR_{i2} \times \ldots \times QR_{ik}\right)^{\left(\frac{1}{k}\right)}$$

qui est la moyenne géométrique des quantités relative des (k) différents gènes de référence pour l'échantillon (i).

Le niveau d'expression du gène (j) pour l'échantillon (i) s'exprime alors sous la forme :

$$\text{Niveau d'Expression}_{ij} = \frac{\text{Quantité relative}_{ij}}{\text{Facteur de Normalisation}_i}$$

Les données aberrantes et isolées, c'est-à-dire ayant une valeur de Ct totalement différente des deux autres répétitions et des autres échantillons, ont été omises.

La macro nous donne alors la possibilité de représenter graphiquement le niveau d'expression des différents échantillons et l'écart-type associé aux 3 synthèses d'ADNc (Fig. Ax-3).

C.5 Analyses Protéiques

Les tissus analysés correspondent aux parties racinaires de l'expérience « Stress Long » et sont identiques à ceux utilisés pour les analyses transcriptomiques. L'étude du protéome a été réalisée par Manon Moreau (formateur : Céline Lalanne) dans le cadre de son Master.

C.5.1 Extraction des protéines et Electrophorèse bidimensionnelle

Le protocole utilisé est inspiré de celui de (Damerval *et al.* 1986), disponible à l'adresse http://www.pierroton.inra.fr/genetics/2D/Proteomevert/Protocoles/protobidinew.pdf .

Pour chaque échantillon de l'expérience « Stress Long », 4 répétitions ont été effectuées, soit un total de 16 gels.

C.5.2 Analyse des gels

Les gels colorés au bleu de coomassie sont scannés et analysés en utilisant la suite logicielle et matérielle Amersham Biosciences (Labscan et ImageMaster Platinum v5.0) (Gion *et al.* 2005). Les paramètres utilisés lors de l'acquisition sont : mode transparent, filtre rouge et sensibilité 300 dpi. Après détection automatique des spots, les gels sont repris manuellement et nettoyés pour éliminer les spots artéfactuels. Seuls les spots détectés sur au moins deux gels du même échantillon sont conservés. Le volume de chaque spot est calculé sur la base de sa surface et des niveaux de gris enregistrés. Un gel de référence créé à partir de l'information de l'ensemble des gels permet de construire une carte générale de tous les spots détectés. Les gels sont comparés à cette carte de référence afin de donner la même identité au même spot présent sur plusieurs gels.

C.5.3 Analyse quantitative

Une analyse de variance (ANOVA) avec les facteurs Provenance (*P*) et Condition (*C*) comme variable dépendante a été réalisée selon le modèle :

$$\mathrm{Y}_{ijk} = \mu + P_i + C_j + P_i C_j + \varepsilon_{ijk}$$

Ce modèle permet d'identifier les spots protéiques significativement spécifiques d'une provenance, d'une condition ou de l'interaction entre provenances et conditions. Le seuil de significativité choisi a été placé à 0.001 pour tenir compte de la multiplicité des tests (120 spots détectés).

C.5.4 Choix des spots à séquencer

Etant donné le nombre conséquent de spots présentant au moins un effet significatif, seuls ceux présentant un effet d'interaction ont été prélevés dans un premier temps. Ces spots peuvent correspondre à des stratégies différentes de réponse au stress entre les deux provenances. Dans un second temps, quelques spots montrant uniquement un effet provenance ont été prélevés ainsi que des spots situés relativement proches les uns des autres sur gel (variation qualitative de type présence/absence) et présentant au moins un effet provenance, dans le but d'identifier des formes allèliques d'une même protéine. Quelques spots montrant un effet traitement ont également été prélevés dans le but d'identifier des

mécanismes généraux de réponse à la sécheresse, indépendamment de la provenance. Enfin, des spots spécifiques des marocains stressés ont été retenus.

C.5.5 Séquençage et identification des protéines

Digestion trypsique et Spectrométrie de masse

Les spots prélevés sont lavés par une solution de bicarbonate d'ammonium 25 mM et d'acétonitrile (ACN) 50% pendant 2H. La solution est alors remplacée par de l'ACN et incubée 10 min pour déshydrater les spots. Après 5 min de séchage au speedvac, les protéines sont digérées une nuit à 37°C par la trypsine, qui coupe au niveau des lysines et des arginines, en présence d'HCl 1 mM et de bicarbonate d'ammonium 50 mM.

Les peptides hydrophiles sont extrait après ajout de 10 µL de bicarbonate d'ammonium 50 mM, incubation 10 min sous agitation, centrifugation 30 s, et récupération du surnageant.

Les peptides hydrophobes sont récupérés par 3 extractions successives commençant par une incubation de 10 min des spots précédents en présence de 10 µL de tampon d'extraction (eau/ACN/Ac acétique 47.5/47.5/5) puis récupération du surnageant.

Les différentes extractions d'un même spot sont réunies puis le volume est ramené à 25 µL environ par évaporation au speedvac à température ambiante avant ajout de 0.006 µL d'acide acétique par µL d'extrait.

La partie séquençage protéique a pu être effectuée en utilisant les équipements disponibles au sein du pôle protéomique de la Plate-forme de Génomique Fonctionnelle de Bordeaux

Les mélanges peptidiques sont analysés par spectrométrie de masse en tandem comme décrit dans l'article de Jean-Marc Gion *et al.* (2005) sur le protéome de la formation du bois chez le pin maritime. Le principe de la technique utilisée se base sur la séparation des peptides par chromatographie liquide en phase inverse en fonction de leur hydrophobicité puis analyse des peptides ionisés par un spectromètre de masse produisant un spectre MS. Certains de ces peptides sont sélectionnés, fragmentés en sous peptides et le plus souvent en acides aminés et ré-analysés pour obtenir des spectres MS/MS. On dispose donc pour chaque spot protéique d'un nombre variable de peptides de séquence connue avec une précision variable. Une analyse informatique est alors nécessaire pour identifier une protéine à partir des séquences peptidiques disponibles. Les spectres MS/MS des peptides sont comparés de façon automatique (programmes Turbo Sequest et Sequest Browser) aux spectres théoriques générés par digestion *in silico* d'une banque d'EST traduits de pin et de la base protéique SwissProt.

C.6 Analyses Bioinformatiques

C.6.1 Analyse et Assemblage des séquences

Les électrophorégrammes ont été analysés par le logiciel « Phred » qui détecte la séquence en acides nucléiques et génère un score de qualité global de la séquence et un score qualité individuel par base. Un outil de comparaison rapide de deux lots de séquence, inclus dans un script Perl écrit par J-M. Frigerio, permet d'enlever les séquences du vecteur pour éditer rapidement et manuellement chaque séquence et obtenir un score Phred 20 de qualité global de la séquence supérieur à 85%. Dans une séquence, Phred attribue à chaque pic de l'électrophorégrammes une probabilité que la base nucléotidique détectée soit réellement la bonne. Le score est généralement un nombre entre 1 et 40 où le nombre représente la puissance de 10. Un score Phred 20 signifie qu'une base n'est pas la bonne une fois sur 100 (10^2) et un score Phred moyen de 20 signifie que la séquence est mauvaise une fois sur 100.

Le principe de l'assemblage bioinformatique d'EST est basé sur le fait que nous ne disposons que de courts fragments de séquences du transcrit d'un gène. En effet, comme nous l'avons déjà évoqué, si l'on considère les transcrits d'un gène disponibles après extraction, tous ne seront pas pleine longueur et donneront donc des ADNc partiels. De plus, la transcription inverse utilise comme amorces des polyT qui vont se fixer sur la queue polyA des ARNm, mais la réaction n'est pas parfaite et produit des ADNc incomplets dans la partie 5' (Fig. C-15). Etant donné que nous avons choisi un séquençage d'EST en 5', nous obtiendrons en général les 600-700 premières paires de bases en 5'. Il en résulte une série de fragment de séquences réparties le long du gène. L'objectif de l'assemblage est de rechercher les homologies de séquences entre EST pour reconstituer la séquence complète du gène à partir des fragments de séquences. Cela est rendu possible de proche en proche grâce aux zones de recouvrements des différents EST. Si des séquences se recouvrent en 5' et d'autres en 3' mais qu'aucune séquence n'est disponible pour la partie centrale, alors deux groupes d'EST seront générés pour un même gène. Plus le nombre de séquences disponibles pour un gène est élevé, plus on a de chance de reconstituer l'intégralité du gène. Les séquences qui n'ont pas de régions suffisamment spécifiques homologues à d'autres EST donneront des singletons. Pour éviter de regrouper les EST sur la base de régions de faibles complexité (ex : régions répétées comme les microsatellites), il est indispensable d'effectuer un masquage de ces régions.

Les séquences de bonne qualité sont injectées dans une suite logicielle nommée EPA, développée et testée en collaboration avec le CBiB (Centre de Bioinformatique de Bordeaux).

Cet outil est construit autour d'une base de donnée MySQL et du module StackPack™ (Electric Genetics www.egenetics.com) qui permet un assemblage rapide et un alignement des séquences. Les séquences sont analysées par le programme RepeatMasker (http://www.repeatmasker.org/) pour masquer les zones de faible complexité comme par exemple les queues polyA ou les microsatellites. Les séquences sont tout d'abord regroupées en **cluster** (CL) par le module D2cluster (Burke *et al.* 1999) qui découpe les séquences en motifs hexanucléotidiques et recherche une suite de motifs identiques entre les séquences (Fig. C-18). Ce premier regroupement permet un assemblage plus fin en **contigs** (CT) par le programme Phrap (http://www.phrap.org/). Des séquences **consensus** (CN) sont alors générées par le module CRAW (Burke *et al.* 1998) qui va analyser les alignements produits par Phrap. Seul le consensus primaire (CN primaire) est conservé pour l'étape d'annotation.
Les EST qui n'ont pu être rapproché d'un contig sont appelés **singletons**.
A titre de comparaison, la stratégie adoptée par le TIGR (http://www.tigr.org/tgi/faq2.shtml) commence par un contrôle et nettoyage de la présence de séquences de vecteur, *E. coli* et/ou poly A/T. Les séquences de taille inférieure à 100 pb ou comportant plus de 3% de N ne sont pas prises en compte. Pour l'assemblage, les EST sont clustérisés sur la base d'un minimum de 40 bases homologues, plus de 94 % d'identité dans la zone de recouvrement et une région d'un maximum de 30 paires de bases n'alignant pas avec les autres EST clustérisés. Le programme utilisé pour la création des séquences TC (Tentative Consensus) est « Paracel Transcript Assembler »

La structure de la base de données donne la possibilité de travailler sur des lots de séquences organisés en projets. Dans notre cas, un premier projet regroupe l'ensemble des séquences de la banque « xylème » (projet « Pinus Pinaster xylem EST ») et un second projet regroupe celles de racines (projet « Pinus Pinaster root EST »). Puis un projet global rassemble l'intégralité des séquences (« Pinus Pinaster Total »).

Avec le CBiB nous avons développé une interface web disponible pour les projets publics à l'adresse (http://cbi.labri.fr/outils/SPAM/COMPLETE/). Elle offre un accès commun aux requêtes et résultats d'homologie (Blast), aux informations de séquences des EST et consensus, ainsi qu'aux annotations fonctionnelles comme KEGG (Kyoto Encyclopedia of Genes and Genomes), FunCat (The Functional Catalogue) et GO (Gene Ontology).

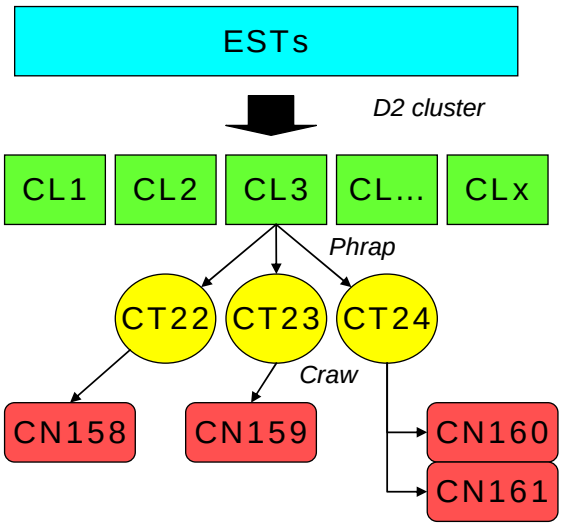


Figure C-18 : Différents niveaux de regroupement des ESTs lors de l'assemblage bioinformatique. Les ESTs sont regroupés en « clusters » (CL) par le programme D2 cluster. L'alignement des ESTs de chaque CL est analysé par le programme Phrap qui génère des « contigs » (CT). Le programme Craw va déterminer une séquence consensus (CN) à partir des ESTs de chaque contig. Si une seule séquence consensus ne peut être déduite, des consensus alternatif sont générés.

