

THESIS / THÈSE

MASTER EN BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE ET CELLULAIRE

Contribution à l'amélioration de la résistance aux pathogènes chez la chicorée industrielle à l'aide de marqueurs moléculaires de classe 1.

Hespeels, Boris

Award date:
2009

Awarding institution:
Université de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



**FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX
NAMUR**

Faculté des Sciences

**Contribution à l'amélioration de la résistance aux pathogènes chez la chicorée
industrielle à l'aide de marqueurs moléculaires de classe 1.**

**Mémoire présenté pour l'obtention
du grade académique de master en biochimie et biologie moléculaire et cellulaire**

Boris HESPEELS

Janvier 2009

Contribution à l'amélioration de la résistance aux pathogènes chez la chicorée industrielle à l'aide de marqueurs moléculaires de classe 1.

HESPEELS Boris

Résumé

La culture de chicorée industrielle, *Cichorium intybus*, est affectée de façon significative par deux champignons : *Eryiphe cichoracearum* et *Puccinia cichorii*. Afin d'améliorer la résistance de la chicorée contre les pathogènes, nous avons participé aux étapes préliminaires nécessaires au développement des marqueurs moléculaires que de type Single Nucléotide Polymorphism (SNP) utilisables en sélection variétale. Pour ce faire, notre stratégie a été de caractériser le polymorphisme présent à l'origine de la sélection de la chicorée industrielle pour préparer des outils utiles pour de futures études d'association.

Les analyses se sont focalisées sur une centaine de gènes candidats. Notre intérêt s'est porté sur les gènes NBS-LRR, sur une série de vingt-et-une séquences de chicorée potentiellement orthologues de séquences de laitue associées à la résistance aux pathogènes (carte génétique du CGPDB), sur des homologues des gènes WAK et d'autres gènes de résistance impliqués dans les voies de réponses aux pathogènes. Ces gènes candidats ont été mis en évidence dans un jeu d'unigènes de chicorée suite à de nombreuses analyses bioinformatiques basées sur des recherches d'homologies, des recherches de domaines ainsi que des recherches bibliographiques. La génération et l'exploitation poussée de banques de données locales se basant sur les séquences d'Astéracée disponibles en ligne a également été réalisée.

Pour chaque candidat, des amorces ont été définies, vingt-sept séquences ont pu être amplifiées et la recherche de polymorphisme de type SNP dans une collection de cultivars anciens a été entreprise expérimentalement pour dix d'entre elles grâce à l'EcoTILLING.

Au terme de ce mémoire, nous disposons d'une liste de gènes candidats, de paires d'amorces, de procédures et de résultats préliminaires concernant le polymorphisme d'une dizaine de gènes impliqués dans la réponse face aux pathogènes. Ces informations pourront, à terme, servir à l'amélioration de la chicorée industrielle.

Mémoire de master en biochimie et biologie moléculaire et cellulaire

Janvier 2009

Promoteur: P. Van Cutsem et N. Dauchot (Co-Promoteur)

Je remercie tout d'abord Monsieur le professeur P. Van Cutsem de m'avoir ouvert les portes de son laboratoire durant cette année de mémoire. Je le remercie pour son investissement et les discussions critiques et constructives que nous avons eues ensemble sur l'élaboration de ce projet. Je lui exprime toute ma reconnaissance.

Je remercie également Nicolas Dauchot dont les talents multiples en marqueurs moléculaires et en bioinformatique ont permis la réalisation de ce travail. Nicolas, ton soutien quotidien, ta créativité, tes qualités de bidouilleur en tous genres ont permis de résoudre bien des problèmes et ont fortement contribué à développer une approche scientifique innovante. Je salue ton enthousiasme et ta disponibilité tout au long de cette année.

Je remercie chacun des membres de l'URBV, Johan Messiaen (je me souviendrai de l'ouverture et de la fermeture du stomate... dans les grandes lignes...), Céline Muys, Aurélien Boland, Juan Carlos Cabrera Pino (pour les découvertes musicales), Ana Maria Bautista Salas, Houda Khamis, Raffael Buonatesta, Pierre Cambier (pour sa grande disponibilité), Catherine Demazy (d'avoir supporté les brots et bizarreries de son voisin de bureau), Céline Léonet, Adrien, et Françoise Lonnoy (pour son aide précieuse). Merci pour la bonne ambiance que vous apportez tous et pour votre accueil chaleureux !

Je remercie aussi Sophie, Stéphanie et Anne Florence avec qui j'ai pu partager les moments heureux et plus difficiles du mémoire. Merci Sophie pour ton aide précieuse lors des manipulations en laboratoire ! Merci Stéphanie de m'avoir partagé avec enthousiasme tes connaissances sur le monde des WAK.

Merci Judith pour ton soutien et l'intérêt que tu as porté à la réalisation de ce travail. Je remercie également les membres de ma famille ainsi que mes amis pour leur curiosité et leurs encouragements.

Je remercie enfin les membres du jury pour l'attention et l'intérêt critique qu'ils porteront à ce travail.

Liste des abréviations

ADN	<i>acide désoxyribonucléique</i>
ADNc	<i>ADN codant</i>
ADNg	<i>ADN génomique</i>
ADP	<i>adénosine diphosphate</i>
AFLP	<i>amplified fragment lenght polymorphism</i>
Amp	<i>ampiciline</i>
APAF-1	<i>apoptotic protease-activating factor-1</i>
ARC	<i>APAF-1 R genes Ced4</i>
ASO	<i>allel specific oligonucleotide</i>
ATP	<i>adénosine triphosphate</i>
Avr	<i>gène d'avirulence</i>
BLAST	<i>Basic Local Alignment Search Tool</i>
CC	<i>coiled-coil</i>
CED-4	<i>Caenorhabditis elegans death-4 protein</i>
CGPDB	<i>Compositae genome project database</i>
COI1	<i>coronatine insensitive 1</i>
ddNTP	<i>didésoxyribonucléotide</i>
dNTP	<i>désoxyribonucléotide</i>
DO	<i>densité optique</i>
EDS1	<i>enhanced disease susceptibility 1</i>
EGF	<i>epidermal growth factor</i>
ERF	<i>ethylene response factor</i>
EST	<i>expressed sequence tag</i>
ET	<i>éthylène</i>
ETI	<i>effector triggered immunity</i>
HR	<i>réponse hypersensible</i>
INA	<i>acide 2,6-dichloroisonicotinique</i>
INDEL	<i>insertion deletion</i>
IPTG	<i>isopropyl-beta-thio-galactoside</i>
JA	<i>jasmonate</i>
JAR1	<i>jasmonate response locus 1</i>
JIN1	<i>jasmonate insensitive 1</i>
LB	<i>luria broth</i>
LRR	<i>leucin rich repeat</i>
MAMP	<i>microbe associated molecular patterns</i>
MIMP	<i>microbe induced molecular patterns</i>
MKS1	<i>MAPK4 Substrate 1</i>
MPK4 ou MAPK4	<i>MAP KINASE 4</i>
NB-ARC	<i>nucleotide binding APAF-1 R genes Ced4</i>
NBS	<i>nucleotide binding site</i>
NBS-LRR	<i>nucleotide binding site-leucin rich repeat</i>

NCBI	<i>National Center for Biotechnology Information</i>
NDR1	<i>nonrace-specific disease resistance</i>
NLS	<i>nuclear localisation sequence</i>
NPR1	<i>non expressor of pathogenesis- related proteins1</i>
PAD4	<i>phytoalexin deficient 4</i>
PAMP	<i>pathogen associated molecular patterns</i>
pb	<i>paires de bases</i>
PCI	<i>phenol: chlorophorm: alcool isoamylique 25:24:1</i>
PCR	<i>polymerase Chain Reaction</i>
PDF	<i>plant defensin</i>
PR	<i>pathogenesis related protein</i>
PTI	<i>PAMP triggered immunity</i>
QTL	<i>quantitative trait loci</i>
<i>R gene</i>	<i>gène de résistance</i>
RFLP	<i>restriction fragment lenght polymorphism</i>
RLK	<i>receptor like kinase</i>
SA	<i>acide salicylique</i>
SAG101	<i>senescence associated gene 101</i>
SCF	<i>skip-cullin-F-box</i>
SNP	<i>single nucleotide polymorphism</i>
SRR	<i>single strand repeat</i>
TAIR	<i>The Arabidopsis Information Resource</i>
TILLING	<i>targeting induced local lesions in genomes</i>
TIR	<i>Toll like interleukin receptor</i>
TTSS	<i>type three secretion system</i>
UTR	<i>untranslated region</i>
UV	<i>ultraviolet</i>
WAK	<i>wall associated kinases</i>
WAKL	<i>WAK-Like</i>
WAKR	<i>WAK-Related</i>

Table des matières

INTRODUCTION

1	La chicorée industrielle.....	1
1.1	Présentation générale.....	1
1.2	Maladies foliaires de la chicorée industrielle	1
2	Interactions plantes- pathogènes	2
2.1	Récepteurs et gènes de résistance	4
2.1.1	Les récepteurs à protéines kinases	4
2.1.2	Les récepteurs WAK : Wall Associated Kinase	5
2.1.3	Les gènes de résistance nucleotide-binding site leucine rich repeat (NBS-LRR)	5
2.2	Réponse hypersensible et résistance systémique acquise	8
2.3	Voies de réponses aux pathogènes	9
2.3.1	La voie de l'acide salicylique (SA)	9
2.3.2	La voie de l'éthylène (ET) et du méthyl jasmonate (JA).....	10
2.3.3	Cross talk entre SA, ET et JA.....	11
3	Les marqueurs moléculaires.....	12
3.1	Recherche de SNPs.....	14
3.1.1	L'EcoTILLING	14
3.1.2	Typage de SNP : analyse microarray.....	16
3.2	L'approche gène candidat et l'utilisation de marqueurs moléculaires : génétiq ue d'association.....	17
3.3	Objectifs du mémoire	18

MATÉRIELS ET MÉTHODES

1	Recherches Bioinformatiques	20
1.1.	Séquences utilisées dans le cadre des recherches bioinformatiques.....	20
1.2.	Format des séquences	20
1.3.	Logiciels installés en local.....	20
1.3.1.	SeqTools 8.4	20
1.3.2.	Vector NTI 10	20
1.3.3.	TreeView 1.6.6	20
1.3.4.	TreeIllustrator	20
1.3.5.	Blast2Go.....	20
1.4.	Logiciels Online.....	21
1.4.1.	Primer3.....	21
1.4.2.	NCBI BLAST	21
1.4.3.	Pfam 23.0	21
1.4.4.	Interproscan v4.1	21
1.4.5.	Spidey	21
1.5.	L'analyse reciprocal best BLAST hit	21
2.	Recherches expérimentales.....	22
2.1.	Matériel végétal	22

2.2. Extraction de l'ADN de chicorée	22
2.3. Définition et validation des amorces PCR.....	22
2.4. Bactéries thermocompétentes TOP 10 F'	23
2.5. Transformation des bactéries thermocompétentes TOP10F'	23
2.6. Préparation du plasmide	24
2.6.1. Maxiprep : QIAGEN Plasmid Maxi kit #12162	24
2.7. Restriction du plasmide	25
2.8. Purification du plasmide restreint et des produits PCR avant clonage.....	25
2.9. Restriction ligation simultanée	26
2.10. Screening	26
2.11. EcoTilling	27
2.11.1. Etape de dénaturation hybridation	27
2.11.2. Etape de digestion	27
2.11.3. Etape de précipitation de l'ADN.....	27
2.12. Préparation séquençage	27
2.12.1. Miniprep.....	27
2.12.2. Agarose Low Melting	28
2.13. Séquençage	29
2.13.1. Réaction de séquençage	29
2.13.2. Lavage et précipitation de l'ADN	29
2.13.3. Préparation des échantillons pour le séquenceur	29

RÉSULTATS ET DISCUSSIONS

1 Recherche de gènes candidats : approche bioinformatique	31
1.1 Recherche d'orthologues entre la laitue et la chicorée	31
1.2 Recherche de gènes candidats de type NBS-LRR et mise en évidence de gènes associés à la pathogenèse.....	32
1.2.1 Gènes de résistance contenant un domaine NBS chez la chicorée	32
1.2.2 Recherche de gènes NBS-LRR au départ de la banque d'A. thaliana.....	33
1.2.3 Recherche de gènes candidats à partir de séquences de tournesol associées aux gènes NBS-LRR.....	34
1.2.4 Data mining : recherche de gènes associés à la pathogenèse dans la banque d'unigènes de chicorée à l'aide d'Interproscan	35
1.2.5 Data mining : mise en évidence de gènes candidats à partir de la banque CGPDB.....	37
1.3 Recherche de gènes clefs dans les voies de défense de la plante : acide salicylique, jasmonate et éthylène	37
1.4 Recherche de gènes WAK (Wall Associated Kinase) chez la chicorée	38
1.4.1 Recherche de gènes WAK sur base de séquences A. thaliana	38
1.4.2 Recherche de WAK par data mining.....	39
1.5 Synthèse de la recherche de gènes candidats impliqués dans les réponses de défenses.....	39
2 Caractérisation du polymorphisme chez C. intybus.....	39
2.1 Constitution de la « Core Collection »	40
2.2 Etape de clonage des gènes candidats	40
2.3 EcoTILLING	41
2.3.1 Approche en POOL	41
2.3.2 Résultats des analyses EcoTILLING sur pool	41
2.3.3 Comparaison analyse individuelle et en pool : CS230 gène NBS-LRR.....	42

2.3.4	Analyse EcoTILLING avec référence	43
2.4	Résultats de séquençage	43
2.4.1	CS58, β 1,3 glucanase.....	43
2.4.2	CS60, homologue NBS-LRR.....	44
2.4.3	CS62, homologue de Avr9/Cf-9 rapidly elicited protein 75	45
2.4.4	CS76, homologue Cf-5 disease resistance protein	45
2.4.5	CS78, homologue Glycine max putative Hs1pro-1-like receptor	46
2.4.6	CS160, homologue ethylene responsive element binding factor 4	47
2.4.7	CS230, gène NBS-LRR	47
2.4.8	CS242, gène NBS-LRR	48
2.4.9	CS284, gène WAK-related.....	49
2.4.10	CS286, gène WAK-related.....	49
2.5	Étude du polymorphisme actuel chez la chicorée industrielle, comparaison avec les résultats obtenus sur la core collection.....	50