Les homologies de séquence des EST et des séquences consensus ont été recherchées dans les différentes bases nucléiques et protéiques publiques. Une stratégie en cascade a été utilisée pour diminuer les temps de calcul. Une première requête BlastX contre la base SwissProt est testée et les meilleurs résultats sont conservés. Si aucune homologie n'est trouvée avec un indice de confiance (E-Value) $\leq 10^{-5}$, la même requête est testée su la base TrEMBL et si toujours aucune homologie n'est trouvée avec une E-Value $\leq 10^{-5}$, le résultat de la requête Blastn contre la base EMBL est retenu.

C.6.2 Annotation des Consensus et Singletons

Chaque singleton et consensus a été annoté en fonction du résultat des recherches d'homologie (BLAST). La concordance de l'annotation choisie pour la séquence consensus avec les homologies trouvées pour les EST composant le contig est vérifiée. Si besoin, une nouvelle recherche d'homologie sur d'autres bases de donné est lancée. Elle doit permettre de déterminer si le conflit d'annotation observé entre l'annotation attribuée à partir du consensus et celle de l'EST vient du défaut de séquence informative dans l'EST par rapport au consensus ou bien de l'assemblage au sein d'un même consensus d'EST correspondant à des gènes différents.

Une fois les EST annotées, il est intéressant de les classer par catégories fonctionnelles. Une annotation fonctionnelle a été entrée chaque fois que possible dans la base en suivant la classification de « Gene Ontology ».

Par ailleurs, nous avons regroupé les informations disponibles sur les séquences du projet LIGNOME dans un tableur Excel pour affiner l'annotation et leur attribuer une catégorie fonctionnelle. Ce travail s'est donc focalisé uniquement sur les informations disponibles pour les tissus racinaires, objet d'étude de cette thèse.

Une même enzyme peut porter des appellations synonymes. Pour éviter les confusions et considérer deux gènes comme différents et non comme membre d'une même famille multigénique, nous avons systématiquement recherché la dénomination officielle des enzymes sur le site Expasy (http://www.expasy.org)». Nous avons choisi de suivre la classification fonctionnelle utilisée par le MIPS (Munich Information center for Protein Sequences) mais sans utiliser les classes « cellular biogenesis », « ionic homeostasis » et « development » pour limiter des erreurs d'affectation de notre part. Nous avons ainsi réparti les EST dans les catégories suivantes : croissance cellulaire/division, organisation cellulaire, énergie,

métabolisme, communication / transduction du signal, transport, trafic intracellulaire, transcription, synthèse de protéines, devenir des protéines, réponse au stress, non classée.

C.6.3 Niveau d'expression *in silico* des EST (Northern electronique)

Le séquençage systématique d'EST de banques non normalisées permet d'accéder à une information sur l'expression des gènes séquencés. En effet, en comptant le nombre de fois que l'ARNm d'un gène donné est séquencé dans une banque, il est possible d'estimer le niveau d'expression de ce gène par rapport à la totalité des EST séquencés. Si plus d'une banque est séquencée, il est alors théoriquement possible de comparer les niveaux d'expression d'un gène entre les différentes banques. Dans notre cas, nous disposions d'une banque de racines témoin et d'une banque de racines stressée. Une analyse statistique de l'abondance des transcrits des différents gènes est donc possible. Ce Northern électronique, encore appelé Northern *in silico*, e-Northern ou digital Northern, permet de détecter les gènes significativement différentiellement exprimés entre les deux conditions. Nous avons utilisé la suite logicielle « IDEG6 » (Identification Differentially Expressed Genes 6 test statistics) spécifiquement développée pour ce type d'analyses et accessible en ligne à l'adresse http://telethon.bio.unipd.it/bioinfo/IDEG6/ (Romualdi *et al.* 2003). Cette suite tient compte du nombre total de séquences de chaque banque pour effectuer une normalisation du nombre d'EST par contig. Les tests qui ont été utilisés sont les suivants : test d'Audic & Claverie, test exact de Fisher, test du Chi², test R de Stekel & Falciani, test du Chi² multiple.

C.6.4 Annotation des spots protéiques

Les spectres MS/MS des peptides sélectionnés sont comparés aux spectres théoriques générés par digestion *in silico* de banques d'EST traduits (Pinus Gene Index du TIGR) et de banques protéiques (Swissprot). La comparaison est automatisée à l'aide des modules de recherche Turbo Sequest et Sequest Browser. Une liste de propositions d'assignations de séquences est générée et différents paramètres vont permettre de sélectionner les propositions les plus pertinentes :

- Le Xcor traduit la corrélation entre les spectres MS/MS expérimental et théoriques et doit supérieur à 1.9 pour un ion monochargé, 2.2 pour un dichargé et 3.75 pour un trichargé.
- Le ΔCn représente la différence de score trouvée entre les candidats consécutifs lors de la recherche d'homologie avec des séquences connues. Il doit être supérieur à 0.1 entre les deux premiers candidats.

- Le nombre d'ions fragments trouvés par rapport au nombre d'ions fragments attendus. Il faut retrouver au moins 2 peptides pour que l'assignement soit considéré comme valable.

L'annotation est alors attribuée sur la base d'une recherche d'homologie (Blast) de la séquence obtenue précédemment sur l'identité des spectres MS/MS avec les bases de données publiques protéiques. L'utilisation des bases de données *Pinus* du TIGR, qui contiennent 320000 EST dont les EST de pin maritime, est utile à l'identification de protéines spécifiques au genre *Pinus*. Toutes les données protéiques des échantillons, aux spectres, en passant par la quantité des protéines, seront prochainement disponibles à l'URL (http://moulon.inra.fr/~langella/proticdemo/Protic/home/index.php?menu=&file=credits) grâce à la base de données PROTICdb (Ferry-Dumazet *et al.* 2005) développées par l'équipe Bioinformatique du Laboratoire Commun de Biologie Moléculaire et de Bioinformatique de l'INRA du moulon en collaboration avec le CBiB et notre unité.

Résultats et discussion

D RESULTATS ET DISCUSSION

D.1 Matériel végétal et paramètres physiologiques

D.1.1 Culture Hydroponique

Les premières levées ont eu lieu 3 semaines après la mise en germination et les plantules de la provenance marocaine ont atteint une taille suffisante pour être transférées en hydroponie au bout de 4 semaines. La taille minimale est celle de la radicelle qui doit être d'environ 5 cm pour atteindre le milieu liquide dans notre dispositif. Les plants d'origine landaise ont levé quelques jours plus tard et ont pu être transférés en hydroponie 5 semaines après la mise en germination. Au moment du transfert, les plants présentaient un flétrissement marqué sur une période de 24h. L'éclairage a donc été coupé pour éviter tout échauffement. Une cinquantaine de plants ont été placés par bac de culture.

Lors des deux premières séries de culture, l'attaque par le *Fusarium sp.* s'est déclarée durant la première semaine d'exposition au stress. La troisième mise en culture sur solution nutritive diluée au 1/5e en présence d'antifongique a permis d'éviter une autre attaque.

Après 8 semaines de culture, les jeunes plants sont très vigoureux. Les feuilles aciculaires (euphylles) d'un vert sombre et intense ne présentent aucun défaut de pigmentation signe d'une chlorose ou autre carence. Les premières pseudophylles se sont développées sur certains individus (Fig. D-1a). Les racines sont très bien développées (Fig. D-1b). Le port du plant diffère par contre nettement de celui des semis de pin maritime en conditions naturelles qui sont élancés, peu ramifiés et à entrenœuds longs (Fig. D-2a). Les plantules sont dans le cas présent très ramassées, à entrenœuds très courts et portent de nombreuses ramifications (Fig. D-2b). Au moment du transfert du milieu nutritif vers le milieu hyper osmotique, les plants présentent un flétrissement rapide qui disparaît progressivement au cours des premières heures.

D.1.2 Potentiel de base

La mesure du potentiel de base étant destructive, cinq individus ont été prélevés pour chacun des écotypes et conditions analysés. Les échantillons ont été prélevés dans les conditions « témoin court » et « 48 heures de stress » pour l'étude de la réponse à court terme et dans les conditions « témoin long » et « stress long » pour l'étude de la réponse à plus long terme. L'obscurité totale a pu être maintenue le temps des prélèvements et des mesures, de 4

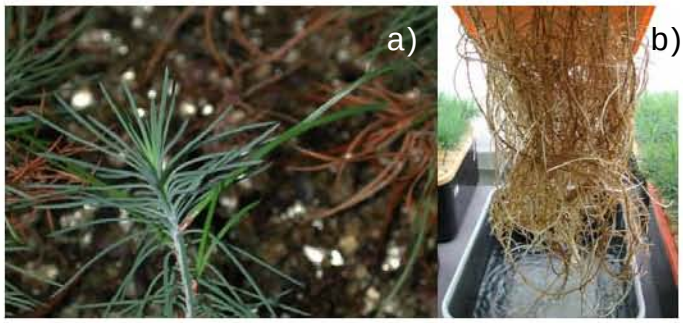


Figure D-1 :
a- Développement des aiguilles secondaires ou pseudophylles
b- Développement du système racinaire de 48 plantules âgées de 7 semaines

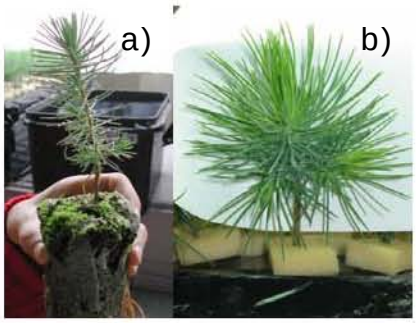


Figure D-2 :
a- Morphologie des jeunes plants de pin maritime élevés en conditions naturelles
b- Développement à entrenœuds courts des plants élevés en hydroponie

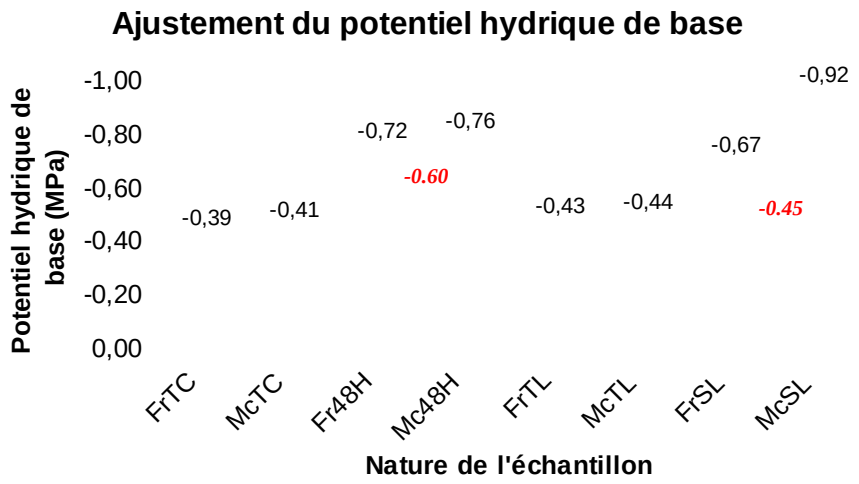


Figure D-3 : Valeurs moyennes et écart-types des potentiels hydriques de base en MPa pour les différents échantillons analysés. (Fr = Provenance Française, Mc = Provenance Marocaine, TC = Témoin Court, TL = Témoin Long, 48H = 48H de Stress, SL = Stress Long)

facteur	Stress Court	Stress Long
Provenance	*0,2712*	**8,95e-07**
Condition	**1,96e-09**	**7,48e-13**
Interaction	*0,8545*	**2,77e-06**

Tableau D-1 : Résultats des analyses de variance (P-Value) pour tester l'hypothèse nulle H0 = « Les potentiels hydriques de base (Ψh) des échantillons ne sont pas différents » sur les mesures de l'expérience « Stress court » et de l'expérience « Stress Long ». Les facteurs testés dans le modèle sont la Provenance (P), la Condition (C) et l'Interaction. (Ψh = P + C + PC + ε)