CONCLUSIONS ET PERSPECTIVES

Conclusions et perspectives.....	52
----------------------------------	----

BIBLIOGRAPHIE

Bibliographie	56
---------------------	----

Ref	Name	Target	ADNc	ADNg	PCR condition	PCR
CS58	1CILE0093PatFw-spe1	beta-1,3-glucanase	353 bp	369	58°C 45" 35X	ok
CS59	1CILE0093PatRv-sal1	beta-1,3-glucanase				
CS60	1CILE0200PatFw-spe1	disease resistance protein (CC-NBS-LRR class)	450 bp	1000	55°C 45" 35X	ok
CS61	1CILE0200PatRv-sal1	disease resistance protein (CC-NBS-LRR class)				
CS62	1CILE0239PatFw-spe1	putative lettuce orthologs associated with resistance	365 bp	363	58°C 45" 35X	ok
CS63	1CILE0239PatRv-sal1	putative lettuce orthologs associated with resistance				
CS74	2CILE0090PatFw-spe1	putative lettuce orthologs associated with resistance	369 bp	800	58°C 45" 35X	ok
CS75	2CILE0090PatRv-sal1	putative lettuce orthologs associated with resistance				
CS76	2CILE0393PatFw-spe1	Leucine Rich Repeat family protein, similar to HcrVf2	431 bp	933	55°C 45" 35X	ok
CS77	2CILE0393PatRv-sal1	Leucine Rich Repeat family protein, similar to HcrVf2				
CS78	2CILE0010PatFw-spe1	disease resistance protein (TIR-NBS-LRR class)	450 bp	450	55°C 45" 35X	ok
CS79	2CILE0010PatRv-sal1	disease resistance protein (TIR-NBS-LRR class)				
CS90	3CILE0092PatFw-spe1	putative lettuce orthologs associated with resistance	377 bp	400	58°C 45" 35X	ok
CS91	3CILE0092PatRv-sal1	putative lettuce orthologs associated with resistance				
CS92	3CILE0149PatFw-spe1	putative lettuce orthologs associated with resistance	351 bp	400	58°C 45" 35X	ok
CS93	3CILE0149PatRv-sal1	putative lettuce orthologs associated with resistance				
CS94	3CILE0317PatFw-spe1	putative lettuce orthologs associated with resistance	361 bp	450	58°C 45" 35X	ok
CS95	3CILE0317PatRv-sal1	putative lettuce orthologs associated with resistance				
CS96	3CILE0362PatFw-spe1	putative lettuce orthologs associated with resistance	352 bp	800	58°C 45" 35X	ok
CS97	3CILE0362PatRv-sal1	putative lettuce orthologs associated with resistance				
CS112	4CILE0279PatFw-spe1	putative lettuce orthologs associated with resistance	364 bp	380	58°C 45" 35X	ok
CS113	4CILE0279PatRv-sal1	putative lettuce orthologs associated with resistance				
CS114	4CILE0402PatFw-spe1	disease resistance protein (CC-NBS-LRR class)	403 bp	non	58°C 45" 35X	
CS115	4CILE0402PatRv-sal1	disease resistance protein (CC-NBS-LRR class)				
CS126	5CILE0091PatFw-spe1	putative lettuce orthologs associated with resistance	355 bp	400	58°C 45" 35X	ok
CS127	5CILE0091PatRv-sal1	putative lettuce orthologs associated with resistance				
CS128	5CILE0114PatFw-spe1	MAPKK Kinase (EDR1)	471 bp			
CS129	5CILE0114PatRv-sal1	MAPKK Kinase (EDR1)				
CS130	5CILE0114PatFw-spe1	MAPKK Kinase (EDR1)	426 bp			
CS131	5CILE0114PatRv-sal1	MAPKK Kinase (EDR1)				
CS132	5CIM5752PatFw-spe1	putative lettuce orthologs associated with resistance	391 bp	600	58°C 45" 35X	ok
CS133	5CIM5752PatRv-sal1	putative lettuce orthologs associated with resistance				
CS160	7CIM6982PatFw-spe1	putative lettuce orthologs associated with resistance	363 bp	400	58°C 45" 35X	ok
CS161	7CIM6982PatRv-sal1	putative lettuce orthologs associated with resistance				
CS198	9CILE0310PatFw-spe1	putative lettuce orthologs associated with resistance	388 bp	600	58°C 45" 35X	ok
CS199	9CILE0310PatRv-sal1	putative lettuce orthologs associated with resistance				
CS200	9CILE0329PatFw-spe1	WRKY family transcription factor	397 bp	400	55°C 45" 35X	ok
CS201	9CILE0329PatRv-sal1	WRKY family transcription factor				
CS202	9CILE0456PatFw-spe1	putative lettuce orthologs associated with resistance	378 bp	500	58°C 45" 35X	ok
CS203	9CILE0456PatRv-sal1	putative lettuce orthologs associated with resistance				
CS212	3CILE0473patFw-spe1	glycosyl hydrolase family 17 (beta-1,3-glucanase)	383 bp	400	58°C 45" 35X	ok
CS213	3CILE0473patRv-sal1	putative lettuce orthologs associated with resistance				
CS216	9CIM4990PatFw-spe1	putative lettuce orthologs associated with resistance	350 bp	400	58°C 45" 35X	ok
CS217	9CIM4990PatRv-sal1	putative lettuce orthologs associated with resistance				
CS218	NCBI1PatPatFw-spe1	AY153807.1	204 bp			
CS219	NCBI1PatPatRv-sal1	AY153807.1				
CS220	NCBI2 PatFw-spe1	36078385.VIS	399 bp			
CS221	NCBI2 PatRv-sal1	36078385.VIS				
CS222	NCBI3 PatFw-spe1	AY153812.1	263 bp			
CS223	NCBI3 PatRv-sal1	AY153812.1				
CS224	NCBI4 PatFw-spe1	AY153815.1	236 bp			
CS225	NCBI4 PatRv-sal1	AY153815.1				
CS226	NCBI5 PatFw-spe1	AY153817.1	207 bp			
CS227	NCBI5 PatRv-sal1	AY153817.1				
CS228	NCBI6 PatFw-spe1	AY153822.1	423 bp	423	58°C 45" 35X	ok
CS229	NCBI6 PatRv-sal1	AY153822.1				
CS230	NCBI7 PatFw-spe1	AY153825.1	424 bp	424	58°C 45" 35X	ok
CS231	NCBI7 PatRv-sal1	AY153825.1				
CS242	TIRNBSLRR1-SpeI	08467385.VIS	416 bp	400		ok
CS243	TIRNBSLRR1-Sal1HF	08467385.VIS				
CS244	TIRNBSLRR2-SpeI	13096385.VIS	300 bp			
CS245	TIRNBSLRR2-Sal1HF	13096385.VIS				
CS246	TIRNBSLRR3-SpeI	15860385.VIS	377 bp			
CS247	TIRNBSLRR3-Sal1HF	15860385.VIS				
CS248	TIRNBSLRR4-SpeI	17666385.VIS	371 bp			
CS249	TIRNBSLRR4-Sal1HF	17666385.VIS				
CS250	TIRNBSLRR5-SpeI	22885385.VIS	423 bp			
CS251	TIRNBSLRR5-Sal1HF	22885385.VIS				
CS252	TIRNBSLRR6-SpeI	25510385.VIS	368 bp			
CS253	TIRNBSLRR6-Sal1HF	25510385.VIS				
CS254	TIRNBSLRR7-SpeI	26173385.VIS	378 bp			
CS255	TIRNBSLRR7-Sal1HF	26173385.VIS				
CS256	TIRNBSLRR8-SpeI	28440385.VIS	378 bp			
CS257	TIRNBSLRR8-Sal1HF	28440385.VIS				
CS258	TIRNBSLRR9-SpeI	28763385.VIS	369 bp			
CS259	TIRNBSLRR9-Sal1HF	28763385.VIS				
CS260	TIRNBSLRR10-SpeI	28802385.VIS	415 bp			
CS261	TIRNBSLRR10-Sal1HF	28802385.VIS				
CS262	TIRNBSLRR11-SpeI	30486385.VIS	407 bp			
CS263	TIRNBSLRR11-Sal1HF	30486385.VIS				

CS264	TIRNBSLRR12-SpeI	30861385.VIS	406 bp				
CS265	TIRNBSLRR12-SalIHF	30861385.VIS					
CS266	TIRNBSLRR13-SpeI	31129385.VIS	372 bp				
CS267	TIRNBSLRR13-SalIHF	31129385.VIS					
CS268	TIRNBSLRR14-SpeI	31299385.VIS	442 bp				
CS269	TIRNBSLRR14-SalIHF	31299385.VIS					
CS270	TIRNBSLRR15-SpeI	31352385.VIS	422 bp				
CS271	TIRNBSLRR15-SalIHF	31352385.VIS					
CS272	TIRNBSLRR16-SpeI	32538385.VIS	435 bp				
CS273	TIRNBSLRR16-SalIHF	32538385.VIS					
CS274	TIRNBSLRR17-SpeI	36078385.VIS	403 bp				
CS275	TIRNBSLRR17-SalIHF	36078385.VIS					
CS276	TIRNBSLRR18-SpeI	FCL007E03.GB	445 bp				
CS277	TIRNBSLRR18-SalIHF	FCL007E03.GB					
CS278	TIRNBSLRR19-SpeI	RLN006A01.FASTA	385 bp				
CS279	TIRNBSLRR19-SalIHF	RLN006A01.FASTA					
CS284	CHICOWAKR1-Fw-SpeI	CHICO WAK RELATED CONTIG6745	494 bp	494 bp	58°C 45" 35X	ok	
CS285	CHICOWAKR1-Rev-SalIHF	CHICO WAK RELATED CONTIG6745					
CS286	CHICOWAKR2-Fw-SpeI	CHICO WAK RELATED CIC.EH691923	507 bp	507 bp	58°C 45" 35X	ok	
CS287	CHICOWAKR2-Rev-SalIHF	CHICO WAK RELATED CIC.EH691923					
CS288	CHICOWAKR3-Fw-SpeI	CHICO WAK RELATED CONTIG4966	461 bp		58°C 45" 35X	ok	
CS289	CHICOWAKR3-Rev-SalIHF	CHICO WAK RELATED CONTIG4966					
CS304	CHICOWAKR4-Fw-SpeI	CHICO WAK RELATED CONTIG 5881	543 bp	450 bp		ok	
CS305	CHICOWAKR4-Rev-SalIHF	CHICO WAK RELATED CONTIG 5881					
WAIT	CHICOWAKR5	CHICO WAK RELATED CI.EH691833					
CS290	NPR1-Fw-SpeI	Cich_inty.EH689598 NPR1	401 bp	410 bp		ok	
CS291	NPR1-Rev-SalIHF	Cich_inty.EH689598 NPR1					
WAIT	JAR1	CONTIG7082					
WAIT	MAPK4	CONTIG5419					
WAIT	COI1	CONTIG3063					
WAIT	JIN1	CONTIG9038					
CS292	CHICOSUN1-Fw-SpeI	EH690542, disease resistance protein	517 bp				
CS293	CHICOSUN1-Rev-SalIHF	EH690542, disease resistance protein					
CS294	CHICOSUN2-Fw-SpeI	CIC.EH693729, NBS-LRR	619 bp				
CS295	CHICOSUN2-Rev-SalIHF	CIC.EH693729, NBS-LRR					
CS296	CHICOSUN3-Fw-SpeI	CIC.EH678163, NBS-LRR	659 bp				
CS297	CHICOSUN3-Rev-SalIHF	CIC.EH678163, NBS-LRR					
CS298	CHICOSUN4-Fw-SpeI	CONTIG4952, NBS-LRR	609 bp				
CS299	CHICOSUN4-Rev-SalIHF	CONTIG4952, NBS-LRR					
CS300	CHICOSUN5-Fw-SpeI	CONTIG6761, disease resistance protein	592 bp				
CS301	CHICOSUN5-Rev-SalIHF	CONTIG6761, disease resistance protein					
CS302	CHICOSUN6-Fw-SpeI	CIC.EH672777, NBS-LRR	502 bp				
CS303	CHICOSUN6-Rev-SalIHF	CIC.EH672777, NBS-LRR					
CS306	CHICOSCAN1-Fw-SpeI	>Cich_inty.EH709933 WRKY	520 bp				
CS307	CHICOSCAN1-Rev-SalIHF	>Cich_inty.EH709933 WRKY					
CS308	CHICOSCAN2-Fw-SpeI	>Cich_inty.EH709264 disease resistance	589 bp				
CS309	CHICOSCAN2-Rev-SalIHF	>Cich_inty.EH709264 disease resistance					
CS310	CHICOSCAN3-Fw-SpeI	>Cich_inty.EH709073 Mildew resistance	604 bp				
CS311	CHICOSCAN3-Rev-SalIHF	>Cich_inty.EH709073 Mildew resistance					
CS312	CHICOSCAN4-Fw-SpeI	>Cich_inty.EH708104 Mildew resistance	504 bp				
CS313	CHICOSCAN4-Rev-SalIHF	>Cich_inty.EH708104 Mildew resistance					
CS314	CHICOSCAN5-Fw-SpeI	>Cich_inty.EH706551 avr-9	529 bp				
CS315	CHICOSCAN5-Rev-SalIHF	>Cich_inty.EH706551 avr-9					
CS316	CHICOSCAN6-Fw-SpeI	>Cich_inty.EH703640 HR	433 bp				
CS317	CHICOSCAN6-Rev-SalIHF	>Cich_inty.EH703640 HR					
CS318	CHICOSCAN7-Fw-SpeI	>Cich_inty.EH703550 PR1	635 bp				
CS319	CHICOSCAN7-Rev-SalIHF	>Cich_inty.EH703550 PR1					
CS320	CHICOSCAN8-Fw-SpeI	>Cich_inty.EH701710 TRANSCRIPTION FACTOR AP2 AVR-9	522 bp				
CS321	CHICOSCAN8-Rev-SalIHF	>Cich_inty.EH701710 TRANSCRIPTION FACTOR AP2 AVR-9					
CS322	CHICOSCAN9-Fw-SpeI	>Cich_inty.EH701125 Mildew resistance	500 bp				
CS323	CHICOSCAN9-Rev-SalIHF	>Cich_inty.EH701125 Mildew resistance					
CS324	CHICOSCAN10-Fw-SpeI	>Cich_inty.EH700709 disease resistance	520 bp				
CS325	CHICOSCAN10-Rev-SalIHF	>Cich_inty.EH700709 disease resistance					
CS326	CHICOSCAN11-Fw-SpeI	>Cich_inty.EH700393 ET response	525 bp				
CS327	CHICOSCAN11-Rev-SalIHF	>Cich_inty.EH700393 ET response					
CS328	CHICOSCAN12-Fw-SpeI	>Cich_inty.EH699232 disease resistance	457 bp				
CS329	CHICOSCAN12-Rev-SalIHF	>Cich_inty.EH699232 disease resistance					
CS330	CHICOSCAN13-Fw-SpeI	>Cich_inty.EH695658 NBS LRR	401 bp				
CS331	CHICOSCAN13-Rev-SalIHF	>Cich_inty.EH695658 NBS LRR					
CS332	CHICOSCAN14-Fw-SpeI	>Cich_inty.EH695155 ET response	511 bp				
CS333	CHICOSCAN14-Rev-SalIHF	>Cich_inty.EH695155 ET response					
CS334	CHICOSCAN15-Fw-SpeI	>Cich_inty.EH693800 WRKY	584 bp				
CS335	CHICOSCAN15-Rev-SalIHF	>Cich_inty.EH693800 WRKY					
CS336	CHICOSCAN16-Fw-SpeI	>Cich_inty.EH693796 Mildew resistance	579 bp				
CS337	CHICOSCAN16-Rev-SalIHF	>Cich_inty.EH693796 Mildew resistance					
CS338	CHICOSCAN17-Fw-SpeI	>Cich_inty.EH692411 ET respnse	698 bp				
CS339	CHICOSCAN17-Rev-SalIHF	>Cich_inty.EH692411 ET respnse					
CS340	CHICOSCAN18-Fw-SpeI	>Cich_inty.EH691945 WOUND-RESPONSE, AP2	502 bp				
CS341	CHICOSCAN18-Rev-SalIHF	>Cich_inty.EH691945 WOUND-RESPONSE, AP2					



Figure 1.1 : Illustration de 3 cultivars actuels de chicorée industrielle. De gauche à droite : Chicorée variété améliorée verte, variété Catalogna et variété Pan di zucchero. Photos Groupement National Interprofessionnel des Semences (<http://www.gnis.fr/>)



Figure 1.2 : Illustration d'une fleur bleue de *C. intybus*. Photo Compositae Genome Project Data Base (<http://cgpdb.ucdavis.edu/>).

Introduction

1 La chicorée industrielle

1.1 Présentation générale

La chicorée industrielle, *Cichorium intybus*, fait partie de la famille des *Compositae/Asteraceae*. Cette famille regroupe entre autres la laitue, le tournesol, le pissenlit, l'artichaut, l'endive, ... Les fleurs sont bleues mais peuvent, dans de rares cas, être blanches ou rosées (Stoddart 2003). Les figures 1.1 et 1.2 illustrent la morphologie de plants de chicorée en champs ainsi que les fleurs bleues. *C. intybus* est présente principalement en Europe mais est également retrouvée aux Etats Unis, en Australie et en Afrique du Sud. (Sulas 2004).

C. intybus est une plante pérenne diploïde ($2n=18$). La chicorée industrielle est capable de s'autoféconder. A l'origine, elle était cultivée pour sa racine qui était utilisée pour fabriquer un ersatz du café. Aujourd'hui, la production de chicorée est principalement destinée à la production d'inuline (un polymère de β -1,2 fructose), une réserve de carbohydrates chez de nombreux *Asteraceae* (Kiers, Mes et al. 2000). L'inuline est une molécule à haute valeur ajoutée, présentant de nombreuses applications et propriétés curatives. L'hydrolyse de l'inuline permet d'obtenir un sirop de fructose, un exhausteur de goût faible en énergie. Des recherches médicales ont mis en évidence que la consommation régulière d'inuline permet une diminution du risque de développer un cancer du colon et entraîne une diminution du taux de cholestérol dans le sang (Cornelis 2003). L'inuline est également connue pour ses propriétés anti-oxydantes (Rani 2003).

La chicorée industrielle est une plante biannuelle. Elle nécessite une phase de vernalisation pour fleurir. Pour germer, les graines ont besoin d'un sol ayant une température supérieure à 10°C et une bonne humidité (Sorensen, Kiaer et al. 2007). La première année de culture, la chicorée apparait durant les deux premiers mois comme une large racine produisant un grand nombre de petites racines latérales, et présente une rosette de 6 à 8 feuilles. Jusqu'au 5^{ème} mois, la racine accumule de l'inuline dont la concentration peut atteindre 80% du poids sec (Druart, De Roover et al. 2001). La production d'inuline était auparavant affectée par le phénomène de montaison qui pouvait toucher jusqu'à 22,7% des plants en champs. Pour monter à fleur, la plante avait recours aux nutriments stockés dans la racine et la teneur en inuline de la racine baissait. Aujourd'hui, la sélection de nouvelles variétés a permis de réduire ce problème (Sorensen, Kiaer et al. 2007).

La chicorée est cultivée pour ses feuilles sous forme d'endive, aussi appelée chicon ou chicorée witloof. Le chicon est obtenu après une induction florale et un développement floral accéléré dans le noir (Quenon 2000).

1.2 Maladies foliaires de la chicorée industrielle

Les producteurs de chicorée industrielle sont confrontés à deux principales maladies affectant le rendement des récoltes : l'oïdium causé par *Erysiphe cichoracearum* et la rouille causée par *Puccinia cichorii*. Certains plants peuvent être également touchés par la bactérie *Pseudomonas syringae* et le champignon *Alternaria* (Cornelis 2003).

Le champignon *Erysiphe cichoracearum* est le principal agent pathogène de la chicorée industrielle. Il peut lourdement toucher la production en affectant l'activité photosynthétique de la plante. L'infection de la plante par le pathogène est caractérisée premièrement par l'apparition de taches blanches sur les feuilles qui vont croître sur l'ensemble de la chicorée jusqu'à la recouvrir d'un duvet blanchâtre. La présence et le développement de ce champignon sont favorisés par une humidité importante. Les premiers symptômes apparaissent dans la seconde partie du mois d'août. La proportion de plants touchés et l'impact sur les cultures varient chaque année. (Trdan, Valic et al. 2004). Illustration voir figure 1.3.

La rouille visible sur les feuilles de la chicorée est causée par le champignon basidiomycète *Puccinia cichorii*. Cette maladie foliaire est très peu décrite et peu de littérature scientifique est disponible à son sujet. L'infection se caractérise par l'apparition de spores à la surface de la feuille qui se présentent sous forme de taches brunâtres. L'augmentation de la surface de ces taches entraîne un flétrissement prématuré de la feuille. Le développement du champignon est favorisé par un temps chaud et humide et se développe parallèlement avec le champignon *Erysiphe cichoracearum*, d'où la difficulté de définir correctement son impact propre sur la production (Hermann 2006). Illustration voir figure 1.4.

2 Interactions plantes- pathogènes

Il n'existe pas d'étude approfondie sur la résistance de la chicorée industrielle à ces pathogènes. Vingt-et-une séquences de chicorée associées à des gènes de résistance sont disponibles dans la banque de données de NCBI et constituent les uniques informations dont nous disposons actuellement sur le « système immunitaire » de la chicorée (Plocik, Layden et al. 2004). Cependant, de nombreuses données sont disponibles chez des plantes modèles comme *A. thaliana*. Nous supposons que *C. intybus* présente, au moins en partie, des homologies avec le système de défense d'*Arabidopsis thaliana*.

Les végétaux peuvent être attaqués par une grande variété de pathogènes tels que des virus, des bactéries, des champignons, des oomycètes, des insectes et des nématodes. Contrairement aux animaux, les végétaux sont incapables de fuir leur agresseur et ne disposent pas de cellules spécialisées dans la défense face aux pathogènes. Chaque cellule végétale doit dès lors être capable de reconnaître et de répondre à une attaque pathogène (McDowell and Simon 2008).

Pour se prémunir des pathogènes, la cellule végétale est protégée par une paroi cellulaire qui lui assure une protection physique et chimique ainsi que par son système immunitaire qui se divise en deux parties. D'une part, une résistance générale dite *pathogen associated molecular patterns triggered immunity* (PTI); d'autre part, une réponse immunitaire spécialisée nommée *effector-triggered immunity* (ETI).

Les PAMP sont des molécules conservées et vitales caractérisant les pathogènes. Ces molécules peuvent difficilement être modifiées et/ou supprimées sans affecter la survie du pathogène. Le lipopolysaccharide est une molécule intrinsèque des bactéries gram négatives. On y retrouve une partie extrêmement conservée, le lipide A, qui est perçu par la cellule végétale comme un PAMP (Schwessinger and Zipfel 2008). Le flagelle bactérien est également une structure indispensable à un grand nombre de bactéries. Le domaine conservé

et peu variable flg22 est reconnu par *A. thaliana* comme un PAMP (Zipfel and Felix 2005). La terminologie PAMP est actuellement remplacée progressivement par d'autres dénominations comme MAMP (*microbe associated molecular patterns*) (Zipfel 2008), ou MIMPS (*microbe induced molecular patterns*) car les réponses de défense peuvent être déclenchées par des molécules induites par la présence du pathogène, pouvant ne pas appartenir au pathogène lui même (Mackey and McFall 2006). Certains pathogènes peuvent, par exemple, sécréter des polygalacturonases ou des pectates lyases capables de dégrader la pectine dans la paroi cellulaire végétale. Des récepteurs sont ensuite capables de reconnaître des fragments pectiques et les perçoivent comme un signal d'activation du système de défense (Humphrey, Bonetta et al. 2007).

Des récepteurs transmembranaires de type serine/thréonine kinase peuvent reconnaître les PAMP par leur domaine extracellulaire composé de structures riches en leucine (domaine *leucin rich repeat*, LRR). Ces récepteurs like kinase (RLK) sont capables, via l'activation de la voie MAPK (*mitogen activated protein kinase*), d'activer des défenses de la plante (Bittel and Robatzek 2007) : la paroi est rigidifiée par la synthèse de callose, il y a sécrétion de phytoalexines, une série de protéines de défense sont sécrétées comme des chitinases, protéases et glucanases qui peuvent s'attaquer directement aux parois des champignons/bactéries (de Wit 2007). La paroi est également renforcée par de la lignine - un polymère de phénylpropanoïde - et des protéines antimicrobiennes, comme les défensines, peuvent être sécrétées. La réponse la plus importante est la réponse hypersensible qui correspond à une mort cellulaire dans la région où le pathogène a été détecté (Nurnberger, Brunner et al. 2004).

La réponse générée par la détection de PAMP est généralement suffisante pour la neutralisation du pathogène. Les pathogènes ont toutefois évolué et des mécanismes ont été sélectionnés pour échapper au système immunitaire de la cellule végétale. Bactéries, champignons et oomycètes sont capables d'injecter dans le cytoplasme de la cellule végétale des effecteurs qui vont inhiber les réponses de défense de la plante. Les bactéries gram-négatives ont recours au système de sécrétion de type 3 (TTSS) pour injecter leurs effecteurs dans le cytoplasme de la cellule végétale. Les champignons, ainsi que les oomycètes, injectent des effecteurs dans la cellule végétale à l'aide d'une structure appelée haustorium, illustrée par la figure 1.5 (McDowell and Simon 2008). Les effecteurs semblent agir comme des « saboteurs » des réponses de défense de la plante (Dangl and McDowell 2006).

La cellule végétale a, à son tour, sélectionné des mécanismes permettant de détecter spécifiquement la présence d'effecteurs et de déclencher des mécanismes de défense dont la réponse hypersensible correspond à une mort cellulaire programmée. La détection d'effecteurs et la réponse spécifique à ces derniers est communément appelée : système de défense ETI « *effector triggered immunity* » (Chisholm, Coaker et al. 2006). Les réponses de défense PTI et ETI sont illustrées par la figure 1.6.

Une étape clef de la réponse hôte pathogène est, comme montré ci-dessus, la reconnaissance entre une protéine de défense et le pathogène ou un effecteur de celui-ci. En 1971, Flor proposa un modèle végétal de résistance aux pathogènes, le modèle *gene for gene*. Le pathogène contient des gènes d'avirulence (*Avr*) qui causent la maladie chez la plante. La plante contient un gène de résistance (*R gene*) qui reconnaît directement le gène d'avirulence du pathogène et confère ainsi une résistance au pathogène. La pression de sélection favorise

l'évolution du gène d'avirulence qui peut échapper à la reconnaissance des protéines de résistance de la plante. Les protéines de résistance évoluent à leur tour pour reconnaître les nouveaux gènes d'avirulence (Flor 1971).

Les gènes de résistance sont des récepteurs aux pathogènes capables, suite à la reconnaissance de ces derniers, de déclencher les voies de défense. En réalité, l'interaction directe entre un récepteur et le pathogène semble être un cas de figure marginal. Les protéines de résistance reconnaîtraient indirectement le pathogène via des molécules adaptatrices détectant des perturbations cellulaires causées par le pathogène et ses effecteurs (Jones and Dangl 2006). Le modèle actuel est appelé hypothèse de garde, « *guard hypothesis* ». Des protéines jouent un rôle de gardien et sont sensibles aux modifications induites par la présence d'un pathogène. La protéine de garde peut être une protéine qui est directement modifiée par le pathogène. Le gardien peut ensuite interagir avec une protéine de résistance qui déclenchera les réponses appropriées au pathogène détecté (Illustration voir figure 1.7) (McDowell and Woffenden 2003).

2.1 Récepteurs et gènes de résistance

Un grand nombre de récepteurs sont impliqués dans la détection de pathogènes. Les principales familles de récepteurs impliquées dans la détection et l'initiation du signal de réponse au pathogène seront décrites ci-dessous.

2.1.1 Les récepteurs à protéines kinases

Chez les animaux, les récepteurs à protéines kinases assurent le transfert d'information du milieu extracellulaire à l'intérieur de la cellule (Shiu and Bleecker 2001). Les récepteurs protéines kinases sont divisés en trois classes principales : les récepteurs à tyrosine kinase qui phosphorylent des résidus tyrosines, les récepteurs à sérine/thréonine kinase qui phosphorylent les résidus sérines et thréonines et les récepteurs à histidine kinases qui phosphorylent les résidus histidines. (Afzal, Wood et al. 2008). En général, les protéines kinases sont présentes en membrane et sont inactives à l'état monomérique. La fixation d'un ligand induit la dimérisation du récepteur et la transduction d'un signal au niveau intracellulaire via des cascades de phosphorylation (Becraft 2002).

Les plantes possèdent dans leur génome un grand nombre récepteurs like kinase (RLK), récepteurs ressemblant aux protéines kinases du monde animal (Becraft 2002). La famille des RLK comprend environ 600 membres chez *A. thaliana* et 1000 membres chez *Oriza sativa*, le riz. L'augmentation et la diversification des RLK se réalisent par réplication en tandem ou lors de duplications complètes de génome (Afzal, Wood et al. 2008). Les RLK sont caractérisés par un domaine extracellulaire, probablement impliqué dans la reconnaissance d'un stimulus, un domaine transmembranaire et un domaine kinase intracellulaire (Montesano, Brader et al. 2003). Alors que la majorité des récepteurs protéines kinases animales sont des tyrosines kinases dans le monde animal, chez les plantes, les RLK sont des serines thréonines kinases (Shiu and Bleecker 2001). Les récepteurs like kinases sont principalement impliqués dans le développement et la croissance cellulaire ainsi que dans la détection et la réponse aux pathogènes (Afzal, Wood et al. 2008). On observe une grande variabilité au niveau du domaine extracellulaire. Celui-ci permet de classer les différents RLK (Montesano, Brader et al. 2003). On dénombre ainsi plus de vingt-et-une classes de RLK chez *A. thaliana*. Le motif extracellulaire le plus fréquent est celui constitué de répétitions de leucine (*leucin rich repeat*, LRR). Ce motif est présent dans plus de la moitié des RLK.

(Becraft 2002). Il existe toutefois aussi quelques récepteurs like kinases cytoplasmiques comme le récepteur Pto qui confère la résistance à *Pseudomonas syringae* (Montesano, Brader et al. 2003). La figure 1.8 illustre la diversité des RLK chez les plantes.

2.1.2 Les récepteurs WAK : Wall Associated Kinase

Les WAK, *wall associated kinases*, appartiennent à la famille des récepteurs like kinases. Tous les WAK contiennent un domaine cytoplasmique serine-threonine kinase, un domaine transmembranaire et un domaine extracellulaire. Ce dernier contient deux domaines similaires à l'EGF (*epidermal growth factor*) (Becraft 2002). Positionnés dans la membrane plasmique, il semblerait que les récepteurs WAK interagissent avec la paroi cellulaire végétale. Des études suggèrent que WAK1 est capable de se fixer de manière covalente à la pectine. On dénombre 5 WAK chez *A. thaliana* ainsi qu'une vingtaine de gènes WAK-like (WAKL) découverts sur base d'homologies de séquences (Humphrey, Bonetta et al. 2007). Certains WAKL ne codent que pour le domaine extracellulaire (Decreux, Thomas et al. 2006). Les 5 WAK sont localisés dans un même cluster de 30kb dans le génome d'*A. thaliana*. On retrouve des WAK dans tous les types cellulaires d'*A.thaliana* et dans d'autres plantes vasculaires (He, Cheeseman et al. 1999). L'expression des WAK est induite chez *A.thaliana* quand la plante est attaquée par un pathogène ou lorsque celle-ci est stimulée par de l'acide salicylique ou un analogue comme l'acide 2,6-dichloroisonicotinique (INA).

WAK1 est le récepteur WAK le plus étudié à l'heure actuelle. Ce récepteur de 78.8kDa pour 713 acides aminés n'est présent que dans les tissus verts d'*A. thaliana*. WAK1 peut lier la pectine. WAK1 est capable de lier des oligogalacturonides ayant un degré de polymérisation supérieur ou égal à 9. Les oligosaccharides sont connus pour agir comme éliciteurs des réponses de défense de la plante (Decreux, Thomas et al. 2006). L'expression de WAK1 est positivement régulée par NPR1 (gène clef de la voie de défense dépendante de l'acide salicylique). L'utilisation d'ARN anti sens montre également que l'expression de WAK1 est nécessaire pour survivre à une stimulation à l'INA. Le déclenchement des réponses aux pathogènes et les changements physiologiques induits par celles-ci obligent la plante à surexprimer des protéines telles que WAK1 pour survivre (Anderson, Wagner et al. 2001). WAK1 est requis chez *A. thaliana* pour résister à l'infection par *Pseudomonas syringae*. L'expression de WAK1 est également surexprimée durant la réponse systémique acquise et est induite par le champignon pathogène *Alternaria brassicicola* ainsi que par l'éthylène et le jasmonate.

Les WAK sont impliqués dans d'autres fonctions que la pathogenèse, dont la croissance cellulaire (Humphrey, Bonetta et al. 2007).

2.1.3 Les gènes de résistance *nucleotide-binding site leucine rich repeat* (NBS-LRR)

La plus grande famille de gènes de résistance rassemble les gènes NBS-LRR : *nucleotide binding site-leucin rich repeat* (Kohler, Rinaldi et al. 2008). Ces gènes sont impliqués dans la détection et la réponse à une grande diversité de pathogènes chez les végétaux : champignons, bactéries, virus, nématodes et insectes (Ayliffe and Lagudah 2004). Le nombre de gènes NBS-LRR varie fortement en fonction des espèces végétales. On dénombre actuellement 150 gènes NBS-LRR fonctionnels chez *A. thaliana*, alors que 400 gènes de ce type ont été trouvés dans le génome de *Populus trichocarpa* et près de 600 gènes NBS-LRR ont pu être mis en évidence dans le génome du riz (Radwan, Gandhi et al. 2008). Aujourd'hui, vingt-et-une

séquences de chicorée ont été associées à cette famille de gènes et sont accessibles via la banque de données NCBI. La grande majorité des gènes de résistance NBS-LRR chez la chicorée reste donc à découvrir.