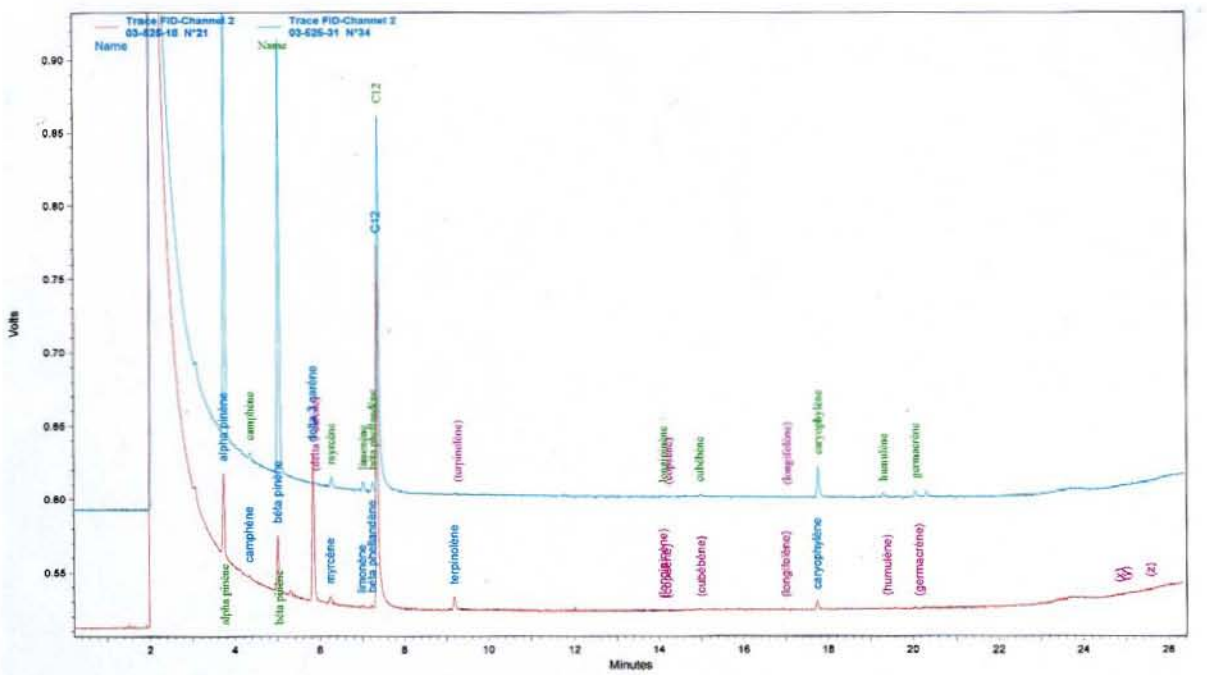


Figure D-4 : Chromatogramme de détection des composés terpéniques obtenus par chromatographie en phase gazeuse. Intensité du signal au niveau du détecteur (en ordonnée) en fonction du temps de rétention (en abscisse).

heures à 10 heures du matin. Les valeurs moyennes obtenues et les écart-types sont représentés dans le graphique (Fig. D-3). Le potentiel hydrique de base reste stable autour de -0.4MPa (soit -4 bars) pour les deux provenances en l'absence de stress osmotique. En situation de stress osmotique violent lors de l'expérience stress court, le potentiel osmotique de la solution nutritive atteint -0.6MPa et les deux provenances répondent en ramenant leur potentiel hydrique de base à une valeur comprise entre -0.72 à -0.75MPa (soit -7.2 et -7.5 bars). L'exposition à un stress de 3 semaines dans une solution nutritive à -0.45MPa permet de différencier les deux provenances quant à leur réponse au stress. Pour les plants d'origine marocaine, le potentiel hydrique passe de -0.44MPa en l'absence de stress à -0.92MPa contre -0.66MPa pour la provenance landaise. L'écotype marocain aura donc plus de facilité pour capter l'eau du milieu. Une analyse de variance sur les valeurs de potentiel hydrique de base avec comme variables explicatives la provenance, la condition et l'effet d'interaction a été conduite séparément sur les données de l'expérience « Stress court » et « Stress long ». Les probabilités de rejet de l'hypothèse nulle sont rapportées dans le tableau D-1. Dans le cas du stress court, seul l'effet condition est significatif. Pour l'expérience « Stress Long », on observe que les effets provenance, condition et interaction sont hautement significatifs. Dans ce cas, le potentiel hydrique de base diffère entre provenances en condition de stress (Test de Student : P-value=1.23^{-7}) mais pas en condition témoin (Test de Student : P-value=0.72). Les deux provenances ne diffèrent donc que pour l'exposition au stress long.

D.1.3 Dosage des terpènes

Des études sur des plantes méditerranéennes, ont montré que la sécheresse / température avait une influence sur l'évolution des concentrations en terpènes (Llusià and Penuelas 1998; Llusia and Penuelas 2000). Une étude sur *Q. ilex* et *P. halepensis* a montré une relation entre la teneur en terpènes émis et respectivement le taux de photosynthèse (PFD) / la température pour le chêne / le pin (Penuelas and Llusia 1999). Une autre étude portant sur *Pinus sylvestris* et *Picea abies* soumis à la sécheresse a montré une augmentation en monoterpène chez ces deux espèces (Turtola *et al.* 2003). Les terpènes sont donc des marqueurs potentiels du stress hydrique.(Llusià and Penuelas 1998)

La technique de détection des terpènes par chromatographie en phase gazeuse permet l'identification et la quantification de 16 composés terpéniques : des monoterpènes $C_{10}H_{16}$ (alpha pinène, beta pinène, delta 3 carène, camphène, limonène, myrcène, beta phellandrène, terpinolène), des sesquiterpènes $C_{15}H_{24}$ (caryophyllène, copaène, cubébène, germacrène, humulène, longifolène, longipinène) et les C12 (Fig. D-4). L'extraction des terpènes par le

pentane concernait uniquement les parties aériennes des plants contrôles et 3 semaines de stress de l'expérience « Stress long » pour laquelle nous avions détecté une différence sur l'ajustement du potentiel de base. Les terpènes en C12, le longipinène et le copaène n'ont pas été détectés dans ces échantillons. Les terpènes les plus abondants sont l'alpha pinène, le beta pinène et le delta-3 carène. Pour chaque condition de chaque provenance, de 5 à 6 plants ont été prélevés et les dosages ont été dupliqués. Les quantités de terpènes mesurées dans les prélèvements ont été ramenées à une teneur par gramme de matière fraîche. Le tableau D-2 regroupe les valeurs de probabilité associées au test F des analyses de variance. Le modèle testé prend en compte un effet de la provenance, de la condition et de l'interaction.

En considérant un seuil de significativité à 5%, l'effet provenance est significatif pour l'alpha pinène, le camphène, le limonène le caryophyllène et l'humulène. L'effet condition n'est mis en évidence pour le delta 3-carène, le myrcène, le terpinolène, le caryophyllène et l'humulène.

Sur la figure D-5a, il apparaît que la teneur en **alpha pinène** est légèrement supérieure dans la provenance marocaine mais c'est surtout le rapport alpha pinène sur béta pinène, nettement plus élevé dans la provenance marocaine qui différencie les deux écotypes (Fig. D-5b).

Le **delta 3-carène** est pratiquement indétectable dans la provenance marocaine contrairement aux plants d'origine landaise qui ont tendance à accumuler ce composé au cours du stress même si une forte variation de concentration est observable (Fig. D-5c). Ce type de profil se retrouve également pour le **terpinolène** (Fig. D-5d).

La teneur en **caryophyllène** est globalement plus élevée dans la provenance marocaine et augmente au cours du stress.

Le **cubébène** et le **germacrène** ne semblent s'accumuler que chez les plants stressés d'origine landaise.

D.1.4 Discussion

La culture de plants de pin maritime en condition hydroponique a été problématique en raison des attaques fongiques rencontrées au cours des deux premières expérimentations. Le maximum de précautions possibles avait pourtant été mis en œuvre pour éviter ce type de contaminations. Les deux premières cultures se sont déroulées dans une salle climatisée et équipée d'un éclairage horticole dont l'accès était restreint. Le nettoyage complet à l'eau de javel de la salle et du matériel, et l'utilisation d'eau distillée associée à l'autoclavage des milieux n'ont pas suffit à éviter le développement de *Fusariu*m. Le fait que l'apparition du pathogène coïncide avec l'application du stress laisse supposer que le *Fusarium* profite d'un affaiblissement du système de défense des plants provoqué par le stress osmotique. Les

Composé	Modèle avec Interaction		
	Provenance	Condition	Interaction
rapport alpha/beta	5,00e-03	0,98	0,46
alpha pinène (C10)	8,00e-04	0,38	0,57
Camphène (C10)	9,00e-04	0,34	0,14
béta pinène (C10)	0,79	0,52	0,38
delta 3 carène (C10)	0,28	0,04	0,14
Myrcène (C10)	0,84	0,03	0,13
Limonène (C10)	0,03	0,24	0,54
béta phellandrène (C10)	0,7	0,15	0,37
Terpinolène (C10)	0,37	0,02	0,09
Longifolène (C15)	0,16	0,14	0,1
Caryophyllène (C15)	0,02	0,04	0,86
Humulène (C15)	0,017	0,027	0,17
Germacrène (C15)	0,51	0	1,00e-04
Cubébène (C15)	*1*	*1,40e-03*	*0,015*

Tableau D-2 : Valeurs de probabilité (P-value) des analyses de variance sur les dosages des composés terpéniques extraits des parties aériennes. Les valeurs significatives à 5% sont indiquées en couleur foncée. Pour le germacrène et le cubébène, très peu de valeurs étaient disponibles.

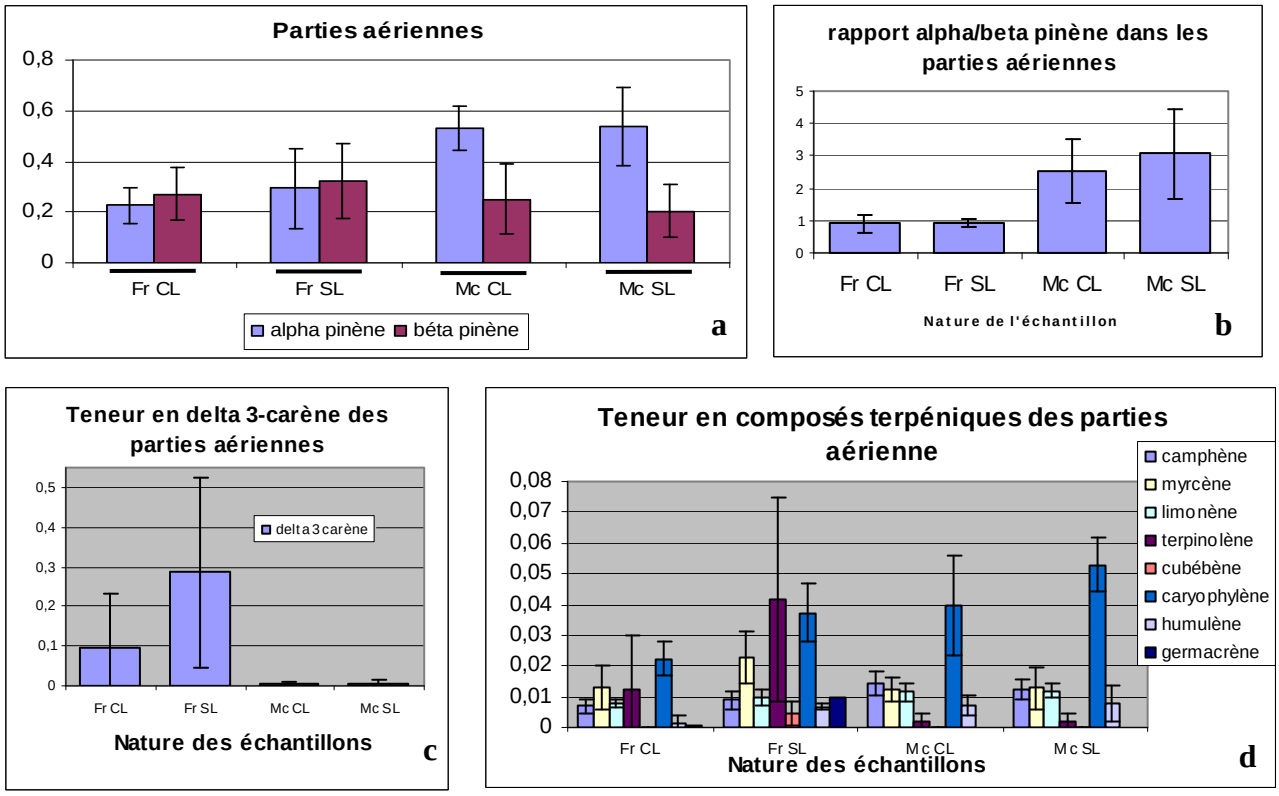


Figure D-5 : Teneur en composés terpéniques pour les provenances Françaises (Fr) et Marocaines (Mc) en condition contrôle (CL) et stress long (SL).

transcrits codants pour des Défensines ou des « pathogenesis related protein (PR)» sont d'ailleurs très souvent surexprimés chez les plants soumis à des stress hydriques. Dubos *et al.*. avaient, par exemple, détectés une PR-10 sur des plants de pin maritime élevés en hydroponie (Dubos and Plomion 2001). Il est également intéressant de noter que ce champignon à l'origine de fusariose par *Fusarium oxysporum lycopersici* (FOL) est un pathogène rencontré couramment dans les cultures de tomates hors sol qui provoque la pourriture de la racine et du collet. Après deux expérimentations exposées au pathogène, la seule alternative restante était d'utiliser un antifongique en traitement préventif. L'ajout de tout produit phytosanitaire peut entraîner une modification de l'expression de certains gènes ou inhiber certaines voies métaboliques impliquées dans la réponse au caractère étudié. On peut cependant penser que l'impact de l'antifongique sera similaire chez les plants contrôles et les plants stressés.