Lorsqu'on a mis en évidence les premières protéines NBS-LRR, on a tout d'abord pensé que ces protéines de résistance étaient principalement localisées dans le cytoplasme de la cellule végétale. Les dernières recherches ont permis de mettre en évidence que les protéines NBS-LRR sont principalement nucléaires. Via un pore nucléaire, des récepteurs peuvent être exportés vers le cytoplasme. Une fois activées suite à la reconnaissance d'un effecteur, des séquences NLS (*nuclear localisation sequence*) permettent une importation de la protéine activée vers le noyau. Les réponses de défense peuvent être alors activées par exemple en inhibant des inhibiteurs des voies de défense de la plante (Caplan, Padmanabhan et al. 2008). Illustration figure 1.9.

Les gènes NBS-LRR sont caractérisés d'une part par le domaine de fixation à un nucléotide, très conservé chez les plantes, d'autre part par un domaine enrichi en leucine (*leucin rich repeat*). Figure 1.10.

Le *nucleotide binding site* (NBS) est capable de fixer une molécule d'ATP ou de GTP (Ayliffe and Lagudah 2004). Ce domaine est repris sous l'appellation de NB-ARC. Celui-ci est partagé par les gènes de résistance NBS-LRR chez les plantes ainsi que par des régulateurs de mort cellulaire chez les animaux. Le domaine NB-ARC apparaît conservé entre APAF-1 (*apoptotic protease-activating factor-1*), les gènes de résistance NBS-LRR chez les plantes ainsi que dans CED-4 (*Caenorhabditis elegans death-4 protein*) (van der Biezen and Jones 1998). Le domaine NB-ARC est lié évolutivement aux protéines NACHT-LRR (NAIP, CIITA, HET-E, TP1) qui sont associées à l'immunité innée des mammifères (Takken, Albrecht et al. 2006). Figure 1.10.

Aucune cristallisation de protéines NBS-LRR n'ayant encore été réalisée, les données structurales actuelles sont donc extrapolées des résultats de la cristallisation de CED-4 et de Apaf-1 (Figure 1.10). Le domaine NB-ARC est défini par 3 sous-domaines chez les plantes: le domaine de fixation du nucléotide, suivi de deux domaines ARC1 et ARC2 (Rairdan and Moffett 2006).

La cristallisation de CED-4 et de Apaf-1 a permis de montrer que le domaine de fixation du nucléotide est une structure en forme de poche où l'ATP ou l'ADP peuvent se localiser. Dans le modèle Apaf-1, c'est la forme ADP qui semble présenter une stabilité thermodynamique optimale. L'inverse est observé chez CED-4. Des mutations dans le site de fixation au nucléotide peuvent altérer soit la fixation de l'ATP, soit l'hydrolyse de l'ATP. L'étape de fixation de l'ATP par le site de liaison entraîne l'activation de la protéine NBS-LRR (DeYoung and Innes 2006). Le domaine de liaison au nucléotide contient une structure p-loop formant le lien avec le troisième phosphate de l'ATP ainsi qu'avec l'ion Mg^{2+} . (Liu, Liu et al. 2007).

Le rôle des domaines ARC a pu être mis en évidence par des expériences de mutagenèse dirigée. Les mutations touchant le domaine ARC1 débouchent sur un phénotype d'auto-activation. Au contraire, les mutations du domaine ARC2 ont tendance à inactiver la protéine. Les mutations fonctionnelles du domaine ARC2 se localisent dans une zone conservée MHD

(méthionine-histidine-aspartate). Le domaine ARC1 permet une interaction et une stabilisation entre le domaine NB-ARC et le domaine LRR. ARC2 interagit également avec le domaine LRR mais constitue la clé de l'activation du gène après la reconnaissance du pathogène par le domaine LRR (van Ooijen, Mayr et al. 2008). (Voir figure 1.11)

Le domaine *leucine rich repeat* (LRR) constitue la partie carboxy terminale du gène NBS-LRR. Ce motif peut être associé à d'autres domaines que le domaine de fixation au nucléotide. Le domaine *leucine rich repeat* constitue ainsi le domaine extracellulaire d'un grand nombre de récepteurs protéines kinases. Il s'agit d'un domaine d'interaction protéine-protéine. La cristallisation de ce domaine révèle la formation d'un grand nombre de feuillettes bêta parallèles suivis d'une hélice alpha (DeYoung and Innes 2006). La région contenant les feuillettes beta est impliquée dans la reconnaissance directe ou indirecte du pathogène. Des variations dans la structure du domaine LRR permettent d'augmenter le nombre de pathogènes reconnus (Rairdan and Moffett 2006). Une simple substitution dans ce domaine peut ainsi conférer une résistance ou une sensibilité à un pathogène donné (Liu, Liu et al. 2007). Contrairement au domaine NBS, le domaine LRR montre une très grande variabilité au sein même de la famille des gènes NBS-LRR. La sélection favorise la diversification de cette région pour accroître le spectre de reconnaissance de pathogènes (Radwan, Gandhi et al. 2008).

La famille des gènes NBS-LRR se subdivise en deux grands groupes caractérisés par la partie N terminale du gène. D'une part, les gènes présentant un domaine homologue aux protéines Toll chez la drosophile et aux récepteurs à l'interleukine 1 chez les mammifères constituent la famille des TIR-NBS-LRR. D'autre part, les gènes NBS-LRR ne présentant pas de domaine TIR sont classés dans le groupe des X-NBS-LRR où le domaine X est souvent un domaine coiled-coil (CC) (Ayliffe and Lagudah 2004). Une plus grande diversité dans la structure et les séquences des gènes X-NBS-LRR laisse supposer que cette famille de gènes de résistance est plus ancienne que celle des TIR-NBS-LRR (Chini and Loake 2005).

Aucune fonction fondamentalement différente entre les gènes TIR-NBS-LRR et X-NBS-LRR n'a pu être mise en évidence. Les gènes TIR-NBS-LRR semblent être liés à la voie de réponse EDS1 (*enhanced disease susceptibility*) par opposition aux X-NBS-LRR qui seraient associés à la voie NDR1 (*nonrace-specific disease resistance*) par des mécanismes encore inconnus aujourd'hui. Les voies EDS1 et NDR1 aboutissent chacune à la même réponse de défense (Glazebrook 2001). Aucun gène CC-NBS-LRR n'a été observé chez les monocotylédones. Chez les dicotylédones, les TIR-NBS-LRR semblent être prédominants par rapport aux X-NBS-LRR (Meyers, Morgante et al. 2002).

Chez la drosophile le domaine *Toll-like receptor* (TLR) permet l'activation des défenses dans la lutte contre les champignons. Chez les mammifères, dix membres de la famille *Toll-like receptor* ont pu être identifiés comme jouant un rôle dans la détection d'effecteurs de réponse de défense. Ce domaine Toll/Interleukin- receptor a une taille d'environ 200 acides aminés et est caractérisé par trois motifs très conservés. On a pu mettre en évidence que des variations dans le domaine TIR étaient liées à la spécificité de la reconnaissance des pathogènes (Ellis, Dodds et al. 2007).

Les gènes de résistance NBS-LRR apparaissent localisés en cluster dans les génomes du riz et d'*A. thaliana* (Figure 1.12). Ces clusters constitueraient des groupes de résistance. Un cluster

contiendrait un ensemble de gènes NBS-LRR qui évolueraient pour reconnaître un pathogène donné. De nouveaux clusters se formeraient dans le génome à l'aide de crossing over et de conversion génique donnant naissance à de nouveaux îlots de résistance (Meyers, Kozik et al. 2003). Les clusters évoluent en se spécialisant vers un motif donné, tandis que de nouveaux clusters apparaissent dans le génome et sont éventuellement sélectionnés s'ils élargissent la gamme de reconnaissance de la plante. Les gènes NBS-LRR sont soumis à une forte sélection positive qui favorise l'augmentation de la diversité chez ces derniers (Mondragon-Palomino, Meyers et al. 2002). Des résultats suggèrent aujourd'hui que la fréquence des crossing over et des remaniements de la structure des gènes NBS-LRR pourrait être augmentée suite à un stress causé par un pathogène. Il existe un mécanisme de diversification qui serait activé lors des stress et qui repose principalement sur la recombinaison homologue (McDowell and Simon 2008). La partie *leucin rich repeat* est particulièrement susceptible à cette génération de diversité, par exemple avec une variation du nombre de domaines LRR ou une variation de la structure de ces derniers (Mondragon-Palomino, Meyers et al. 2002).

Les récepteurs NBS-LRR peuvent reconnaître un pathogène suivant trois modèles. Premièrement, l'interaction entre l'effecteur pathogène et le récepteur NBS-LRR peut être directe et se produit généralement au niveau du domaine LRR. Deuxièmement, la reconnaissance peut également se faire de manière indirecte. La molécule élicitrice peut être reconnue par un adaptateur de la cellule végétale. Une fois l'éliciteur reconnu par la molécule adaptatrice, celle-ci peut se lier au récepteur NBS-LRR au niveau de la région N terminale ainsi qu'au niveau LRR et activer ce dernier. Troisièmement, un adaptateur peut également venir se fixer sur le gène NBS-LRR au niveau du domaine N terminal et permet ainsi la reconnaissance d'un éliciteur par le nouveau complexe formé (Caplan, Padmanabhan et al. 2008). Voir figure 1.13.

A l'état inactif, les gènes NBS-LRR sont liés à une molécule d'ADP. Une fois sa cible reconnue, un changement de conformation permet la libération de l'ADP du site de fixation au nucléotide et le remplacement de ce dernier par une molécule d'ATP. Ce dernier induit un deuxième changement de conformation notamment au niveau de la partie N-terminale. L'activité ATPase de la protéine permet le retour à l'état inactif du récepteur (Takken, Albrecht et al. 2006). Illustration à la figure 1.11.

2.2 Réponse hypersensible et résistance systémique acquise

La détection d'un pathogène par la plante se traduit d'une part par un renforcement des tissus au niveau de l'infection, la synthèse locale de métabolites anti-microbiens, l'expression de gènes de résistance, et d'autre part par une mort cellulaire appelée réponse hypersensible (McDowell and Dangl 2000). La réponse hypersensible, HR, empêche le développement du pathogène (Glazebrook 2001).

Après avoir reconnu le pathogène, le premier événement détectable est un flux d'ions à travers la membrane et la production d'intermédiaires d'oxygène réactif (ROI) comme le superoxyde O_2^- et le peroxyde d'hydrogène H_2O_2 . Cette première étape, appelée aussi burst oxydatif, est nécessaire pour l'activation des autres réponses de défense (McDowell and Dangl 2000).

Parallèlement, une réponse systémique chez la plante est activée, la *systemic acquired resistance* (Castiglioni, Rizzi et al.). La plante développant une SAR montre une résistance

accrue aux pathogènes. Cette réponse de défense comprend l'accumulation d'acide salicylique et l'expression de gènes de résistance (Glazebrook 2001).

2.3 Voies de réponses aux pathogènes

Trois voies de réponses principales sont activées par les gènes de résistance suite à la détection d'un pathogène. Nous détaillerons ci-dessous les voies de signalisation de l'acide salicylique (SA), de l'éthylène (ET) et du jasmonate (JA). Nous montrerons enfin comment ces trois voies de signalisation interagissent les unes avec les autres.

2.3.1 La voie de l'acide salicylique (SA)

L'acide salicylique est un métabolite secondaire naturel des plantes connus pour ses propriétés analgésiques. C'est avec l'acide salicylique que l'on synthétise la molécule de l'aspirine, l'acide acétylsalicylique (Hopkins 2003).

Le rôle essentiel de l'acide salicylique dans la réponse aux pathogènes chez les plantes a pu être mis en évidence grâce à l'expression de plantes transgéniques surexprimant le gène bactérien NahG, salicylate hydroxylase, qui convertit l'acide salicylique en catéchol. L'acide salicylique ne peut donc être accumulé. Ces plantes transgéniques montrent une susceptibilité accrue à tous types de pathogènes. D'autres mutants impliqués dans la voie de synthèse de l'acide salicylique présentent également le phénotype de sensibilité aux pathogènes (Pieterse 2001).

La première voie de synthèse de l'acide salicylique à avoir été décrite partait de la phénylalanine. Cependant, cette voie apparut marginale lorsque l'on s'aperçut qu'elle ne pouvait fournir une quantité suffisante d'acide salicylique à la plante et qu'il devait exister des voies parallèles (Shah 2003). La seconde voie implique l'enzyme isochorismate synthase et l'isochorismate : pyruvate lyase qui permettent la synthèse d'acide salicylique à partir de chorismate, ce dernier provenant de la voie shikimate (Shah 2003). Cette seconde voie est principalement utilisée lors de la réponse systémique acquise (Vasyukova and Ozeretskovskaya 2007). La mutation du gène SID2, codant pour l'isochorismate synthase, entraîne une chute de la concentration en acide salicylique confirmant que la voie de synthèse principale de ce dernier est basée sur l'isochorismate plutôt que sur la voie des phénylalanines (Glazebrook 2005). L'acide salicylique est stocké sous une forme glycosylée dans les vacuoles. L'acide salicylique 2-O-beta-D-glucoside est ensuite hydrolysé lors de l'activation des défenses de la plante. L'acide salicylique peut aussi être modifié en méthyle salicylate, forme volatile qui joue un rôle dans la transmission du signal de défense à travers la plante (Loake and Grant 2007).

Les gènes ci-dessous, impliqués dans la voie de défense dépendante de l'acide salicylique, sont illustrés en figure 1.14.

Deux gènes, PAD4 (*phytoalexin deficient 4*) et EDS1 (*enhanced disease susceptibility 1*), sont requis pour l'activation de l'accumulation d'acide salicylique. Ces deux gènes codent pour des protéines interagissant entre elles. Malgré leurs homologues respectives avec des triacyl-glycerol lipases, cette activité n'est pas requise pour le fonctionnement de ces protéines (Glazebrook 2005). EDS1 est activé par les gènes de défense TIR-NBS-LRR. Quelques gènes CC-NBS-LRR peuvent également interagir avec EDS1. En plus d'être activé par PAD4, EDS1 est capable d'interagir avec SAG101, *senescence associated gene 101*. Ces

trois gènes sont requis pour la réponse hypersensible (Liu and Coaker 2008). L'acide salicylique est capable d'induire l'expression de EDS1 et de PAD4 par une boucle rétro activatrice (Wiermer, Feys et al. 2005).

EDS5 est également nécessaire à la production d'acide salicylique. Ce gène code pour un transporteur MATE qui pourrait assurer le transport d'intermédiaires de la voie de synthèse de l'acide salicylique.

La protéine codée par le gène NPR1, *non expressor of pathogenesis- related proteins1*, agit en aval de l'acide salicylique. Lorsque le taux d'acide salicylique est normal, NPR1 se trouve à l'état oligomérique dans le cytoplasme de la cellule. L'augmentation du taux d'acide salicylique entraîne la dissociation du complexe NPR1 en monomère de NPR1 par la réduction de ponts disulfures. Grâce à son domaine NLS (Liu and Coaker 2008), NPR1 sous sa forme monomérique peut rentrer dans le noyau où il peut interagir avec les facteurs de transcription de type TGA qui sont, entre autres, impliqués dans l'activation des gènes de résistance PR1 (Glazebrook 2005). NPR1 est également nécessaire à l'activation de facteurs de transcription de type WRKY (Liu and Coaker 2008). L'existence de voies alternatives à NPR1 est suggérée par l'observation de réponses différentes bloquées par le mutant NPR1 ou par l'observation d'une plante incapable d'accumuler l'acide salicylique (Glazebrook 2001).

2.3.2 La voie de l'éthylène (ET) et du méthyl jasmonate (JA)

Des voies alternatives à celle de l'acide salicylique existent dans le cadre des réponses aux pathogènes chez les végétaux. La voie de l'éthylène et du méthyl jasmonate sont deux voies alternatives à celle de l'acide salicylique.

L'acide jasmonique (ou jasmonate), ainsi que son ester méthylé (ou méthyl jasmonate) présent sous forme gazeuse, sont synthétisés à partir d'un composé fréquent dans les membranes cellulaires, l'acide α -linoléique. Ces deux molécules sont impliquées entre autres dans la sénescence foliaire et les réponses de défense aux herbivores et insectes. (Hopkins 2003). L'éthylène est un hydrocarbure simple ($H_2C=CH_2$) synthétisé sur base de méthionine. Outre son rôle dans la réponse aux pathogènes, l'éthylène est également connu entre autres pour : son effet inhibiteur de l'accroissement des cellules ; son rôle dans la maturation des fruits ; il favorise l'abscission des feuilles (Raven 2007).

La voie du méthyl jasmonate est déclenchée par une blessure de la plante ou par les attaques d'insectes. Une partie de la voie de l'éthylène interagit avec celle du méthyl jasmonate dans le cadre des réponses de défense. Ces deux voies mènent à des réponses différentes de celles déclenchées par l'acide salicylique comme la réponse systémique acquise (Dong 1998).

Le rôle du méthyl jasmonate dans la réponse aux pathogènes a pu être mis en évidence par l'observation de phénotypes de mutants pour de la voie de synthèse du jasmonate ainsi que pour la perception à ce dernier. Les mutants montrent une sensibilité accrue aux pathogènes.

Le gène central de la voie du jasmonate est COI1. Dans toutes les espèces végétales où le gène a été étudié, COI1 est impliqué les réponses dépendantes du jasmonate. La protéine codée par COI1 (*coronatine insensitive 1*), est présente dans un complexe E-3 ubiquitine ligase. Ce complexe cible des protéines à dégrader par exemple RPD3b, une histone deacétylase. La dégradation de RPD3b par le protéasome via COI1 permet une activation des

voies du jasmonate. Le complexe SCF, *skip-cullin-F-box*, avec lequel COI1 interagit est régulé par des protéines venant d'autres voies de signalisation que le jasmonate. AXR1 et JAI4 régulent SCF étant respectivement liés à l'auxine et à l'éthylène.

Les voies du jasmonate et de l'éthylène peuvent coopérer ou être antagonistes pour agir de manière ciblée à un stress. Les mutants pour les voies de synthèse du jasmonate et de l'éthylène ainsi que pour la perception de ceux-ci montrent une sensibilité accrue aux pathogènes. Deux facteurs de transcription jouent un rôle clef dans la régulation entre les deux voies : ERF1 (*ethylene response factor 1*) et JIN1 (*jasmonate insensitive 1*).

L'expression d'ERF1 est induite par l'infection de pathogènes nécrotrophes et régule l'expression de gènes de défense comme PDF1.2 (*Plant Defensin 1.2*). L'expression constitutive d'ERF1 permet l'augmentation de la résistance à une grande diversité de champignons nécrotrophes.

JIN1 encode pour un facteur de transcription qui régule deux voies différentes du jasmonate. Il régule ainsi négativement l'expression de gènes de défense mais est capable d'activer l'expression de gènes de défense en réponse à une blessure mécanique ou biotique de la plante (Lorenzo and Solano 2005).

2.3.3 Cross talk entre SA, ET et JA

Les voies du jasmonate et de l'acide salicylique agissent de manière antagoniste l'une envers l'autre. De manière générale, l'application simultanée d'acide salicylique ou d'un homologue ainsi que de jasmonate montre une inhibition globale des réponses de défense chez la plante (Thaler 2002). Les différentes interactions entre les voies de l'acide salicylique, de l'éthylène et du jasmonate décrites ci-dessous sont illustrées par la figure 1.15. Les interactions principales entre des gènes clefs des trois voies sont détaillées ci-dessous et ne constituent qu'un aperçu de la diversité et de la complexité de la régulation entre les différentes voies.

NPR1 est un acteur clef entre la voie de l'acide salicylique et les voies du méthyl jasmonate. La voie de l'acide salicylique est incapable d'inhiber celle du méthyl jasmonate dans un mutant NPR1. La localisation nucléaire de NPR1 n'est pas requise pour inhiber la voie du méthyl jasmonate. NPR1 est également capable de réguler négativement la voie de l'acide salicylique. NPR1 inhibe ainsi la production d'acide salicylique durant l'attaque d'herbivores et permet à la voie du méthyl jasmonate de s'exprimer pleinement contre les herbivores (Koornneef and Pieterse 2008).

La MAP KINASE 4 (MPK4) est une protéine kinase régulant l'interaction entre les voies de l'acide salicylique et du méthyl jasmonate. Elle est activée rapidement après la reconnaissance du pathogène (Turner, Ellis et al. 2002). Le mutant *mpk4* est caractérisé par une surexpression des gènes PR, l'accumulation d'acide salicylique et l'inhibition de la voie du méthyl jasmonate. La MPK4 inhibe les voies de l'acide salicylique et active celles du méthyl jasmonate. Un des substrats de la MPK4 est MKS1 (*MAPK4 Substrate 1*), dont la phosphorylation réprime les voies de l'acide salicylique (Koornneef and Pieterse 2008). MKS1 interagit avec deux facteurs de transcription de type WRKY (WRKY25 et 33) et peut empêcher leur activation par MPK4 (Andreasson, Jenkins et al. 2005). Les facteurs de transcription WRKY constituent une famille de facteurs de transcription principalement impliqués dans la régulation de la réponse aux pathogènes, pouvant agir comme régulateurs

positifs ou négatifs des réponses de défense. Les facteurs de transcription sont des cibles d'effecteurs privilégiés permettant le contournement de la réponse de défense par les pathogènes (Eulgem and Somssich 2007). La sur ou sous expression de MKS1 n'affecte pas l'expression de PDF1.2 et indique que la MAPK4 a d'autres cibles que MKS1 permettant la régulation entre autres du méthyl jasmonate (Koornneef and Pieterse 2008). Le mutant *mapk4* empêche l'expression de gènes impliqués dans la voie du jasmonate comme PDF1.2 et THI2.1 (Turner, Ellis et al. 2002).

Le facteur de transcription WRKY70 joue un rôle clef dans la régulation des voies de l'acide salicylique et du jasmonate. La surexpression de WRKY70 est caractérisée par une expression constitutive de la voie de l'acide salicylique et de l'expression de gènes de résistance. A l'inverse, l'inexpression de WRKY70 se traduit par l'activation de la voie du jasmonate. WRKY régule positivement la voie de l'acide salicylique et négativement celle du jasmonate (Lorenzo and Solano 2005).

3 Les marqueurs moléculaires

La différence entre individus d'une même espèce réside dans le polymorphisme du génome, certaines séquences étant conservées, d'autres présentant des variations de séquence. On distingue ainsi trois types de polymorphismes: le polymorphisme de séquence, le polymorphisme d'insertion- délétion et le polymorphisme de nombre d'unités de répétitions dans les régions répétées.

Il est possible de mettre en évidence le polymorphisme de séquence directement par un séquençage de fragments homologues. Le recours au séquençage dans une approche haut débit n'est actuellement pas financièrement envisageable. Les marqueurs moléculaires apparaissent ainsi comme une solution alternative pour caractériser des génomes.

Un marqueur moléculaire est défini comme un locus polymorphe capable de renseigner sur le génotype de l'individu qui le porte et d'informer sur le génotype d'un ou de plusieurs locus voisins. Le marqueur a une position définie dans le génome et doit idéalement présenter les caractéristiques suivantes :

- Le marqueur doit être polymorphe, c'est-à-dire qu'il doit posséder plus d'un allèle au moins dans la population étudiée.
- Le marqueur idéal est codominant, ce qui signifie qu'un hétérozygote peut être différencié de l'homozygote au locus en question.
- Il est non épistatique, c'est-à-dire que le génotype peut être lu à partir de son phénotype sans influence du génotype des autres locus. Il y a une absence d'interactions intra et inter locus.
- Le marqueur est neutre, une modification des locus marqueurs n'a pas d'autres effets phénotypiques que ceux qui permettent de déterminer son génotype.
- Le génotype peut être inféré à partir du phénotype quel que soit le milieu, un « bon » marqueur moléculaire est donc insensible au milieu (Vienne 1998).

Différents types de marqueurs moléculaires sont actuellement utilisés, dont les marqueurs RFLP (Botstein D 1980), AFLP (Vos P 1995), SRR (Liu 2004), ... Dans le cadre de ce travail, nous nous focaliserons uniquement sur les marqueurs moléculaires de type SNP.

Les marqueurs moléculaires sont classés en deux catégories : les marqueurs moléculaires de type I sont associés à un gène de fonction connue. Les marqueurs moléculaire de type II sont liés à des parties anonymes du génome. Les SNP font majoritairement partie des marqueurs moléculaires de classe II. Cependant les SNP mis en évidence sur base d'EST (*expressed sequence tag*) seront définis comme des marqueurs moléculaires de classe I (Liu 2004).

Les SNP (prononcés « snip ») ou *Single Nucléotide Polymorphism* constituent la majorité du polymorphisme du génome (90% du polymorphisme du génome humain (Brookes 1999)). Les SNP sont des mutations qui ne concernent qu'un seul nucléotide. Remarquons avant tout la différence entre polymorphisme et mutation : un SNP provient d'une mutation dans une population. Mais lors de l'observation d'un variant pour une séquence donnée dans une population, le SNP observé n'est plus une mutation car il correspond à un évènement depuis longtemps révolu, il n'est plus qu'« un variant de séquence » (Gibson 2004).

On observe un SNP lors du changement d'une base en une autre. Les insertions et délétions d'une seule base sont également reconnues comme SNP. Le changement de base résultant d'une mutation peut être une transition soit d'une purine en purine ($A \leftrightarrow G$) ou d'une pyrimidine en pyrimidine ($C \leftrightarrow T$). Les transversions sont la mutation d'une purine en pyrimidine ou une pyrimidine en purine ($A \leftrightarrow C$, $A \leftrightarrow T$, $G \leftrightarrow C$, $G \leftrightarrow T$) (Vignal, Milan et al. 2002).

La classification la plus courante des SNP se réfère à la nature de la base modifiée. Ainsi, les SNP non codants se trouvent dans les régions 5' et 3' UTR et dans les introns. Les SNP codants peuvent être des polymorphismes de substitution - changement de l'acide aminé codé - ou des polymorphismes synonymes - changement du codon mais l'acide aminé reste le même. Les fréquences des différentes combinaisons de SNP possibles dans le génome ne sont pas les mêmes. Ainsi deux tiers des SNP impliquent une transition de C en T (Brookes 1999).

Les SNP sont distribués de manière régulière à travers l'ensemble des génomes. Chez l'Homme, la fréquence moyenne des SNP est de un SNP/ 500 paires de bases (Christine KEYSER 2006).

Ce nombre élevé de SNP permet la création de cartes génétiques de haute densité. Les SNP permettent encore l'étude de recombinaisons et de réarrangements chromosomiques (Pflieger 2001; Gibson 2004). Avec leur faible taux de mutations par génération (10^{-8}) les SNP peuvent servir de marqueurs pour suivre l'évolution des espèces (Crow et al., 1995; Li et al., 1996 ; Morin 2004). Le faible taux de mutation par génération permet également l'utilisation des SNP pour réaliser des tests de paternité (Sobrino, Brion et al. 2005; Wang 2006).

Actuellement, les SNP du génome humain sont étudiés principalement dans le cadre du projet HapMap (voir www.hapmap.org). Ce projet HapMap définit des SNP comme étant des marqueurs uniques d'un haplotype. Le but du projet HapMap est de pouvoir définir des haplotypes, soit des régions conservées, pouvant par exemple être retrouvées chez plusieurs individus d'une même origine. Un haplotype peut être défini comme une combinaison d'allèles caractérisant un individu. Ainsi, dans une population tirée au hasard, on peut trouver différents haplotypes en différentes proportions. Le projet HapMap identifie les haplotypes courant dans quatre populations issues de divers endroits du monde. « C'est ainsi qu'en détectant les SNP marqueurs d'une personne (génotypage), les chercheurs seront en mesure

de caractériser tous les haplotypes présents dans le génome de cette personne. On estime que le nombre de SNP nécessaires pour contenir la plupart de l'information sur la variation génétique se situe entre 300,000 et 600,000, soit beaucoup moins que les 10 millions de SNP courants ». A long terme, on espère pouvoir faire le lien entre les haplotypes et un phénotype donné. Ainsi par exemple, dans le domaine médical, un haplotype donné correspondrait à un facteur tendant vers un certain type de pathologie.

Comme introduit ci-dessus le projet HAPMAP, les SNP s'inscrivent en génomique fonctionnelle. Ils peuvent également être utilisés pour des études de génomique fonctionnelle de gènes candidats, ceux-ci étant des gènes susceptibles d'être impliqués dans une fonction d'intérêt que l'on va pouvoir étudier grâce aux SNP qui les caractérisent. Certains SNP, se trouvant dans des régions codantes, affectent le phénotype en changeant un acide aminé en un autre. On a pu ainsi montrer que la présence de certains SNP caractérisait une maladie ou une propension à développer une maladie (maladie d'Alzheimer, Martin et al, 2000 ou le diabète type II Horikawa et al, 2002). La présence d'un SNP dans une région régulatrice, promotrice, peut aussi affecter la transcription d'un gène. La majorité des SNP n'ont cependant pas d'effet direct sur le phénotype !

3.1 Recherche de SNPs

La découverte de SNP peut être réalisée à l'aide de différentes approches. La plus simple nécessite le séquençage d'une région donnée chez différents individus et consiste à mettre en évidence la présence de SNP par alignement de séquences. Le coût actuel du séquençage empêche de pouvoir utiliser cette approche dans le cadre d'analyses à haut débit (Vignal, Milan et al. 2002). Des méthodes alternatives ont été mises au point pour permettre la détection de nouveaux SNP. Dans le cadre de ce travail, nous aborderons la technique de l'EcoTilling.

3.1.1 L'EcoTILLING

L'EcoTILLING est une adaptation de la technique du TILLING, *Targeting Induced Local Lesions in Genomes*, qui permet la mise en évidence du polymorphisme, dont des SNP, dans une population.

Tout d'abord, le TILLING est une méthode combinant des mutations induites chimiquement et le crible d'une séquence d'intérêt par PCR. Une région d'ADN allant jusqu'à 1kpb est amplifiée avec des amorces fluorescentes pour un ensemble de souches d'individus mutés. Les parties mutées correspondent à des mauvais appariements entre bases. Ces mutations sont reconnues par l'endonuclease CEL I ou ENDO I (Triques 2008) qui va cliver l'ADN aux endroits de mutations. De cette manière, on génère des fragments fluorescents de tailles différentes. La position de la mutation déterminera la distance de migration des fragments sur un gel d'électrophorèse (Colbert 2001). Le TILLING peut être facilement automatisé pour être utilisé dans des approches à haut débit (McCallum 2000).

La technique de l'EcoTILLING va se servir de la capacité du TILLING à détecter des mutations et va être adaptée à la mise en évidence du polymorphisme principalement de type SNP. L'EcoTILLING peut être appliqué à tout individu (du monde animal, végétal ou bactérien). Si le TILLING permet la mise en évidence de mutations induites par l'expérimentateur, l'EcoTILLING met en évidence le polymorphisme naturel de type SNP, INDEL (*insertion deletion*), SSR (*single strand repeat*). L'EcoTILLING est une technique

qui permet de mettre rapidement en évidence le polymorphisme et de dresser des haplotypes dans une population. L'EcoTILLING apparait comme une alternative au séquençage systématique de différents individus pour permettre la mise en évidence du polymorphisme dans une population (Barkley, Wang et al. 2008).