Concernant les conditions d'éclairage utilisées pour cette expérience, la croissance des plantes indique que du point de vue quantitatif, nous répondons aux besoins des plants. Par contre, du point de vue qualitatif, le port des plants élevés en hydroponie diffère de celui des plants en conditions naturelles par la présence d'entrenœuds courts. Cela confirme l'hypothèse d'un éclairage à valeur de phi élevée (RC/RL important). Il faudrait donc prévoir un apport de rayonnement RL par des sources incandescentes à raison de 10%-20% de l'énergie totale utilisée en prêtant attention à l'apport calorifique engendré.

Le flétrissement observé visuellement au moment de l'application du stress osmotique suivi d'une récupération rapide indique que le stress est suffisant pour induire une réponse de la plante. L'évolution des valeurs de potentiel hydrique de base nous confirme que les plants subissent effectivement un stress osmotique et répondent en adaptant de façon significative leur potentiel hydrique. En l'absence de stress, les plants ont une turgescence optimale. Le potentiel hydrique mesuré sur l'ensemble de l'appareil aérien et de la tige est proche de -0.4MPa pour une solution nutritive dont le potentiel osmotique est de -0.08MPa ((Dubos 2001)), le potentiel hydrique est donc nettement inférieur au potentiel osmotique. La solution hyper osmotique provoque bien une diminution du potentiel hydrique mais qui est de l'ordre de 0.33MPa au bout de 48H ce qui est inférieur à la variation du potentiel osmotique de la solution qui est de 0.60MPa. L'intensité du stress ou le temps d'application ne permettent peut être pas aux plants de mettre en place l'ensemble des mécanismes de réponse. Le potentiel hydrique des plantes reste cependant inférieur au potentiel osmotique de la solution, ce qui contribue à la disparition du flétrissement. Dans le cas de l'expérience « stress long », l'état d'équilibre doit être atteint et la faible mortalité (moins de 5%) semble indiquer que

l'intensité du stress est supportée par la plante. L'abaissement de potentiel hydrique n'est que de 0.23MPa pour l'écotype landais alors que la variation de potentiel osmotique de la solution atteint 0.45MPa. Si le flétrissement est absent, l'abaissement du potentiel hydrique ne suit pas celui de la solution. Par contre, l'écotype marocain abaisse son potentiel hydrique dans les mêmes proportions que l'abaissement du potentiel osmotique de la solution (0.47MPa). Ces mesures mettent en évidence que les deux provenances régulent leur potentiel hydrique de façon significativement différente. L'écotype marocain a clairement un ajustement osmotique plus efficace que la provenance landaise, ce qui confirme les résultats de Nguyen (1989b). Les conditions expérimentales appliquées devraient donc permettre de mettre en évidence des réponses au stress chez les deux provenances.

Les monoterpènes sont des composés très volatils, il est même possible de mesurer et suivre l'évolution des teneurs atmosphériques en alpha-pinène, beta-pinène, delta-3-carène ou limonène émis par les pins maritimes (Simon *et al.* 1994). Les teneurs atmosphériques mesurées par Simon *et al.* (1994) montrent des variations journalières corrélées à de nombreux facteurs tels que la température, l'intensité lumineuse ou la pression de vapeur d'eau. Une part de la variabilité intra-échantillon observée peut provenir d'une hétérogénéité lors du prélèvement, de la volatilité des composants et des faibles teneurs en terpènes des échantillons. Les analyses statistiques permettent cependant de détecter un effet provenance pour plusieurs composés. La variabilité génétique de teneur en monoterpènes entre différents écotypes de *Pinus pinaster* a déjà été constatée (Tognetti *et al.* 2000). A ce titre, le rapport de concentration entre l'alpha-pinène et le beta-pinène est un très bon marqueur de distinction des provenances landaises et marocaines (P. Pardon, communication personnelle). La régulation des teneurs par le stress hydrique est moins nette. Mis à part le caryophyllène dont les quantités diffèrent à la fois pour l'effet provenance et pour l'effet condition, l'analyse statistique avec prise en compte de l'effet d'interaction a permis de détecter des effets conditions. Il semble que la production de plusieurs composés terpéniques soit modifiée lors de la réponse au stress hydrique, notamment le delta-3-carène, le myrcène, l'humulène et le caryophyllène dont la production augmente au cours du stress. Les monoterpènes (C_{10}) dérivent du GPP tandis que les sesquiterpènes (C_{15}) sont synthétisés à partir du FPP (Fig. D-6). Nous avons déjà indiqué dans la synthèse bibliographique que la synthèse de l'ABA et des GA dérivait d'isoprènes comme le GGPP ou les carotènes. On peut donc s'attendre à observer une régulation de certaines enzymes de la voie de biosynthèse des terpènes. Comme nous le verrons lors de la présentation des résultats des transcrits, nous avons détecté des gènes

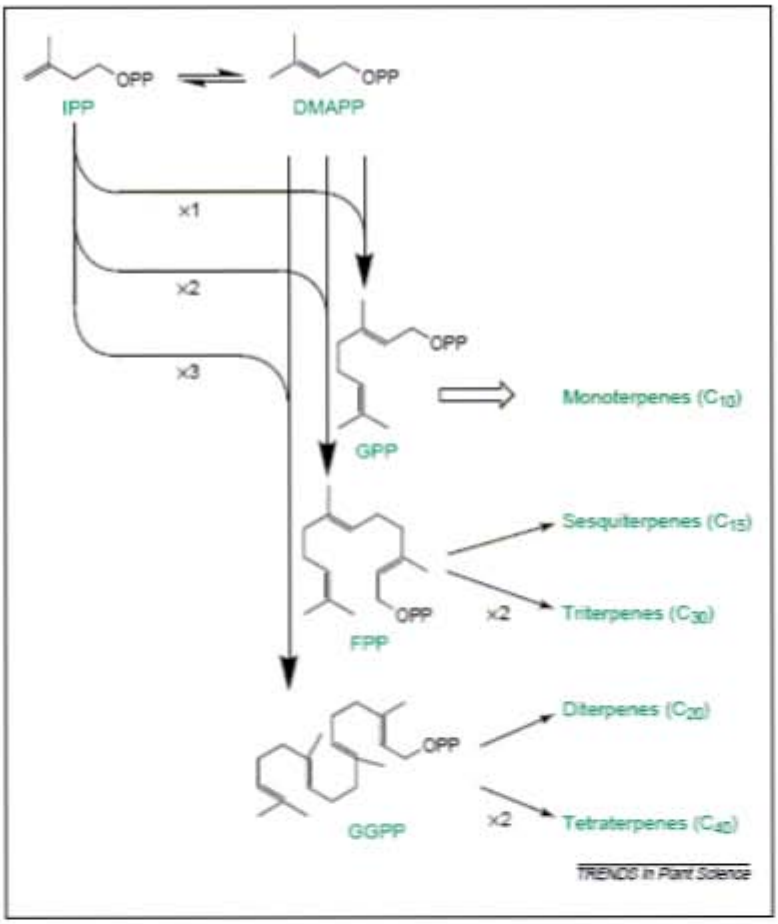


Figure D-6 : Voie de biosynthèse des isoprènes. La condensation avec le DMAPP d'un IPP permet la synthèse de GPP, de 2 IPP donne le FPP et de 3 IPP donne le GGPP. Le GGPP est un précurseur de la synthèse de l'Acide Abscissique et des Gibbérellines (Mahmoud, 2002). Abréviations : IPP : Isopentenyl diphosphate, DMAP : Dimethylallyl monophosphate, GPP : Geranyl Pyrophosphate, FPP : Farnesyl Pyrophosphate, GGPP : Geranyl Geranyl Pyrophosphate.

présentant des homologies avec la pinène synthase et la GPP synthase. De plus les composés terpéniques peuvent jouer un rôle dans la résistance aux pathogènes. Le limonène présent à des concentrations légèrement supérieures chez la variété marocaine aurait un pouvoir répulsif sur des insectes pathogènes comme la chenille processionnaire (Tiberi *et al.* 1999). Les terpènes sont également à l'origine des caroténoïdes, leur implication dans la réponse au stress osmotique n'est donc pas étonnante. Ce type de stress se traduit également par un stress oxydatif face auquel la plante peut répondre par la production de composés antioxydants tels que les caroténoïdes. Pour confirmer cette hypothèse, il serait nécessaire de réaliser un nouveau plan d'expérience dans ce sens.

D.2 Analyse et annotation des séquences

Ce chapitre présentera les caractéristiques des séquences générées et les résultats des traitements bioinformatiques qui ont été appliqués. Nous commencerons par l'assemblage des séquences puis leur annotation et leur répartition en catégories fonctionnelles. Nous continuerons par la présentation des données d'expression in silico. Nous terminerons par la réalisation de l'Unigène et la discussion des résultats obtenus.

D.2.1 Séquençage

Au début de la thèse, nous disposions de 6984 séquences issues d'une banque d'ADNc de xylème ainsi que 721 séquences de banques de parties aériennes de plantules élevées en hydroponie dont 242 provenaient de plants contrôles et 479 de plants stressés. Le projet de séquençage d'EST des banques de racines que nous avons mené au cours de la première année de cette thèse a aboutit à l'étalement de plus de 14000 clones. Après élimination des séquences de moins de 100 paires de bases, plus de 9500 séquences ont été générées. La première phase de traitement qui a suivi consistait à nettoyer individuellement chaque séquence en fonction de l'indice de qualité global et par base pour obtenir un score Phred20 supérieur à 85%. Cette démarche permet de s'assurer de la qualité des séquences avant injection dans le pipeline bioinformatique (cf Matériel et méthodes) mais prend un temps considérable. En analysant a posteriori nos critères de décision et la logique appliquée sur ce lot de données, cette étape manuelle a pu être entièrement automatisée. Un script Perl se charge de détecter et d'éliminer alternativement les zones de basse qualité aux extrémités de la séquence en prenant en compte le score moyen et le nombre de bases de basse qualité consécutives comme nous le faisions manuellement. La présence de séquences correspondant au vecteur de clonage est finalement recherchée en vu de leur élimination.

Deux projets distincts, l'un d'EST issus d'une banque d'ADNc construite à partir de xylème, l'autre construite à partir des racines et parties aériennes, et un projet global (ensemble des EST de pin maritime) ont été construits et injectés dans le pipeline. Dans le cadre de cette thèse nous nous sommes intéressés plus particulièrement aux séquences issues de racines (Fig. D-7).

Ce projet comprend 10238 séquences obtenues sur 9711 clones différents (plusieurs passes pour certains clones) dont 8996 ont pour origine les banques de racines et 715 sont issus des parties aériennes. Au total, plus de 4.4 Mb ont été séquencées. Les EST des racines stressées représentent 4324 séquences contre 5193 séquences pour les EST des racines contrôles.

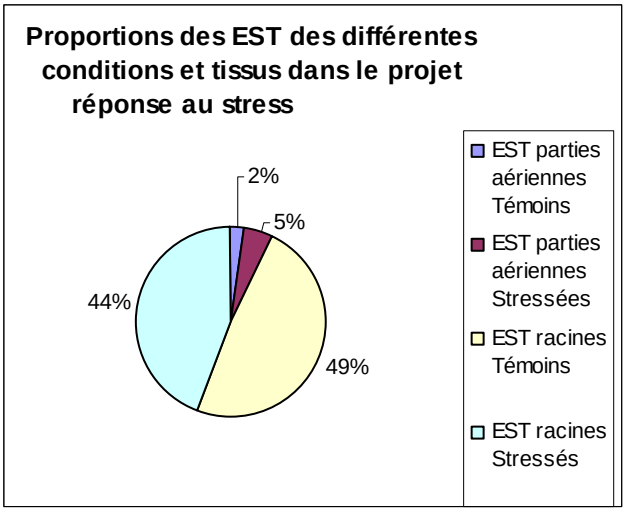


Figure D-7 : Représentativité des différents tissus et conditions dans le projet réponse au stress.