Un individu de référence est choisi et un pool constitué d'individus différents, nommés écotypes, est formé. L'écotype sera comparé à l'individu de référence (Comai, Young et al. 2004). Le principe de l'EcoTILLING est illustré par la figure 1.16

On amplifie une séquence d'ADN pouvant aller jusqu'à 1500 pb par PCR de l'individu de référence ainsi que des différents individus de l'écotype (Gilchrist, Haughn et al. 2006). On utilise des amorces fluorescentes qui permettent l'obtention d'un produit marqué. Dans chaque puits d'une plaque multi-puits sont déposés des fragments amplifiés de l'individu de référence, ainsi que d'un écototype. Chaque puits de la plaque correspond à une comparaison entre un individu de l'écotype et l'individu de référence. Après dénaturation, les brins d'ADN sont réhybridés, formant des hétéroduplex entre l'ADN de l'individu de référence et l'ADN de l'individu étudié.

Un SNP, soit une différence d'un nucléotide dans la séquence d'un individu par rapport au modèle de référence, apparaîtra dans cet hétéroduplex comme un mauvais appariement de bases. L'ajout d'une endonucléase (CEL I, ENDO1) comme dans le TILLING, qui coupe l'ADN à chaque mésappariement entraîne la création de fragments de tailles différentes. Les différents fragments du produit de digestion sont passés dans un analyseur de taille.

Si l'individu testé ne présente aucune variation génétique par rapport à l'individu de référence dans la séquence amplifiée, les deux fragments d'ADN s'hybrideront parfaitement, l'ADN ne sera pas digéré et les fragments auront une taille égale à celle du fragment amplifié par PCR. Si au contraire, l'individu étudié présente un SNP, un changement de base à un endroit donné par rapport à l'individu de référence, il y aura hybridation des deux fragments d'ADN mais avec mésappariement de base au niveau du SNP. Après digestion de l'ADN par l'enzyme de restriction, la taille du fragment résultant sera fonction de la position du SNP dans la séquence amplifiée. La présence de plusieurs SNP dans la séquence se traduira par la présence de fragments de plusieurs tailles détectés. La taille des fragments permet de localiser les SNP sur la séquence.

L'EcoTILLING permet la détection rapide de différents haplotypes ; les haplotypes d'intérêts pouvant être séquencés par la suite. L'EcoTILLING peut être appliqué à la mise en évidence non seulement de SNP, d'insertion/délétion mais également de microsatellites. L'EcoTILLING permet une étude facile du polymorphisme avec un excellent taux de détection. Il est possible de mettre en évidence plusieurs SNP dans une séquence définie en une seule étape. Cette méthode est applicable à n'importe quel organisme, hétérozygote et polyploïde, soit présentant un ou plusieurs polymorphismes sur deux ou plusieurs allèles. Un autre avantage de l'EcoTILLING est que cette méthode ne nécessite aucune connaissance préalable de la séquence à analyser, excepté pour la création des amorces PCR (Comai, Young et al. 2004).

Dans le cadre d'études à haut débit, l'EcoTILLING peut être réalisé en pool. Huit individus différents peuvent ainsi être analysés simultanément. Dans une analyse sur pool, le taux de

détection de SNP est toujours élevé mais avec un taux de faux positifs de 5% et 4% de faux négatifs (Till, Zerr et al. 2006).

3.1.2 Typage de SNP : analyse microarray

L'EcoTILLING permet la mise en évidence de nouveaux SNP. Le typage de SNP ou *SNP typing* correspond aux méthodes utilisées pour mettre en évidence des SNP connus. Des plateformes de typage de SNP permettent d'analyser et caractériser à haut débit une population d'individus testés.

Les techniques actuelles à haut débit s'inspirent principalement de trois méthodes : l'hybridation allèle spécifique, l'extension d'amorces et la ligation d'oligonucléotides. Nous détaillerons d'abord ces trois méthodes (illustration figure 1.17) et nous décrirons ensuite la technique Illumina (voir figure 1.18) permettant le typage de SNP à haut débit sur support microarray.

3.1.2.1 Typage de SNP : Allèle spécifique hybridation

Cette méthode de typage de SNP se base sur la mise en évidence d'un SNP connu par hybridation d'une sonde spécifique au SNP et est également appelée hybridation spécifique d'oligonucléotides (ASO) (Sobrinho, Brion et al. 2005). Un SNP peut être mis en évidence à l'aide d'une sonde nucléotidique complémentaire à la séquence contenant le SNP par hybridation spécifique. Un SNP étant par définition bi allélique, deux sondes sont suffisantes pour discriminer la nature du SNP. La sonde et les conditions expérimentales sont optimisées pour l'hybridation de la sonde spécifique (Gut 2001). (Figure 1.17 A.)

3.1.2.2 Typage de SNP : Extension d'amorces

Cette méthode se base sur l'extension d'une sonde oligonucléotidique située en amont du SNP à typer. Deux versions de cette méthode sont utilisées. La première utilise deux sondes dont le dernier nucléotide en 3' se trouve sur le SNP. Si le dernier nucléotide de la sonde est complémentaire du SNP, il y a hybridation parfaite. L'extension de la sonde peut se réaliser par l'ajout de nucléotides et de l'ADN polymérase comme lors d'une réaction PCR. Si le dernier nucléotide de la sonde n'est pas le complémentaire du SNP, il y a un mismatch et l'ADN polymérase n'est pas capable d'ajouter les nucléotides après la sonde. La deuxième version du typage de SNP par extension de primer est appelée extension de base simple (SBE). La sonde est construite pour aller s'hybrider juste avant la position du SNP à caractériser. On utilise ensuite l'ADN polymérase pour incorporer un ddNTP marqué. L'utilisation de ddNTP ne rend possible l'incorporation que d'un seul nucléotide. Chaque nucléotide ayant été marqué de manière différente, il est possible de savoir la nature du nucléotide complémentaire du SNP et par extension la nature du SNP typé (Vignal, Milan et al. 2002). (Figure 1.17 B.)

3.1.2.3 Typage de SNP : Oligonucléotides ligation

L'enzyme ADN ligase est capable de réparer des coupures dans l'ADN en reformant un pont diester. La méthode de ligation d'oligonucléotides nécessite trois sondes : deux sondes allèles spécifiques et une sonde locus spécifique. Les deux sondes allèles spécifiques sont variables pour le dernier nucléotide en position 3'. La sonde locus spécifique est conçue pour s'hybrider directement après la position du SNP à typer. L'enzyme ADN polymérase est capable de lier la sonde allèle spécifique complémentaire du SNP testé à la sonde locus spécifique. Si l'hybridation de la sonde allèle spécifique n'est pas parfaite, le mismatch en position 3' de la sonde empêche toute ligation. En discriminant l'une et l'autre sonde allèle

spécifique, il est possible de connaître la nature du SNP typé (Sobrinho, Brion et al. 2005). (Figure 1.17 C.)

3.1.2.4 Typage de SNP sur support solide : Illumina GoldenGate assay

Les trois méthodes ci-dessus peuvent être adaptées sur support liquide ou solide. La méthode *Illumina GoldenGate assay* permet déjà de typer simultanément 683,000 (info constructeur) SNP chez l'Homme par microarray. Elle pourrait être adaptée pour le typage de SNP mis en évidence chez la chicorée. Nous détaillerons ci-dessous le principe général de cette technique.

La séquence d'ADN à typer pour un SNP défini est mise en présence de deux sondes oligonucléotidiques allèle spécifique, chacune étant spécifique d'un allèle différent du SNP à son extrémité 3', et d'une sonde locus spécifique. Cette dernière s'hybride spécifiquement quelques nucléotides après le SNP à déterminer. Les deux sondes allèles spécifiques contiennent respectivement un site d'amplification P1 ou P2. La sonde locus spécifique contient elle aussi un site d'amplification P3. Une étape d'hybridation, extension et ligation permet de lier la sonde allèle spécifique du SNP avec la sonde locus spécifique. En fonction de l'allèle du SNP, une combinaison sonde allèle spécifique et locus spécifique est obtenue. Le complexe est amplifié simultanément avec les amorces P1 et P2, marquées en fonction de l'allèle, et avec l'amorce P3 non marquée. Une séquence d'adressage (zip code) comprise dans la sonde locus spécifique permet la fixation spécifique de l'hybride sonde allèle spécifique et locus spécifique sur une séquence complémentaire localisée à un point donné sur un microarray. Le produit PCR est déposé sur la puce microarray et va s'hybrider spécifiquement sur la séquence complémentaire du zip code. En observant la nature de la fluorescence de ce point, il est possible de mettre en évidence la nature du SNP (allèle 1 et/ou allèle 2). L'ensemble de la technique est illustrée dans la figure 1.18.

3.2 L'approche gène candidat et l'utilisation de marqueurs moléculaires : génétique d'association

L'amélioration d'un trait quantitatif d'une plante peut être abordée par différentes approches comme par exemple des approches de mutagenèse dirigée. La fonction d'un gène (par exemple la résistance ou la sensibilité à un pathogène) est mise en évidence par l'inactivation de celui-ci. Cependant, ce type d'approche est lourd à mettre en place pour l'analyse d'un grand nombre de gènes. Une approche alternative à la mise en évidence de gènes d'intérêts est l'approche « gène candidat ». Cette approche est réalisable chez l'Homme, les animaux et les plantes. Elle se base sur l'hypothèse que le polymorphisme d'un gène candidat est lié à une variation phénotypique. En d'autres termes, le polymorphisme d'un gène est associé à un trait phénotypique comme par exemple la résistance ou la sensibilité à un pathogène. Un gène candidat est défini comme tout gène polymorphe pouvant être impliqué dans la variation d'un trait d'intérêt.

Il existe des gènes candidats fonctionnels, soit des gènes dont la fonction est associée à la problématique abordée, et des gènes candidats positionnels mis en évidence sur base de parties du génome qui co-ségrègent avec un phénotype donné (Pflieger 2001; Varshney, Graner et al. 2005). Des gènes candidats fonctionnels peuvent être aisément mis en évidence sur base d'homologies de séquences. Un gène présentant une grande homologie avec un gène déjà décrit a une certaine probabilité d'être associé à cette fonction. L'analyse BLAST est actuellement une des méthodes utilisées pour la mise en évidence d'homologies de séquences (Varshney, Graner et al. 2005). Dans le cas de gènes candidats positionnels, il est nécessaire de disposer préalablement d'analyses QTL. Un QTL, *quantitative trait loci*, est un locus

pouvant contenir plusieurs gènes associés à un caractère quantitatif. Un gène présent dans un QTL peut être un gène candidat positionnel car il est lié avec le phénotype même si sa fonction n'est a priori pas liée au phénotype (Morgante and Salamini 2003; Varshney, Graner et al. 2005).

Le polymorphisme du gène candidat est caractérisé à l'aide de marqueurs moléculaires. Comme introduit précédemment, les SNP apparaissent comme des marqueurs de choix pour la caractérisation des gènes candidats : grande densité de SNP à travers tout le génome ; faible taux de mutation ; possibilité de détection à haut débit (Zhu 2008).

Les études d'association permettent de relier le polymorphisme mis en évidence dans les gènes à un phénotype observé. Deux approches permettent de faire le lien entre polymorphisme et phénotype : l'approche QTL présentée ci dessus et l'approche statistique. L'approche statistique repose sur la mise en évidence de la corrélation entre le polymorphisme caractérisé chez plusieurs individus et leurs phénotypes respectifs. Selon les données disponibles on aura recours à des analyses de variance (ANOVA), de régression linéaire, Chi carré, ... (Pflieger 2001; Dauchot 2007; Zhu 2008).

Les études d'association peuvent être réalisées sur l'entièreté d'un génome. Ces analyses *Whole-Genome Scan* à très haut débit reposent sur le séquençage de nombreuses régions du génome de différents individus, permettant la mise en évidence de SNP en comparant ces derniers. Les études d'associations sur base de gènes candidats sont actuellement plus faciles à mettre en application (d'un point de vue pratique et financier) que l'approche *Whole-Genome Scan* limitée par le coût de séquençage et des plateformes de typage de SNP à très haut débit (Zhu 2008).

3.3 Objectifs du mémoire

Comme indiqué plus haut, la culture de la chicorée industrielle, *Cichorium intybus*, est principalement affectée par deux maladies : l'oïdium causé par *Erysiphe cichoracearum* et la rouille causée par *Puccinia cichorii*.

Notre premier objectif est de mettre en évidence et de caractériser le polymorphisme de gènes candidats associés à la résistance de la chicorée présents dans les cultivars à l'origine de la sélection de la chicorée industrielle.

L'approche gène candidat et des études d'association ont déjà été réalisées avec succès dans la mise en évidence d'allèles d'intérêts associés à la résistance de la plante à un pathogène (Pflieger 2001).

La recherche de gènes candidats fonctionnels de la résistance aux pathogènes chez la chicorée s'articule en trois axes : la recherche de gènes NBS-LRR (fréquemment désignés comme des gènes candidats fonctionnels de choix (Michelmore 2000; McDowell and Woffenden 2003; Liu, Liu et al. 2007; McDowell and Simon 2008)), la recherche de gènes WAK et la mise en évidence de gènes impliqués dans les voies de réponse aux pathogènes. Nous avons également recherché des orthologues de la laitue positionnés sur une carte génétique en utilisant la méthode du *reciprocal best BLAST hit*.

Nous avons également souhaité annoter la banque d'unigènes de chicorée à l'aide du logiciel Interproscan permettant la recherche de domaine protéique. L'annotation permettra une mise en évidence et une recherche facile de séquences associées à un mot clef et ainsi de trouver de nouveaux gènes candidats associés à la résistance aux pathogènes chez la chicorée.

La mise en évidence du polymorphisme de gènes candidats se fera sur base de la technique de l'EcoTILLING. A partir des résultats obtenus, des clones seront séquencés afin de confirmer la présence de SNP et de caractériser ces derniers. Une analyse EcoTILLING sur des individus actuels sera réalisée afin de déterminer si du polymorphisme ancien est toujours présent dans les variétés de chicorées actuelles.

Nous essayerons d'optimiser au maximum la caractérisation du polymorphisme dans une approche haut débit et abordable.

A plus ou moins long terme, les résultats obtenus lors de ce mémoire serviront à la réalisation d'une puce microarray de type Illumina et serviront à la réalisation études d'associations statistiques d'une population (d'individus sensibles et résistants) phénotypée et génotypée à l'aide du microarray qui devront permettre la mise en évidence de SNP associés à la résistance aux pathogènes. L'objectif final est de pouvoir sélectionner ou réintroduire des allèles de résistance perdus lors de la sélection classique de la chicorée dans les variétés actuelles. Il est également possible de voir si les SNP caractérisés dans la core collection et toujours présents actuellement peuvent être associés à la sensibilité ou à la résistance de variétés actuelles.

1 Recherches Bioinformatiques

1.1. Séquences utilisées dans le cadre des recherches bioinformatiques.

Les séquences EST utilisées dans le cadre de la banque d'unigènes de chicorée proviennent du projet PHYTOMOL (FUNDP, Namur), de la banque de données de Lille (<http://www.lille.inra.fr/>) et du CGPDB (Compositae genome project database, <http://cgpdb.ucdavis.edu/>). Les séquences utilisées dans ce travail peuvent également avoir pour origine la banque de données d'NCBI (<http://www.ncbi.nlm.nih.gov/>) et la banque de données d'*A. thaliana* (<http://www.arabidopsis.org/>).

1.2. Format des séquences

Les séquences nucléotidiques et protéiques utilisées sont manipulées en format FASTA (extension .fasta). Ce format est défini un « header » débutant par le symbole « plus grand que », « > », suivi de la référence de la séquence ainsi qu'une brève description de la séquence. Le header est ensuite suivi à la ligne par la séquence nucléotidique ou protéique. Les fichiers FASTA sont éditables facilement à l'aide d'éditeurs de texte (bloc-notes, Word, ...) et compatibles avec un grand nombre de logiciels bioinformatiques.

1.3. Logiciels installés en local

La liste des logiciels ci-dessous reprend ceux nécessitant une installation pour être utilisés. Système d'exploitation utilisé : Windows Vista Premium SP1

1.3.1. SeqTools 8.4

Disponible sur le site de son concepteur (S. W. Rasmussen) www.seqtools.dk/, le logiciel SeqTools 8.4 a été utilisé pour la création de banques de données locales et l'analyse BLAST. Ce programme permet l'exportation de séquences en format FASTA et l'exportation de résultats BLAST en format Excel.