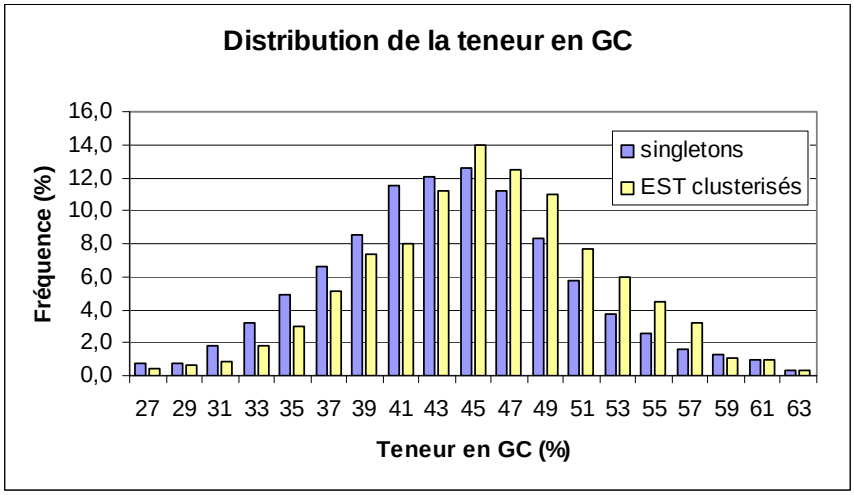


Figure D-8 : Distribution des concentrations en bases G et C des ESTs pour les singletons et les séquences clustérisées.

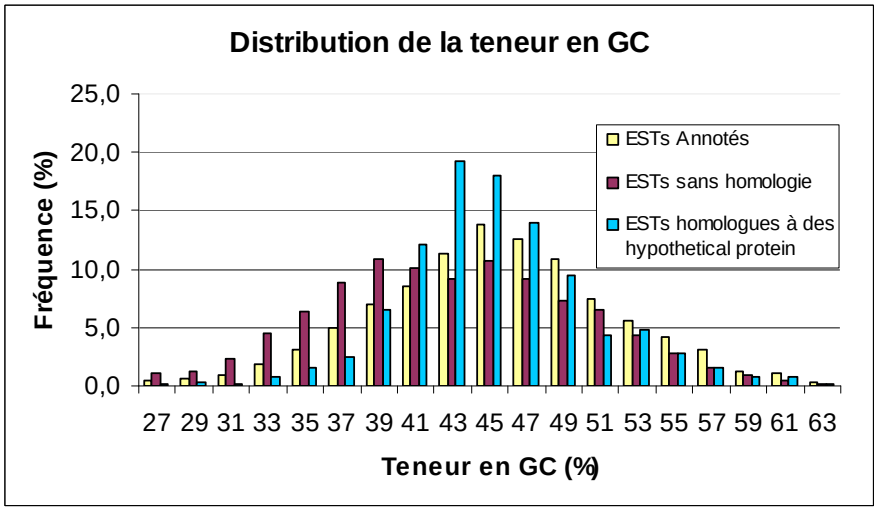


Figure D-9 : Distribution des concentrations en bases G et C des ESTs pour les séquences annotées, les homologues à des « hypothetical proteins » et les séquences sans homologies.

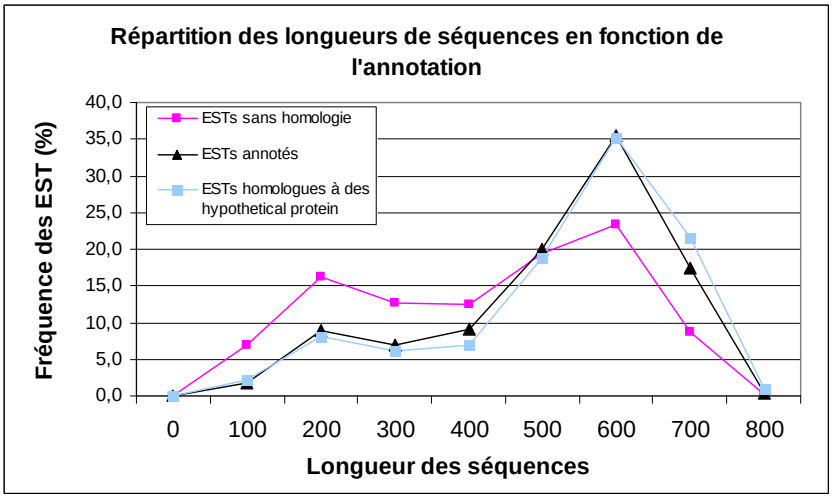


Figure D-10 : Répartition des longueurs des séquences en fonctions de l'annotation.

Si l'on considère les résultats de l'assemblage et de l'annotation que nous aborderons dans les chapitres suivants, nous pouvons calculer un indice de redondance définit comme :

$$\mathrm{Re}\,dondance = \frac{\text{nombre d'EST clustérisés - nombre de consensus}}{\text{nombre total d'EST}}$$

Ce projet étant constitué de 7355 EST clustérisés dans 2261 consensus, la redondance est donc de 49,7 % ce qui correspond à 5144 clones différents (2883 singletons et 2261 consensus) pour un total de 2.5 Mb non redondantes.

La teneur en GC est globalement de 44.9% mais n'est pas distribuée de façon homogène selon le type de séquence étudié. La figure D-8 présente les distributions des teneurs en GC des singletons et des EST clustérisés. Il apparaît que la distribution des singletons est centrée sur une valeur de GC plus faible (43.7 %) que celle des EST clustérisés (45.3 %). L'hétérogénéité de la distribution est aussi constatée en distinguant les séquences selon leur annotation (Fig. D-9). La distribution de la concentration en GC pour les séquences annotées est similaire à celle des EST clustérisés de la figure D-8. La distribution des EST présentant une homologie avec des « hypothetical proteins » est asymétrique. Les faibles teneurs en GC sont sous représentées mais la classe la plus abondante se situe autour de 43-45 %. La distribution des EST sans homologie avec des séquences connues est bimodale, la partie droite suit la distribution des séquences annotées et la partie gauche est centrée sur la classe 39-41 %.

La longueur moyenne des séquences obtenues après nettoyage et élimination du vecteur est de 445 paires de bases. Comme pour la teneur en GC, les distributions varient en fonction du type d'annotation rencontrée. La figure D-10, montre que la répartition des longueurs est similaire pour les EST annotés et les EST homologues à des « hypothetical proteins ». Le profil est bimodal, mais plus de 75 % des séquences ont une taille supérieure à 400 pb, la classe 600-700 pb étant la plus abondante. La répartition des longueurs pour les EST sans homologie montre un profil bimodal plus marqué, la partie droite correspondant aux courbes que nous venons de décrire, la partie gauche étant centrée sur la classe 200 à 300 pb. Une analyse statistique (ANOVA et test-T) indique que la courbe de longueur des EST sans homologie diffère significativement des deux autres.

Le profil bimodal des courbes de longueur des EST reflète les difficultés rencontrées pour optimiser les performances de la réaction de séquence avec la chimie utilisée. Lorsque les conditions sortent légèrement des conditions optimales, cela se traduit par la génération de séquences courtes avec une chute de la qualité. Nous obtenons soit des séquences longues soit des séquences courtes en fonctions de la réaction de séquence ou des paramètres d'injection.

Nous venons de voir que les courbes de distribution de longueur et de teneur en GC différaient en fonction du type d'annotation assigné aux séquences. Les séquences courtes sont plus abondantes dans les EST sans homologie et leur teneur en GC est plus basse que celle des autres EST. Nous avons observé une distribution bimodale pour la teneur en GC des EST sans homologies avec un pic à 39% et un pic à 45 %. Des données similaires ont été obtenues sur les EST de *Solanum tuberosum* avec une teneur de GC inférieure de 4% pour les EST dont les résultats Blast $<10^{-2}$ (Crookshanks *et al.* 2001). Si l'on s'intéresse à la teneur en GC calculée à partir des génomes d'*A. thaliana* et *O. sativa* (Yu *et al.* 2002), les valeurs de GC sont plus basses pour les introns que pour les exons, respectivement (0.324 – 0.370) et (0.432 – 0.514), et la teneur en GC décroit de l'extrémité 5' vers l'extrémité 3' dans les gènes du riz. Ainsi, de faibles valeurs de GC pour les EST sans homologies pourraient être reliées avec une grande proportion de séquences de la région 3'. Cette hypothèse concorderait avec la stratégie de séquençage adoptée. La région 3' étant connue pour déterminer la spécificité des membres d'une famille multigénique plutôt que la fonction de la protéine. Les recherches d'homologie sont plus difficiles en 3'. Une autre interprétation serait une plus forte occurrence de séquences de la région UTR (UnTranslated Region) que l'on suppose plus proches des régions introniques pour leur teneur en GC. Les séquences courtes rendent également plus difficile la recherche d'homologie et l'assemblage, ce qui peut également expliquer leur plus forte proportion dans la catégorie des EST sans homologie.

D.2.2 Assemblage des séquences

L'assemblage bioinformatique des séquences permet d'identifier et d'aligner les séquences des EST issus d'un même gène (cf. Matériel et méthode). Les trois assemblages bioinformatiques, accessibles à l'URL : http://cbi.labri.fr/outils/SAM2/COMPLETE/, sont les suivants :

- L'injection des 10238 séquences du projet « réponse au stress » dans le pipeline a conduit à leur regroupement en 2118 contigs contenus dans 1719 clusters et 2883 singletons. Des consensus primaires ont été générés pour chaque contig. Des consensus supplémentaires, au nombre de 143, ont également été générés pour 58 contigs (Fig. C-18). Ce projet est nommé « Pinus Pinaster root EST ».
- Le projet EST de xylème, accessible sous le nom « Pinus pinaster xylem EST », est organisé en 995 clusters contenant 1226 contigs et 2948 singletons.
- Le projet « Pinus pinaster total » comprenais l'ensemble des EST soit 18498 séquences réparties en 2697 clusters, 2893 contigs et 5051 singletons. Pour ce projet,

4003 consensus dont 2893 consensus primaires ont été générés. C'est à partir de ces 2893 contigs et 5051 singletons que nous avons choisi les séquences de l'Unigène dont nous parlerons plus loin. Une nouvelle version du projet a été mise récemment en ligne et contient maintenant 17770 séquences.

Pour le projet « racines », 72 % des séquences sont clustérisés contre 73 % pour le projet total. La fusion des deux projets xylème et racines n'a donc pas permis de réduire significativement la proportion de singletons. Cela peut signifier qu'il existe un grand nombre de gènes faiblement exprimés pour lesquelles nous ne disposons que de séquences sans recouvrement. Les deux banques concernent cependant des tissus très différents et l'ajout de séquences ne provoque donc pas une simple augmentation du nombre d'EST par gène ce qui permettrait de réduire le nombre de singletons, mais l'apport d'EST correspond à des gènes ou membres de familles « spécifiquement » représentés dans chacun des tissus. L'ajout des séquences de xylème permet de passer de 2118 à 2893 contigs. De façon simplifiée, cela signifie que 775 nouveaux « gènes » sont détectés par ajout des séquences de xylème. Dans le même temps, si la proportion de consensus alternatifs restait très faible au sein du projet « racines », dans le projet total, nous avons 1110 consensus alternatifs pour 2893 consensus primaires. Les consensus alternatifs sont générés dans le cas où l'alignement des EST constitutifs d'un contig présente des régions polymorphes. Les EST sont alors classés en plusieurs groupes de séquences similaires et un consensus est généré pour chaque sous-groupe. Une région est considérée polymorphe si elle atteint plusieurs dizaines de bases (selon paramètres de l'algorithme). Cela ne concerne donc pas les erreurs ponctuelles de séquençage ou des polymorphismes de quelques bases isolées (SNP). La forte proportion de consensus alternatifs indique donc qu'un même gène présente un polymorphisme sur une région de quelques dizaines de paires de bases entre les transcrits des racines des plants landais utilisés en hydroponie et les transcrits des arbres corses utilisés pour l'étude du « xylème ». Ce polymorphisme peut correspondre à différents membres de familles multigéniques ou des formes allèliques très différentes.

Le tableau D-3 présente les annotations des contigs constitués de plus de 20 EST dans le projet « racines ». Le contig le plus représenté correspond à la sous unité 25S ribosomique qui regroupe 157 EST soit un peu plus de 1.5 % des transcrits séquencés. Les ARN ribosomaux font partie des transcrits les plus abondants de la cellule. Certains membres de cette famille sont exprimés de façon constitutive et stable dans la cellule tel que l'ARN 18S

Tableau D-3 : Nombre d'ESTs de racines et Annotation des Contigs comportant plus de 20 ESTs dans le projet « racines ». Seules les 8996 ESTs issus de racines sont pris en considération.