1.3.2. Vector NTI 10

Commercialisé par Invitrogen, Vector NTI10 est utilisé pour l'alignement de séquences (AlignX), la formation de contigs (ContigExpressProject), l'édition de séquences nucléotidiques, la visualisation des sites de restriction, l'analyse d'amorces (Oligo Testing), l'exportation d'arbres phylogéniques, ...

1.3.3. TreeView 1.6.6

Freeware disponible sur www.taxonomy.zoology.gla.ac.uk/rod/treeview.html, TreeView (auteur : D. M. Roderic) permet la visualisation et l'édition d'arbres phylogéniques.

1.3.4. TreeIllustrator

Disponible sur le site <http://nexus.ugent.be/geert/>, TreeIllustrator permet la visualisation et l'édition d'arbres phylogéniques de manière plus complète que TreeView. Nécessite Java virtual machine.

1.3.5. Blast2Go

Nécessitant l'installation préalable de Java virtual machine, Blast2Go est disponible gratuitement à l'adresse suivante : <http://blast2go.bioinfo.cipf.es/>. Le logiciel est utilisé pour la manipulation et l'extraction de séquences provenant de grandes bases de données.

1.4. Logiciels Online

Les logiciels ci-dessous sont accessibles via le navigateur web et ne nécessitent aucune installation.

1.4.1. Primer3

Disponible gratuitement à l'adresse suivante: <http://frodo.wi.mit.edu/>. Primer3 est un logiciel entièrement paramétrable pour la création d'amorces PCR.

1.4.2. NCBI BLAST

NCBI BLAST permet différentes analyses BLAST (BLASTX, TBLASTX,...) contre de multiples banques de données. Il est accessible sur <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.

1.4.3. Pfam 23.0

Développé par l'institut Sanger, Pfam est un logiciel en ligne permettant la prédiction de domaines sur base de séquences protéiques. Il est disponible à l'adresse suivante : <http://pfam.sanger.ac.uk/>.

1.4.4. Interproscan v4.1

Interproscan est un outil permettant la prédiction de domaines et est disponible sur www.bioinformatics.nl/iprscan/iprscan. Interproscan permet l'analyse simultanée de séquences suivant une sélection d'algorithmes de prédiction de domaines et se réfère à différentes bases de données simultanément. Une interface graphique permet de visualiser les domaines. Il est également possible d'exporter les résultats dans un fichier texte. Une version locale du logiciel est disponible gratuitement mais requiert un système d'exploitation Unix et *perl* ainsi qu'un espace libre de minimum 20 Go pour les bases de données.

1.4.5. Spidey

Permet l'alignement entre la séquence ADNc et la séquence d'ADN génomique sur <http://www.ncbi.nlm.nih.gov/spidey/spideyweb.cgi>.

1.5. L'analyse *reciprocal best BLAST hit*

Basée sur des analyses BLAST, la méthode du *reciprocal best BLAST hit* permet la mise en évidence d'orthologues entre deux espèces. Pour rappel, deux séquences présentes dans deux espèces différentes sont orthologues si ces deux séquences sont apparues suite à un événement de spéciation.

Deux séquences A et B sont orthologues si et seulement si elles remplissent toutes les conditions ci dessous:

- ✓ La séquence A appartient à la banque de données X (espèce1)
- ✓ La séquence B appartient à la banque de données Y(espèce2)
- ✓ La séquence A est le meilleur résultat de BLAST quand la séquence B est blastée contre la banque de donnée X.
- ✓ La séquence B est le meilleur résultat de BLAST quand la séquence B est blastée contre la banque de donnée Y.

Les analyses BLAST (TBLASTX, paramètres par défaut) ont été réalisées avec SeqTools 8.4.

2. Recherches expérimentales

2.1. Matériel végétal

Le matériel de base se compose de jeunes pousses de chicorée de 4 semaines présentant le cotylédon ainsi que de jeunes feuilles. Les échantillons proviennent de 29 cultivars de chicorée utilisés à l'origine de la sélection de la chicorée industrielle. Pour chaque cultivar, nous disposons de trois échantillons. Ceux-ci ont été fournis par SA Warcoing en eppendorf stockés à -20°C.

2.2. Extraction de l'ADN de chicorée

Le protocole d'extraction de l'ADN des échantillons de chicorée est divisé en trois étapes principales : destruction mécanique de la paroi cellulaire végétale, lyse de la membrane cellulaire et extraction de l'ADN. Le protocole se base sur le kit d'extraction et de purification Nucléon Phytopure *Amersham Biosciences* #RPN 8510.

Deuxième purification réalisée suivant le système Nucleospin Extract II (*Macherey-Nagel* #740 609.50)

Les ADN purifiés ont été stockés à -20°C.

2.3. Définition et validation des amorces PCR

- Taille du produit amplifié sur ADNc comprise entre 350 et 500 pb,
- 50% de GC
- *GC clamp*=2. (Si aucune amorce n'est trouvée suivant ces paramètres le *GC clamp* est descendu à 1)
- Tester les propriétés thermodynamiques des amorces avec VectorNTI10. Les amorces ne sont pas validées lorsqu'elles forment des dimères ou lorsqu'il y a formation de boucle(s) à l'intérieur du primer et que ces interactions présentent des valeurs de ΔG inférieures à -8 kcal/mol.
- Ajouter un site de restriction SpeI (ACTAGT) à chaque extrémité 5' des amorces *forward*. La séquence « GG » est également ajoutée pour améliorer la restriction.
- Ajouter un site de restriction SalI (GTCGAC) à l'extrémité 5' de chaque amorce *reverse*. La séquence « GCG » est également ajoutée pour améliorer l'efficacité de la restriction enzymatique (Moreira and Noren 1995).
- Re-suspendre les amorces dans de l'eau, du Tris 10mM ou du TE 10/0,1 pH8.0 afin d'obtenir une solution stock à 100µM.
- Tester l'amplification avec le mix PCR suivant : 0,5 µL dNTP (10mM); 0,4 µL Taq (5 U/µL); 0,3 µL amorce *forward* (100 µM); 0,3 µL amorce *reverse* (100µM); 1µL Pool DNA (5 ng/µl); 10µL Buffer 5X; 38 µL H₂O. Les paramètres PCR utilisés sont: 4 minutes à 94°C (étape de dénaturation longue), 45 secondes à 92° (étape de dénaturation), 45 secondes à 55°C (étape d'hybridation des amorces), 45 secondes à 72°C (étape d'élongation), 10 minutes à 72°C (étape d'élongation longue), 4°C à l'infini (conservation du produit PCR). Les étapes de dénaturation, d'hybridation et d'élongation sont répétées 35 fois.

- Dans le cas de non amplification ou d'amplifications non spécifiques, une PCR sur gradient (températures étalées entre 55 et 68°C) est réalisée afin de déterminer les paramètres optimums pour les amorces testées.

2.4. Bactéries thermocompétentes TOP 10 F'

Le protocole suivant permet la fabrication et la transformation de bactéries thermocompétentes TOP10F' (*Invitrogen* #C303003). Le protocole utilisé a été adapté du protocole disponible à l'adresse suivante : [http:// openwetware.org/wiki/TOP10_chemically_competent_cells](http://openwetware.org/wiki/TOP10_chemically_competent_cells).

Le protocole nécessite l'utilisation de tampon CCMB80 dont la composition est : KOAc 10 mM pH 7,0; CaCl₂.2H₂O 80 mM (VWR #1226); MnCl₂.2H₂O 20 mM (*Merck* #5934); MgCl₂.6H₂O 10 mM (*Merck* #1.05833.0250); 10% glycerol (*Sigma* #G-7757). Le tampon est ajusté à pH 6,4 et est stérilisé via l'utilisation d'un filtre 0,22 µm avant d'être stocké à 4°C.

Les bactéries TOP10F' sont sensibles aux détergents. On veillera tout au long du protocole à utiliser des récipients lavés à l'eau désionisée.

- Lancer une préculture overnight de bactéries TOP10F' dans 10 mL de milieu LB (*BD* #240230) avec de la tétracycline (50 µg/mL; *Sigma*)
- Sur base de la préculture, lancer une culture de jour dans 400 mL de milieu LB avec une concentration en tétracycline de 50 µg/mL.
- Incuber à 37°C et 180 rpm dans un incubateur thermostaté (C24 Incubator Shaker (*New Brunswick Scientific*) jusqu'à obtenir une DO de 0,3 à 590 nm (Spectrophotomètre UV-1601PC, *Shimadzu*)
- Transférer les bactéries en solution dans 4 tubes Falcon de 50 mL.
- Centrifuger à 4500 rpm 10 minutes à 4°C
- Eliminer le surnageant et resuspendre les culots dans un volume total de 80 mL de tampon CCMB80 stocké à 4°C
- Répartir les bactéries dans deux tubes contenant chacun 40mL et les conserver sur glace pendant 20 minutes.
- Centrifuger à 4,500 rpm 10 minutes à 4°C
- Resuspendre dans 8 mL de tampon CCMB80
- Diluer 200 µL de bactéries TOP10F' dans 800 µL de milieu LB.
- Diluer dans du tampon CCMB80 jusqu'à obtenir une DO comprise entre 1.0 et 1.5 à 590 nm.
- Incuber 20 minutes sur glace et aliquoter les bactéries en eppendorf par 50 µL
- Stocker à -80°C

Remarque : la capacité de transformation des bactéries TOP10F' est affectée par des congélations décongélations répétées.

2.5. Transformation des bactéries thermocompétentes TOP10F'

Ce protocole utilise du milieu SOC stérile à pH 7 contenant : 2% bacto-peptone (*BD* #211677); 0,5 % yeast extract (*BD* #212750); KCl 2,5 mM (VWR #1591); MgCl₂ 10 mM (*Merck* #1.05833.0250); MgSO₄ 10 mM (VWR #1433). Mise à volume d'1 L avec de l'eau et mise à pH 7. Après autoclave ajouter 20 mM de glucose (*Riedel-deHaën* #16301) stérilisé sur filtre 0,22 µM.

- Ajouter 1 µL de plasmide pBlueScript II KS+ (500ng) à 50 µL de bactéries thermocompétentes

- Garder 30 minutes sur glace.
- Heat Shock 30 seconde à 42°C.
- Garder 2 minutes sur glace.
- Ajouter directement 250 µL de milieu SOC (T° ambiante).
- Incuber 1 h à 37°C pour permettre aux bactéries de récupérer.
- Etaler les bactéries sur des boîtes de LB Agar (*Fluka* #05038) avec ampicilline (60 µg/ml; *Sigma* #A0166), X-gal (87,5 µg/ml; *Promega* #V3941) et IPTG (41,7 µg/ml) et incuber overnight à 37°C

2.6. Préparation du plasmide

Le plasmide utilisé pour le clonage des gènes candidats amplifiés est le plasmide pBluescript II KS+. Ce plasmide est produit dans des bactéries TOP10F' (*Invitrogen* #C303003). Le plasmide est extrait avec le kit « *QIAGEN Plasmid Maxi kit* ».

2.6.1. Maxiprep : *QIAGEN Plasmid Maxi kit* #12162

- Au départ d'une colonie isolée de bactéries TOP 10F' transformées préalablement avec le vecteur pBlueScript II KS+, lancer une préculture dans 20 mL de milieu LB avec de l'ampicilline (60 µg/mL) sous hotte à flux laminaire vertical.
- Incubation overnight de la préculture à 37°C à 180 rpm dans un incubateur C24 Incubator Shaker (*New Brunswick Scientific*).
- Dans un Erlen stérile de 500 mL, ajouter les 20 mL de la préculture dans 400mL de milieu LB (Amp 60 µg/mL). Incuber à 37°C et à 180 rpm jusqu'à obtenir une DO comprise entre 0,4 et 0,8 à 590 nm (Spectrophotomètre UV-1601PC, *Shimadzu*)
- Diviser la culture dans 2 biberons de 250 mL et centrifuger durant 15 minutes à 4°C et 6000 g (Centrifugeuse J2-21, rotor JA-14, *Beckman Coulter*).
- Eliminer le surnageant et resuspendre les culots dans 5 ml de tampon P1 (RNase) stocké à 4°C et ajouter 5 ml de tampon P2 (NaOH SDS) dans chaque biberon.
- Après 5 minutes d'incubation à température ambiante, le contenu de chaque biberon est transféré dans un tube correx stérile de 30 mL
- Ajouter 10 ml de tampon P3 (acétate) stocké à 4°C et inverser plusieurs fois les tubes. Utiliser un parafilm.
- Laisser reposer les échantillons sur glace durant 20 minutes.
- Culotter les protéines en centrifugeant 30 minutes à 4°C à plus de 20,000 g (Centrifugeuse J2-21, rotor JS-13.1, *Beckman Coulter*)
- Récolter le surnageant dans un nouveau tube
- Activer la colonne de fixation des plasmides QIAGEN-tips 500 en faisant passer 10 mL de tampon QBT à travers cette dernière. Laisser le tampon migrer par gravité.
- Passer le surnageant sur la colonne activée et rincer deux fois à l'aide de 30 mL de tampon QC.
- Eluer l'ADN fixé sur la colonne avec 15 mL de tampon QF et récolter le tout dans un tube correx stérile de 30 mL
- L'ADN est précipité en ajoutant 10,5 ml d'isopropanol (*Sigma*) stocké à température ambiante
- Centrifuger 30 minutes à 4°C à plus de 15,000 g (Centrifugeuse J2-21, rotor JS-13.1, *Beckman Coulter*)
- Laver le culot avec 5 ml d'éthanol 70% à température ambiante (*Sigma*)

- Centrifuger 10 minutes à 4°C à plus de 15,000g (Centrifugeuse J2-21, rotor JS-13.1, *Beckman Coulter*)
- Eliminer le surnageant en séchant au SpeedVac durant 5 à 10 minutes
- Suivant la DO initiale, resuspendre le culot dans 50 à 100 µl de Tris 10 mM pH 8 pour diluer la solution jusqu'à obtenir une concentration finale d'1 µg/µl. Le dosage s'effectue à l'aide du spectrophotomètre.

2.7. Restriction du plasmide

Le plasmide pBluescript ksII+ est linéarisé avec les enzymes de restriction XhoI (*NEB* #R0146S) et XbaI (*NEB* #R0145S). Le mix suivant est incubé 4h30 à 37°C dans un thermo cycler : 4 µg (XµL) pBluescript II ks+ ; 2 µL Buffer react2 10X ; 2 µL mix enzymes ratio 2/1 : XhoI (13,3 U)/ XbaI (6,67 U) ; 0,3µL BSA 100X ; Monter à 20µL avec de l'eau

Afin d'empêcher le vecteur de se recirculariser, le plasmide restreint est traité à la phosphatase alcaline (Calf Intestinal Alkaline Phosphatase = CIAP; 1 U/µl; *Promega* #M1821). Ce traitement a pour objectif de catalyser l'hydrolyse des groupements phosphates en 5' et de limiter la recircularisation du vecteur.

Composition du mix traitement phosphatase alcaline: 20 µL de produit de restriction (soit 4 µg de vecteur) ; 5 µL tampon CIAP 10X ; 1 µL (+1 µL) CIAP ; 23 µL H₂O (Vf= 50 µL)

- Incuber 30 minutes à 37°C avec 1 µL de CIAP
- Après 30 minutes ajouter à nouveau 1 µL de CIAP
- Incuber 30 minutes

2.8. Purification du plasmide restreint et des produits PCR avant clonage

Les enzymes (Taq polymérase et enzymes de restrictions) sont inactivées par une extraction PCI (phenol: chlorophorm: alcool isoamylique 25:24:1; *Sigma* #087K1890). Ce traitement est appliqué au vecteur restreint ainsi qu'aux produits PCR avant clonage et évite aux enzymes d'interférer avec l'étape de restriction-ligation.