Nombre d'ESTs	Annotation
157 (1.75 %)	25S ribosomal subunit (AJ271027)
58	Phosphopyruvate hydratase (EC 4.2.1.11)
42	Naringenin-chalcone synthase (EC 2.3.1.74)
34	Elongation factor 1-alpha
30	Avr9/Cf-9 rapidly elicited protein 146 (Q9FQZ6)
30	Methionine adenosyltransferase (EC 2.5.1.6)
30	Ubiquitin
28	Translationally controled tumor protein
28	Late embryogenesis abundant protein
26	60S ribosomal protein L41 (U26255)
23	Tonoplast intrinsic protein (P24422)
23	Proline-rich protein (Q9LLZ6)
21	ADP-ribosylation factor (P51823)
20	Metallothionein-like protein EMB30 (Q40854)
20	Polcalcin (O64943)
20	Defensin (O65740)
20	Catalase (EC 1.11.1.6)
20 (0.2%)	putative protein

dans les racines de pin (Sathyan 2004) tandis que d'autres ont leur expression régulée au cours du stress comme des 40S ou 60S (Kawasaki *et al.* 2001; Ozturk *et al.* 2002).
On rencontre également des transcrits de la Phosphopyruvate hydratase (EC 4.2.1.11) ou enolase qui intervient dans la glycolyse en catalysant la réaction de déphosphorylation du 2-phosphoglycérate au Phosphoenol-pyruvate. Il est donc confortant de retrouver cet enzyme qui intervient dans la synthèse d'un composé clé de la production d'énergie pour la cellule et qui est indispensable à la production de l'Acetyl-CoA qui intervient dans de nombreuses voies métaboliques.
Les EST homologues à la Naringenin-chalcone synthase (EC 2.3.1.74) sont rencontrés 42 fois. Cet enzyme est nécessaire à la biosynthèse des flavonoïdes. Ces composés jouent le rôle d'antioxydant pour la cellule et évite les dégradations des constituants cellulaires. La catalase (EC 1.11.1.6) présente dans la liste avec 20 EST va aussi intervenir dans les phénomènes de prévention de l'oxydation. Cet enzyme catalyse la réaction de conversion du peroxyde d'hydrogène en molécules d'oxygène et d'eau.
Nous trouvons en quatrième position un contig contenant 34 EST homologues au facteur d'élongation 1-alpha. Ce type de transcrits faisait également parti des 6 transcrits les plus abondant dans le transcriptome des tubercules de pomme de terre avec une abondance atteignant 2.6 % des transcrits (Crookshanks *et al.* 2001).
Parmi ces contigs, il y a des fonctions pouvant intervenir dans la réponse au stress hydrique comme une AVr9/Cf-9, une LEA, une protéine riche en proline et une Polcalcine (Tab. D-3).

Les contigs contenant le plus d'EST concernent donc le métabolisme cellulaire (ex : constituants des ribosomes, enzymes de la glycolyse), la protection contre l'oxydation des composant cellulaires (flavonoïdes, catalase) et des enzymes de réponse au stress (LEA, Protéine Riche en Proline, Polcalcine, Défensine).

Un nouveau projet de séquençage du pin maritime vient juste d'être terminé. Il s'agit de 8131 séquences de bourgeons réalisées en partenariat avec le Centre National de Séquençage (CNS). Nous disposons donc depuis l'automne 2005 de 26629 séquences. L'assemblage de ces nouveaux EST avec ceux de racines et du xylème s'organise en 7358 singletons et 3874 Clusters regroupant les 4182 Contigs. Le nombre de consensus passe à 4640 dont 458 sont des consensus alternatifs. Par rapport aux résultats de l'assemblage « Pinaster total », le nombre de consensus alternatifs a fortement diminué et rejoint les proportions du projet racine initial. Le nombre de contig augmente par contre de 2893 à 4182.

L'ajout de séquences de bourgeons a vraisemblablement permis d'éclater les contigs contenant des formes alternatives d'un gène –représentés par les consensus alternatifs– en contigs différents. La proportion de singletons reste stable avec 27 % des séquences.

D.2.3 Annotation des séquences

Le rôle de cette étape est d'attribuer une fonction aux singletons et aux gènes dont la séquence consensus partielle a été obtenue par l'assemblage informatique. L'annotation a été réalisée pour les séquences des projets « xylème » et « racines ». Une attention particulière a été portée au projet « racines » en attribuant aux annotations les noms officiels des enzymes.

D.2.3.1 Annotation des consensus et des singletons

La première étape d'annotation a été réalisée en se servant de l'interface d'annotation du pipeline. Notre travail d'annotation sur les séquences de pin maritime a permis de faire évoluer les fonctionnalités et l'ergonomie de cette interface web entre la version de développement mise en place au début de la thèse et la version actuellement disponible à l'URL (http://cbi.labri.fr/outils/SAM2/). La figure D-12, montre le résultat de cette collaboration avec le Centre de Bioinformatique de Bordeaux. L'annotation se fait, pour les EST clustérisés, au niveau de la séquence consensus puis l'annotation est héritée sur les EST composant le contig. Les résultats de recherche d'homologie sont indiqués en bleu dans la partie inférieure droite de l'interface et leur intitulé alimente le champ « proposition » en haut à gauche. L'annotateur peut sélectionner l'annotation choisie et modifier l'intitulé dans le champ « Annotation ». La seule prise en considération des résultats du Blast et des valeurs de probabilité associées est insuffisante pour assigner une annotation correcte. Il faut également tenir compte de l'étendue de l'homologie trouvée. Une forte homologie avec une séquence connue mais sur quelques dizaines de paires de bases uniquement ne signifie pas que la fonction du gène est celle de l'homologue mais uniquement que les deux gènes ont un domaine commun qui peut éventuellement correspondre à un site actif mais aussi à un domaine sans rapport avec la fonction réelle du gène.

Cela soulève un des risques majeurs de l'annotation par inférence électronique. Une séquence peut présenter une forte homologie avec la séquence d'un gène codant une protéine de fonction connue et vérifiée biochimiquement. Si l'homologie concerne la région codant le site actif de la protéine, l'héritage de l'annotation du gène homologue sur le gène recherché ne pose pas de problème. Dans le cas contraire, si une forte homologie est rencontrée, rien de garantit que la fonction biochimique du gène recherché soit la même que celle du gène

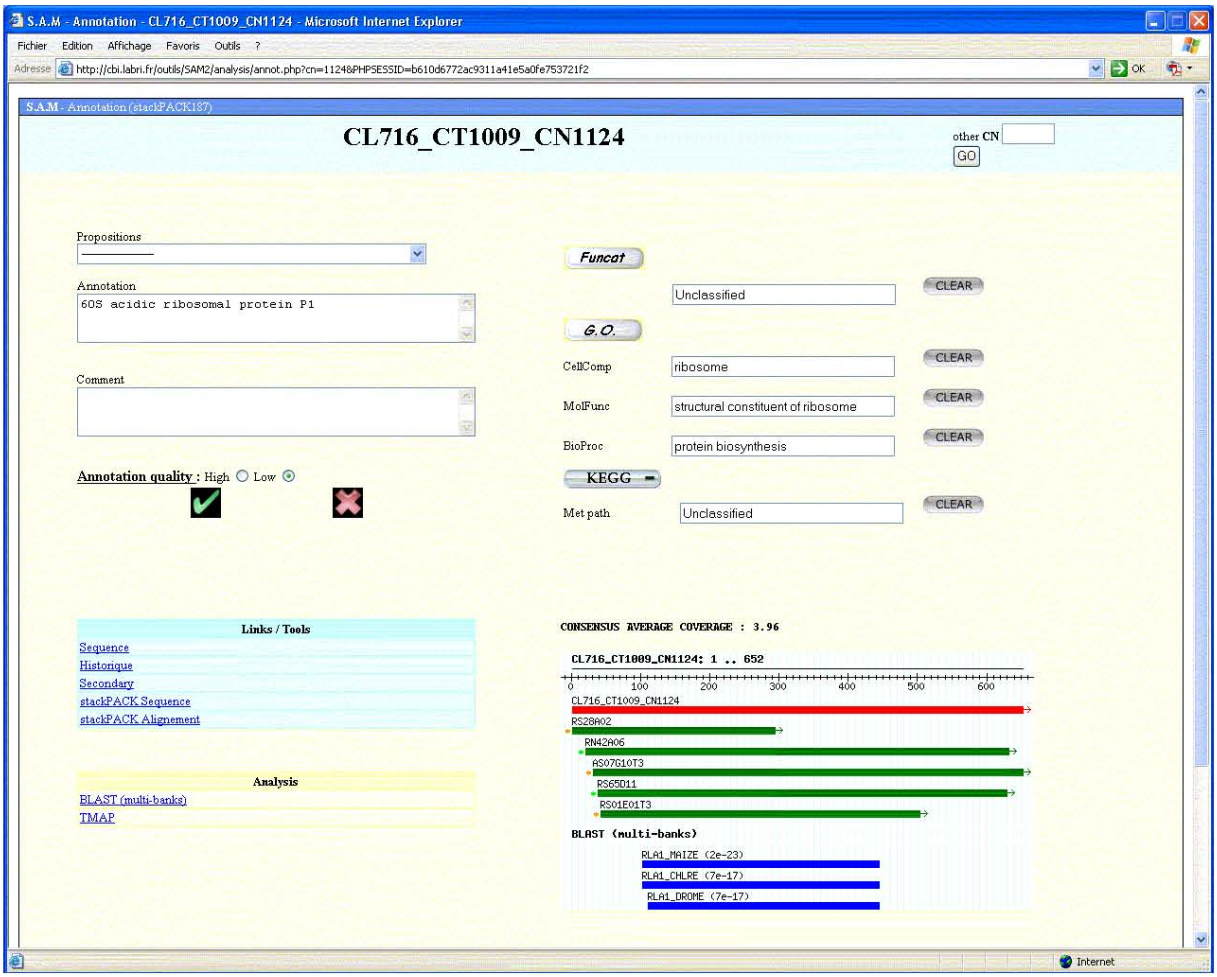


Figure D-12 : Dernière version de l'interface d'annotation du pipeline d'annotation des ESTs.

		Confiance	Quantité info
Blast Prot	SwissProt	++	-
	Trembl	+	++
	TIGR (2005)	+	+/-
Blast Nucl	EMBL	-	++

Figure D-13 : Bases de données publiques utilisées pour la recherche d'homologie. La qualité des annotations, la quantité d'information disponible et le type données (protéique/nucléique) sont indiqués.

homologue et il peut en résulter une annotation erronée. Une autre séquence pourra présenter une forte homologie avec ce gène mal annoté. Dans le cas où l'homologie trouvée ne concerne pas la région qui était homologue à la séquence du gène de départ de fonction vérifié biochimiquement, l'annotation peut également être erronée. De proche en proche, la fonction attribuée au gène recherché peut n'avoir rien de commun avec la fonction réelle de la protéine codée. Une annotation est toujours délicate à réaliser et le maximum de précautions doit être pris pour s'assurer de son exactitude avant de la publier sur les bases de données publiques, ce qui est loin d'être toujours le cas.

La figure D-13 synthétise les bases qui ont été utilisées pour les recherches d'homologie et leurs caractéristiques en termes de confiance dans les annotations indiquées, quantité d'information disponible et type de données proposées. La stratégie de recherche d'homologie en cascade sur les bases SwissProt, TrEMBL puis EMBL (cf. Matériel et méthodes) donne une priorité à la base protéique SwissProt dont la qualité d'annotation est la plus élevée, les fonctions des protéines ayant le plus souvent été vérifiées biochimiquement. Le nombre de séquences disponibles et le nombre d'espèces représentées est cependant plus faible que dans les autres banques (http://www.expasy.org/sprot/). La banque TrEMBL est une traduction de la banque EMBL. La quantité d'information est donc très élevée. Les annotations sont par contre moins fiables que celles de SwissProt. La recherche d'homologie sur des séquences traduite en acides aminés permet de s'affranchir des variations nucléotidiques silencieuses inter espèce et de considérer la fonction du gène. La base EMBL permet la recherche d'homologie contre un maximum de séquences disponibles sur la base de séquences nucléotidiques.

La stratégie en cascade, si elle permet de réduire considérablement le temps nécessaire à la recherche d'homologie, a toutefois un inconvénient en privilégiant essentiellement les résultats obtenus sur les premières bases testées. La valeur de la p-value seuil doit être bien choisie. Dans notre cas, fixée à 10^{-5}, elle s'est avérée trop lâche. Cela se traduisant parfois par un résultat d'homologie faible (ex : 10^{-6}) sur SwissProt car la protéine codée par le gène n'est pas présente dans la base mais possède des similarités faibles avec une autre protéine. Une recherche sur TrEMBL peut par contre renvoyer une très forte homologie avec des séquences connues. Ce type de situation se rencontre notamment pour des gènes déjà séquencés chez *Pinus taeda* qui diffèrent fortement des autres espèces et qui par conséquent ne sont pas présents dans la base SwissProt. Un autre cas de figure concerne une différence de résultats entre le consensus et les EST qui le composent. Cela s'explique par la quantité d'information disponible. Une séquence consensus est plus informative qu'un EST. Plus la longueur de la séquence recherchée est importante, plus précise sera la recherche. Une homologie avec la

séquence consensus peut être trouvée alors qu'aucune homologie correcte n'est observée pour les EST pris individuellement. Comme dans le cas expliqué précédemment, un consensus peut présenter une homologie faible avec des accessions SwissProt alors qu'il existe une forte homologie avec une accession de la base TrEMBL. Dans ce cas, nous avons fréquemment observé que les EST composant le contig présentaient l'homologie avec l'accession de TrEMBL.

C'est pour cela qu'une vérification systématique de la concordance des homologies trouvées pour les EST d'un contig et celles du consensus a été faite. Un système de point de couleur verte ou orange devant la barre verte foncée symbolisant les EST du contig indique si le résultat du Blast de l'EST concorde avec celui du consensus (Fig. D-12). Il est possible de relancer spécifiquement un Blast pour le consensus ou un EST sur une base en particulier.

L'assemblage bioinformatique peut être à l'origine de chimères. Dans le cas du projet « racines », nous n'avons pas été confrontés à ce problème. Ce phénomène intervient lorsqu'un insert est constitué par la juxtaposition de deux EST de gènes différents. Cette séquence chimérique effectue un lien entre l'assemblage des deux gènes. Sur la représentation graphique de l'interface, il apparaîtrait deux blocs d'EST alignés reliés par une seule séquence. Dans ce cas, les résultats de Blast montreraient des homologies avec des gènes différents pour les deux blocs.

Le TC57901 du TIGR est un exemple de ce type de configuration. Il est consultable à l'URL http://www.tigr.org/tigr-scripts/tgi/tc_report.pl?tc=TC57901&species=pinus. Les 80 accessions sont organisées en deux blocs. L'alignement des séquences du premier bloc qui comprend la plupart des accessions couvre les 675 premières paires de bases du consensus. Le second alignement couvre la région des bases 700 à 2263. Une seule séquence permet de relier ces deux alignements. Les séquences du premier bloc ont été inversées pour l'alignement tandis que l'orientation de celles du second bloc a été conservée. La recherche d'homologie donne une Déhydrine pour le second bloc, la base 807 correspondant à la première base du match. Le premier bloc présente une homologie avec une « unknown protein » d'*Arabidopsis* en sens complémentaire, la base 576 correspondant à la première base de la protéine homologue. L'annotation des EST de ce TC a été donnée à partir du résultat d'homologie du second bloc qui ne regroupe que 33 des 80 accessions.

Pour procéder à l'annotation, le résultat du Blast doit donc être analysé en tenant compte de la base de données d'où il provient, l'étendue de l'homologie doit être notée ainsi que la concordance des résultats de Blast des EST et leur position dans l'alignement. Si les informations disponibles ne sont pas suffisantes, une nouvelle recherche d'homologie peut

être lancée. Le champ d'annotation peut alors être alimenté. Avec l'intégration de nos séquences de pin maritime dans le projet d'assemblage « *Pinus* » du TIGR, une annotation de nos séquences intégrerait aujourd'hui les bases de données *Pinus* du TIGR.

Au cours de la phase d'annotation, il arrive fréquemment que des résultats de blast ayant de bons scores correspondent à des fonctions d'intitulés différents alors qu'il s'agit en réalité de la même enzyme. Cela vient du fait que les dénominations des enzymes ont évolué au cours des années. A titre d'exemple, la SAM synthase (EC 2.5.1.6) ou S-adenosylmethionine synthase est aussi appelée AdoMet synthetase et son nom officiel est Methionine adenosyltransferase. Nous avons donc exporté l'ensemble des annotations réalisées dans un tableur Excel et recherché le nom officiel des enzymes sur le site web http://www.expasy.org afin d'uniformiser les annotations. Cette étape très laborieuse s'est avérée extrêmement utile pour les études d'expression *in silico*. Nous avons également pu vérifier la cohérence des annotations des contigs d'un même cluster.

Nous avons vu dans les chapitres précédents que certaines séquences ne présentaient pas d'homologie avec des gènes de fonctions connues. Dans les bases de données nous pouvons rencontrer les annotations « putative protein », « unknown protein » ou encore « hypothetical protein ». L'annotation « putative protein » est généralement rencontrée pour des séquences de transcrits (ex : EST) mais pour lesquelles la traduction en protéine n'est pas montrée. Les séquences annotées « unknown protein » correspondent normalement à des transcrits traduits en protéines mais dont la fonction est inconnue. Le terme « hypothetical protein » est plus ambiguë et peut parfois prendre la définition donnée pour « putative protein » ou parfois celle de « unknown protein ». Certaines accessions sont ainsi notées « Hypothetical 10.4kDa protein (Expressed protein) » ce qui indique que la protéine était bien exprimée lors de sa découverte mais de fonction inconnue et d'autres accessions « Hypothetical 85.6 kDa protein » peuvent provenir d'une modélisation de la protéine potentiellement produite.

Nous avons donc choisi dans notre cas de désigner par « putative protein » ou « no hit » les séquences ne présentant pas d'homologie (> 1e-05) à l'issue de la stratégie en cascade. Nous sommes en présence de séquences d'EST, il s'agit donc de séquences de transcrits. Il est probable mais pas certain que la séquence soit traduite en protéine. Dans le cas où une homologie avec d'autres transcrits annotés « hypothetical protein » est trouvée, nous avons considéré deux situations. Si une précision était indiquée sur la nature de la protéine hypothétique (ex : 28.2kDa) nous avons conservé l'annotation de l'homologue. Dans le cas

contraire, nous désignerons les séquences comme « homologue to hypothetical protein » ou « unknown protein ». Cette notation est toutefois discutable et dans le cas où une publication des annotations sur les bases publiques était envisagée, il serait préférable de donner une annotation du type : « similar to » suivi de l'annotation de la séquence homologue pour lever toute ambiguïté. Il est également possible de suivre le TIGR qui a choisit d'ajouter un qualificatif avant l'annotation pour préciser la qualité de l'homologie obtenue (Tab. D-4).

La répartition des annotations des EST (Fig. D-14) montre que la proportion d'EST annotés, sans homologie et homologues à des protéines hypothétiques est similaire pour les deux conditions au sein d'un même tissu. Bien que la proportion des EST issus des parties aérienne soit très faible (Fig. D-7), on peut toutefois remarquer la plus forte proportion de séquences ne présentant pas d'homologie dans les EST de racines. Une hypothèse pour expliquer cette différence serait que la quantité d'EST générés étant beaucoup plus faible pour les parties aériennes, la probabilité de séquencer des transcrits abondant donc mieux connus est plus forte. Il se peut également que les transcrits de racines soient moins bien décrits que ceux des parties aériennes.

D.2.3.2 Assignation des catégories fonctionnelles

L'objectif de la classification est d'organiser les transcrits des gènes identifiés dans un nombre limité de catégories fonctionnelles et de suivre un vocabulaire contrôlé pour les termes utilisés dans la description des fonctions des gènes. Les trois ressources principales, faisant référence et intégrées à l'interface d'annotation (Fig. D-12), sont les bases FunCat (The Functional Catalogue), GO (Gene Ontology) et KEGG (Kyoto Encyclopedia of Genes and Genomes). Les bases FunCat et GO ne sont pas des bases de séquences, il s'agit de bases permettant de rechercher une terminologie associée aux termes recherchés suivant celle du MIPS (Munich Information Center for Protein Sequences). La base KEGG en revanche repose sur le principe du Blast. Une recherche est lancée sur une base protéique contenant uniquement des protéines très bien identifiées, annotées dont les fonctions et les positions dans les voies métaboliques sont connues. L'utilisation du KEGG peut être automatisé mais le nombre de protéines dans la base est restreint et ne permet d'assigner une classe fonctionnelle qu'aux enzymes les plus connues. Notre choix s'est donc porté sur la classification GO qui s'articule autour de 3 niveaux d'organisation : la fonction moléculaire, le processus biologique et la localisation cellulaire. A chaque fois que cela était possible, les champs ont été renseignés mais une même enzyme peut avoir des fonctions multiples dans la cellule et

Tableau D-4 : Qualificatifs utilisés par le TIGR pour l'annotation.

critère	qualificatif
100% > % d'identité >= 90%	*homologue to*
90% > % d'identité >= 70%	*similar to*
% d'identité < 70%	*weakly similar to*
couverture protéine > 98%	*complete*
couverture protéine <= 98%	*partial*

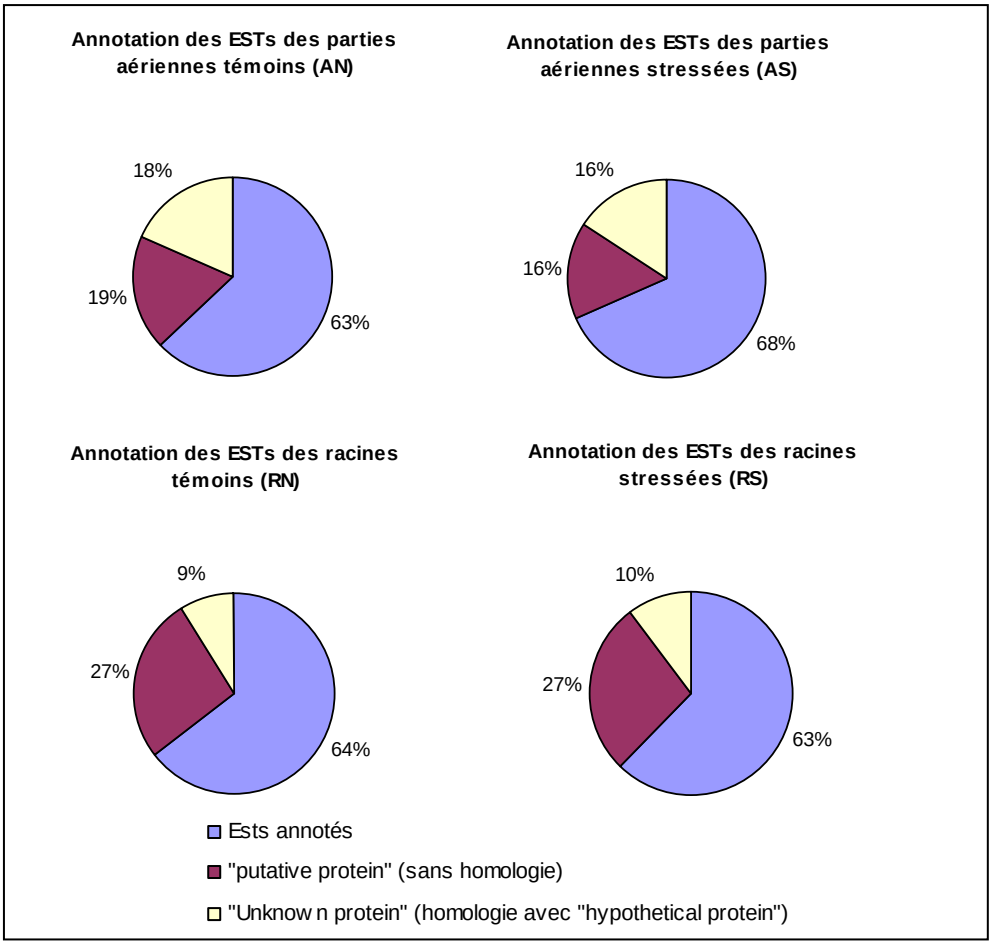


Figure D-14 : Proportion d'ESTs annotés, sans homologie et homologues à des protéines hypothétiques pour les deux tissus et les deux conditions dans le projet « réponse au stress ».

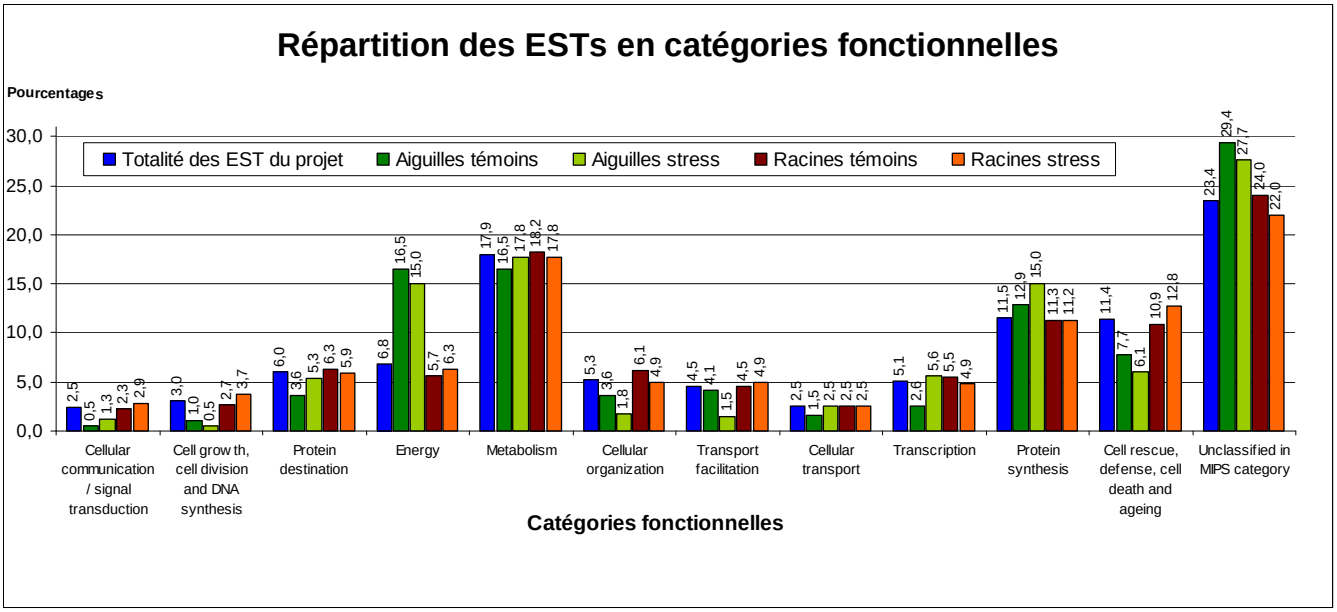


Figure D-15 : Répartition des ESTs en catégories fonctionnelles.

l'assignation à une classe particulière ne peut pas toujours être tranchée. Il reste donc une part de subjectivité de l'annotateur dans cet exercice.

La répartition des annotations de GO en 12 grandes catégories de l'ensemble des EST pour lesquels une annotation a pu être donnée, –c'est-à-dire mise à part les EST sans homologie et homologues à des « hypothetical protein »- est indiquée à la figure D-15. Comme pour la proportion des EST annotés « putative protein » ou « unknown » (Fig. D-14), la répartition des EST dans les différentes catégories fonctionnelles est très proche pour les deux conditions d'un même tissu. Par ailleurs, la différence de répartition dans les catégories est nette entre les deux tissus. Les EST de la classe énergie sont beaucoup plus abondants dans les parties aériennes. Cela est dû à l'abondance des transcrits homologues à des protéines du photosystème, de la chlorophylle et de la Rubisco. La sur-représentativité de cette classe se fait au détriment des classes « communication cellulaire et transduction du signal », « Organisation cellulaire », « Transport » et « Trafic intracellulaire ».

Pour les racines, la plus grande part des EST soit près de 23% n'ont pu être affectés à une classe spécifique. La classe la plus abondante est ensuite celle du métabolisme, suivie de la classe correspondant à la synthèse des protéines. La classe regroupant les gènes de mort cellulaire, de défense et de réponse au stress atteint plus de 11% des EST de racines. Les classes les moins représentées concernent la communication cellulaire et la transduction du signal, la croissance et division cellulaire et le trafic intracellulaire.

D.2.4 Northern électronique

Cette méthode d'estimation du niveau d'expression des transcrits entre plusieurs banques a été appliquée aux banques de racines stressées et contrôles. Cette approche repose sur le nombre d'EST séquencés pour un gène donné. Pour estimer correctement l'expression d'un gène, il faudra être en mesure d'identifier quels sont les EST issus des transcrits de ce gène. Une première hypothèse possible est de considérer que chaque contig constitue un gène différent, ce qui soulève trois incertitudes majeures :

- Nous avons vu que l'assemblage repose sur la similarité des séquences et les zones de recouvrement entre EST pour la génération des clusters et des contigs. Il est donc possible que les transcrits d'un même gène soient répartis dans deux contigs distincts si aucune séquence ne permet de relier les deux alignements.
- Concernant les singletons, une situation similaire peut se présenter. Une séquence est classée dans les singletons si elle n'a pu être alignée avec d'autres séquences. Cela ne signifie pas pour autant que nous soyons en présence d'un gène supplémentaire. Les

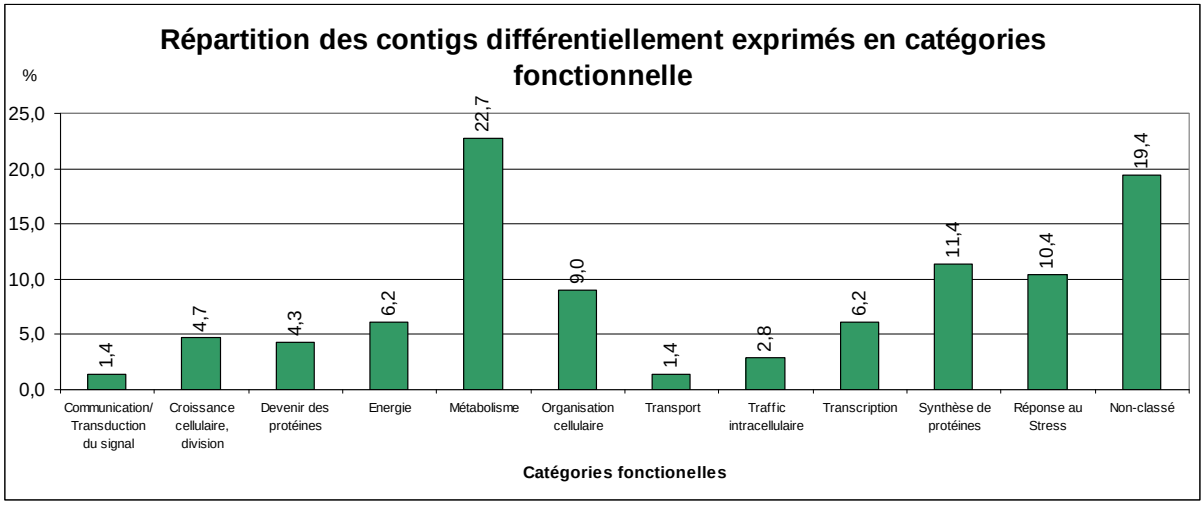


Figure D-16 : Répartition en catégories fonctionnelles des contigs détectés comme différentiellement exprimés *in silico*.

EST des transcrits d'un gène peuvent être regroupés au sein d'un contig mais ne couvrir que sa partie 5'. Un transcrit pour lequel l'EST est positionné en 3' sera alors considéré comme un singleton.

- La situation inverse est également possible, les EST de différents membres d'une famille multigénique peuvent être groupés sous le même contig avec d'autant plus de risques que la séquence en 3' n'est pas connue. Dans ces situations, l'estimation de l'expression du gène par comptage sera biaisée.

L'analyse statistique par IDEG6 a été réalisée en utilisant comme jeu de données les EST issus de racines en éliminant l'intégralité des singletons et des éventuelles séquences multiples d'un même clone. Le comptage des EST de chaque banque (stressé vs contrôle) a été fait en se plaçant au niveau du contig. Dans ce cas, nous considérons qu'un contig correspond à un gène, et aucune distinction n'est faite entre les EST du consensus primaire et ceux des éventuels consensus alternatifs. Le tableau Ax-1 (voir en annexe) donne la liste des 279 contigs détectés comme différentiellement exprimés par au moins un des tests statistiques sur les 2022 contigs analysés. Ils sont classés des plus fortes surexpressions dans la librairie 1 (racines contrôle) aux plus fortes sous expressions. La première constatation concerne le faible nombre d'EST par banque constituant les contigs. On peut se poser la question de la représentativité des données pour un décompte qui dépasse rarement 6 EST. Lorsque la normalisation (sur le nombre d'EST de chaque banque) est appliquée et qu'aucun EST n'a été obtenu dans une des banques, les différences d'expression sont artificiellement très élevées. Il faut donc prendre les résultats avancés avec précaution.

La figure D-16 nous présente les catégories fonctionnelles auxquelles les contigs détectés appartiennent. Par comparaison avec la répartition en catégories fonctionnelles de l'ensemble des EST (Fig. D-14), il apparaît que la proportion de « gènes » des classes « métabolisme », « organisation cellulaire » et « croissance cellulaire/ division » est plus élevée dans les « gènes » détectés comme différentiellement exprimés. A l'inverse, les classes « devenir des protéines », « transport » et « non-classé » sont sous-représentées.

Plusieurs contigs ont des annotations pouvant avoir un rôle dans la réponse au stress hydrique. Cependant, des transcrits de fonctions généralement mises en évidence lors du stress hydrique sont ici détectés comme plus abondants dans la banque de racines contrôles.

Nous avons l'Inositol 1-alpha-galactosyltransférase (EC 2.4.1.123) (CT469) avec un ratio de plus de 25 fois supérieur dans le contrôle. Cet enzyme permet la conversion du myo-

Inositol en Galactinol qui est un précurseur de la voie des RFO qui, comme nous l'avons vu, peuvent être accumulés lors du stress. La surexpression chez les racines contrôles ne va donc pas dans ce sens. C'est aussi le cas des contigs (CT477) et (CT219), avec respectivement des facteurs de 14 et 7 fois, homologues à des facteurs de liaison aux éléments de réponse à l'éthylène (EREBF). Une LEA (CT276) semble aussi régulée par un facteur 2.45 en faveur des contrôles. La prévention des oxydations est représentée par plusieurs peroxydases (CT168, CT1176 et CT235) et une L-ascorbate peroxydase (CT529) ainsi qu'une Flavonoide 3',5'-hydroxylase (CT1039) et « Metallothionein-like protein EMB30 » (CT431). Si ces fonctions ne sont pas des marqueurs spécifiques de la réponse au stress hydrique, il est courant de les mettre en évidence en situation de stress hydrique qui engendre aussi un stress oxydatif (Rabbani *et al.* 2003). Les peroxydases, comme les laccases peuvent également intervenir dans la polymérisation des monolignols en lignine et donc avoir un rôle dans la synthèse des parois. La sous-expression lors du stress d'un transcrit homologue à une « Metallothionein-like protein EMB30 » conforte les résultats de Dubos *et al.* sur les racines de pins maritime en stress osmotique (Dubos and Plomion 2003). Nous pourrons enfin citer une Polcalcin (T56), un précurseur d'Alpha-expansin (CT859) et une Défensine (CT166) qui sont des acteurs de la réponse au stress pour respectivement la transduction du signal, l'extensibilité des parois et le système de défense contre les pathogènes.

A l'opposé, parmi les contigs comportant plus de séquences de la banque stressée, nous trouvons également une liste de fonctions intéressantes concordant plus avec le sens de variation attendu en condition de stress hydrique. Parmi les plus forts ratios d'expression, nous retrouvons une LEA (CT275), des protéines de liaison au calcium (CT59)(CT60) et une Flavonoide 3',5'-hydroxylase (CT346). Les contigs en rapport avec la réponse à l'oxydation et à la détoxication sont plus abondants ce qui concorderait avec l'hypothèse d'un stress oxydatif plus fort chez les plants en stress hydrique. Les représentants sont des « Metallothionein-like protein » (CT312 et CT314), des sous-unités du complexe endopeptidase du Proteasome (CT1596 et CT1609) ainsi qu'un précurseur de « cathepsin B-like cysteine proteinase » (CT1841) qui est une protéine localisée dans le lysosome. Des enzymes intervenant à plusieurs niveaux dans la voie de synthèse des Flavonoïdes sont également détectés avec une 3',5'-hydroxylase (EC 1.14.13.88) (CT346, CT342, CT344) et une Flavonol 3-O-glucosyltransférase 5 (EC 2.4.1.91) (CT1670) (http://www.genome.ad.jp/dbget-bin/show_pathway?map00941+1.14.13.88+2.4.1.91). Dubos *et al.* avaient également trouvé une accumulation de transcrits homologues à la Flavonoïde 3',5'-hydroxylase (EC 1.14.13.21). Nous trouvons aussi des fonctions en rapport avec la structure des parois, comme une