- Ajouter 50 µL de PCI à l'ependorf contenant l'échantillon à purifier
- Mélanger par inversion
- Centrifuger durant 5 minutes à 4°C et 13,000 rpm (Biofuge A, *Heraeus*)
- Récolter la phase aqueuse supérieure contenant l'ADN
- Ajouter 10 µL de tampon de charge
- Faire migrer les échantillons sur un gel d'agarose 1% et 5 µL de Smart Ladder

La deuxième étape de purification consiste à purifier l'ADN extrait sur agarose après migration en utilisant le kit NucleoSpin Extract II (#740 609.50 *Macherey-Nagel*).

- Découper la bande correspondant au produit PCR ou au plasmide à purifier.
- Peser la bande de gel.
- Ajouter 200 µL de tampon NT par 100 mg de gel.
- Faire fondre le gel d'agarose dans un bain préchauffé à 50°C durant 10 minutes.
- Fixer l'ADN sur la colonne Nucleospin en passant le gel liquide sur la colonne et centrifuger 1 minute à 11,000 g (Biofuge 13, *Heraeus*).

- Eliminer le liquide dans le fond de l'éppendorf
- Laver la membrane en ajoutant 600 µL de tampon NT3 et centrifuger durant 1 minute à 11,000 g (Biofuge 13, *Heraeus*)
- Eliminer le liquide
- Centrifuger à 11,000 g durant 2 minutes pour sécher la membrane
- Placer la colonne dans un nouvel eppendorf
- Eluer l'ADN en ajoutant entre 15 et 50 µL de tampon d'élution préchauffé (37°C) au bain marie en fonction de la concentration finale souhaitée
- Incuber 1 minute à température ambiante
- Centrifuger 1 minute à 11,000 g et récolter l'ADN purifié dans le fond de l'éppendorf.

2.9. Restriction ligation simultanée

La restriction ligation simultanée des produits PCR non digérés et du vecteur digéré est réalisée suivant le ratio 3 :1.

- Incuber le mix suivant dans un thermo cycler : 2 µL tampon T4 DNA ligase 10X (*NEB* #B0202); 1 µL NaCl 1M; 0,3 µL BSA 100X; 0,5 µL SpeI (*NEB* #R0133S); 0,5 µL SalI-HF (*NEB* #R3138S); 0,3 µL T4 DNA ligase *NEB* 2000 U/µL (*NEB* #M0202S); 150 ng de produit PCR non digéré; 50 ng de vecteur pBluescript II ks+ linéarisé par XhoI, XbaI et déphosphorylé; 11,4 µL H₂O pour un volume final de 20 µL.
- Utiliser le programme suivant : 2h30 à 37°C; 3h à 20°C; 16°C à l'infini

2.10. Screening

L'insertion du gène candidat (produit PCR) dans le vecteur disrups le gène de la β-galactosidase. Sur un milieu contenant du X-Gal (homologue chromogène du substrat de la β-galactosidase), les bactéries ayant l'insert apparaissent blanches en opposition aux bactéries contenant le vecteur sans insert qui sont bleues.

Un screening est effectué sur 64 clones transformés par un plasmide ayant le gène candidat d'intérêt. Le screening est réalisé à l'aide d'amorces marquées T3 et T7, amplifiant les régions bordant l'insert. L'utilisation d'amorces marquées D4 ou D3 (WellRed) est requise pour l'étape d'EcoTilling suivante.

- Composition du master mix pour 64 PCR de 20µL : 28 µl amorce forward T7 10 µM ; 28 µl amorce reverse T3 10 µM ; 14µl dNTP 10mM (*Fermentas*) ; 11,2 µl Taq Polymerase (*Promega*) ; 280 µl tampon Taq 5x (*Promega*) ; 1038,8 µl H₂O
- Piquer dans une colonie blanche avec un tips
- Repiquer sur une boîte de Pétri contenant du LB Agar + antibiotiques (Amp)
- Tremper le tips dans un puits de la plaque de screening
- Changer de tips et recommencer pour les autres clones
- Sur environ 80 clones étalés sur boîte, les 64 premiers sont amplifiés par PCR
- Lancer une PCR avec les paramètres suivants : 10' à 94°C ; (45'' à 92°C ; 45'' à 58°C; 45'' à 72°C) X35 ; 10' à 72°C ; 4°C à l'infini.

Le screening permet de vérifier la présence de l'insert dans le clone amplifié et l'obtention de produits PCR marqués utilisables pour l'analyse EcoTilling.

2.11. EcoTilling

Le protocole d'EcoTilling est divisé en trois parties : l'étape de dénaturation- hybridation, l'étape de digestion et une étape de purification de l'ADN.

2.11.1. Etape de dénaturation hybridation

- Mettre la même quantité (en ng) d'ADN à analyser que de références dans un volume final de 10 µL
- Chauffer à 95°C puis refroidir graduellement jusqu'à 4°C

2.11.2. Etape de digestion

- Prendre 4 µL de produit d'hybridation et monter à un volume de 10 µL avec de l'eau miliQ.
- La quantité d'ADN ne peut pas dépasser 7 µL
- Pour chaque tube ajouter 5 µL du mix suivant : 1.5 µl d'enzyme endo1 diluée 50x (à l'aide du tampon de dilution) ; 1.5 µl de tampon de réaction 10x ; 2 µl d'eau milliQ
- Incuber 2h à 37°C

2.11.3. Etape de précipitation de l'ADN

Stop solution : 2 µl d'acétate de sodium (3 M pH 5.2) ; 1.6 µl d'eau miliQ ; 1 µl de glycogen (20 mg/ml); 0.4 µl d'EDTA (0.5 M pH 8)

- Stopper la réaction en ajoutant 5 µL de Stop solution et 60 µL d'éthanol absolu stocké à -20°C.
- Centrifuger 1h à 3.500 rpm à 4°C
- Éliminer le surnageant
- Ajouter 200 µl d'éthanol 70% stocké à -20°C
- Centrifuger 10 min à 3500 rpm et éliminer le surnageant
- Répéter un second lavage à l'éthanol 70%
- Éliminer le surnageant
- Sécher l'ADN au Speedvac pendant 30 minutes
- Resuspendre dans 40 µl de Sample Loading Solution (*Beckman Coulter #608082*) contenant un standard de taille (Standard de taille 1000, dilution 170X)
- Pipeter 25 X pour resuspendre le culot
- Transférer le contenu de la plaque PCR dans une plaque « échantillons » prévue pour le séquenceur

2.12. Préparation séquençage

Le séquençage des gènes candidats peut se faire sur plasmide ou directement sur produit PCR.

2.12.1. Miniprep

Suite aux résultats EcoTilling, des clones sont sélectionnés pour être séquencés afin de valider les résultats EcoTilling. La réaction de séquençage peut être réalisée sur plasmides extraits suivant le protocole ci-dessous :

La composition des tampons P1, P2 et P3 nécessaires au protocole est décrite ci-dessous :

- Tampon P1 : 100 µg/ml Rnase A (Sigma); 50 mM Tris HCl pH 8 (Sigma); 10 mM EDTA (Sigma)
 - Tampon P2 : 200 mM NaOH; SDS 1%
 - Tampon P3 : 3 M KAc pH 5,5
-
- Dans un falcon de 15 mL, inoculer 4 mL de milieu LB contenant de l'ampicilline (60 µg/mL) avec un échantillon du clone à séquencer ayant été repiqué préalablement.
 - Incuber overnight à 37°C à 180 rpm (C24 Incubator Shaker, *New Brunswick Scientific*)
 - Prélever 150 µL de culture et ajouter 150 µL de glycérol 40% stérile. Les clones peuvent ainsi être stockés dans des cryotubes à -80°C (Stock glycerol).
 - Centrifuger les falcons de 15 mL durant 4 minutes à 4°C et 3,500 rpm (Allegra X-22R, *Beckman Coulter*)
 - Eliminer le surnageant et resuspendre dans 300 µL de tampon P1 stocké à 4°C.
 - Incuber 10 minutes à température ambiante
 - Ajouter 300 µL de tampon P2 stocké à température ambiante pour lyser les bactéries
 - Incuber 5 minutes à température ambiante
 - Précipiter les protéines en ajoutant 300 µL de tampon P3 stocké à 4°C
 - Incuber durant 30 minutes à 4°C
 - Centrifuger 10 minutes à 4°C et 13,000 rpm (Biofuge A, *Heraeus*)
 - Transférer le surnageant dans un nouveau tube et précipiter l'ADN en ajoutant 0,7 volume d'isopropanol stocké à -20°C
 - Homogénéiser par inversion et incuber durant 30 minutes à température ambiante
 - Centrifuger 15 minutes à température ambiante à 13,000 rpm (Biofuge 13, *Heraeus*)
 - Eliminer le surnageant
 - Laver l'ADN en ajoutant 500 µL d'éthanol 70%
 - Centrifuger à 4°C durant 5 minutes (Biofuge 13, *Heraeus*)
 - Eliminer le surnageant d'éthanol avec une pipette Pasteur
 - Sécher les échantillons 30 minutes au Speedvac
 - Resuspendre l'ADN dans 25 µL de Tris 10 mM pH 8 (*Sigma*)
 - Placer sur agitateur rotatif et laisser resuspendre à 4°C overnight.
 - Le dosage de l'ADN est réalisé par électrophorèse.

2.12.2. Agarose Low Melting

La miniprep est une méthode lourde lorsqu'il faut manipuler un grand nombre de clones simultanément. Pour pallier à ce problème, nous avons adapté le protocole de « Hao Ma and Stephen DiFazio, 2008 » permettant la purification de produit PCR après électrophorèse sur gel d'agarose. L'agarose low melting reste liquide à des températures inférieures à celle de l'agarose normal. Des mises au point ont montré qu'il était possible de prélever le produit PCR à la pipette dans un gel d'agarose low melting 0,2% (BP1360-100 *Fisher Scientific*) et de le séquencer directement (Ceq8000 *Beckman Coulter*). Le protocole a été mis au point pour être utilisé sans bromure d'éthidium. La durée totale du protocole est approximativement 3h.

- Un gel classique d'agarose 1% est coulé (1:75000 SyberSafe S33102 *Invitrogen* 10000X). Chaque puits peut être chargé avec 15 µL de produit PCR.

- Découper une fenêtre dans le gel d'agarose à l'aide d'un scalpel. La bande est découpée 3 centimètres sous les puits et la hauteur n'excède pas 2 cm.
- La fenêtre est remplie d'agarose low melting 0,2% à la pipette pasteur en veillant à ne pas boucher les puits.
- Laisser solidifier l'agarose low melting en refroidissant le gel pendant 1,5 h à 4°C
- Charger les puits et lancer l'électrophorèse du gel sur glace
- Prélever directement sous UV la bande (volume = 25 µL) avec les produits PCR lorsque ceux-ci se trouvent dans la fenêtre d'agarose low melting. Stocker les échantillons dans des tubes PCR
- Centrifuger les échantillons à 17,000 g et 16°C durant 5 minutes
- 17 µL de supernageant peuvent être prélevés par échantillon et directement utilisés pour la réaction de séquençage.

2.13. Séquençage

Le séquençage des gènes candidats suit le protocole du kit Ceq DTCS quick start, Beckman Coulter. Le protocole est divisé en trois étapes principales : la réaction de séquençage, le lavage et la précipitation de l'ADN et la préparation des échantillons pour le séquenceur.

2.13.1. Réaction de séquençage

- Préparer un mélange contenant 1 µl d'amorces (10 pmoles /µl) et 100 ng d'ADN. Ajuster le volume à 10 µl avec de l'eau milliQ.
- Dénaturer le mélange à 95 °C pendant 15 minutes. Mettre immédiatement sur glace pendant 3 minutes.
- Pendant ce temps, faire une solution contenant 32 µl de CEQ DTCS quik start kit et 50 µl d'eau milliQ (quantité pour 8 réactions de séquençage)
- Ajouter 9.5 µl de cette solution dans chaque tube
- Lancer une réaction de séquençage avec le programme suivant : 92°C pendant 1'30''; 40 cycles définis comme suit : 20'' à 96°C ; 20'' à 50°C ; 4' à 60°C
Ensuite, 4°C à l'infini.

2.13.2. Lavage et précipitation de l'ADN

La solution Stop utilisée dans le protocole ci-dessous a la composition suivante : 2 µl d'acétate de sodium (3 M pH 5.2) ; 1.6 µl d'eau milliQ ; 1 µl de glycogène (20 mg/ml); 0.4 µl d'EDTA (0.5 M pH8)

- Stopper la réaction, en ajoutant 5 µl de Stop solution + 60 µl d'éthanol absolu. Stocké à -20°C
- Centrifuger 1h à 3,500 rpm à 4°C
- Éliminer le surnageant
- Ajouter 200 µl d'éthanol 70% stocké à -20°C
- Centrifuger 10 min à 3,500 rpm et éliminer le surnageant
- Répéter un second lavage à l'éthanol 70%
- Éliminer le surnageant en secouant une seule fois de manière ferme la plaque multi-puits
- Éliminer l'éthanol restant en plaçant les échantillons 30 minutes au Speedvac

2.13.3. Préparation des échantillons pour le séquenceur

- Resuspendre dans 40 µl de SLS. Placer dans un puits d'une plaque de séquençage et couvrir avec une goutte d'huile minérale.

- Prendre une nouvelle plaque, et ajouter dans chaque puits 8 gouttes de SEQ sequencing separation buffer. Placer cette plaque et celle contenant les échantillons à l'endroit prévu dans l'appareil de séquençage.
- Lancer la méthode de séparation LFR-a (CEQ 8000 Genetic analysis system) dont les paramètres sont repris ci-dessous :
 - température des capillaires lors de la migration : 50°C
 - température de dénaturation : 90°C durée : 120 secondes
 - injection : voltage : 2 kV, durée : 15 secondes
 - séparation voltage : 4 kV, durée : 110 minutes
- Analyser les séquences avec les paramètres d'analyse par défaut.

Résultats et discussions

1 Recherche de gènes candidats : approche bioinformatique

La recherche de gènes candidats associés à la pathogénèse est divisée en quatre approches. Premièrement, la recherche d'orthologues entre la laitue et la chicorée sur base des banques d'unigènes disponibles pour les deux espèces. Deuxièmement, la recherche de gènes NBS-LRR, ceux-ci jouant un rôle majeur dans la détection des pathogènes et l'activation de réponses spécifiques à ces derniers. Troisièmement, la recherche de gènes clefs dans les voies de défense de la plante. Finalement, nous avons recherché des homologues des gènes WAK chez la chicorée. Ces derniers sont nécessaires aux voies de défense chez *A. thaliana*.

1.1 Recherche d'orthologues entre la laitue et la chicorée

La laitue est la plante modèle chez les *Asteraceae* et est phylogénétiquement très proche de la chicorée. Contrairement à la chicorée industrielle qui reste très peu étudiée, de nombreuses données sont disponibles pour la laitue. Plusieurs cartes génétiques de laitue sont publiées dans la banque de données du "*Compositae Genome Project Data Base (CGPDB)* ». Plusieurs séquences EST associées à la résistance de la laitue ont été cartographiées sur l'ensemble des 9 chromosomes de la laitue. Notre but est de trouver des orthologues de ces gènes chez la chicorée. Après une caractérisation du polymorphisme de ces orthologues, ceux-ci pourront être positionnés sur une carte génétique de chicorée et permettront de futures études comparatives entre le génome de la laitue et de la chicorée industrielle. La cartographie comparée permet de réaliser entre autres des études de synténie. Nous espérons pouvoir transposer des données intéressantes de la laitue vers la chicorée et vice versa.

Nous avons utilisé l'approche du *reciprocal best BLAST hit* pour mettre en évidence des orthologues de la laitue dans notre banque d'unigènes de chicorée. Cette méthode est basée sur l'algorithme BLAST et permet la détection d'orthologues tout en limitant les faux positifs. Cependant, la technique du *reciprocal best BLAST hit* montre certaines difficultés à discriminer les orthologues des paralogues. Cette méthode présente l'avantage principal d'être basée sur une approche relativement simple à appliquer au départ des analyses BLAST. Contrairement aux méthodes alternatives de mise en évidence d'orthologues, le *reciprocal best BLAST hit* ne nécessite pas le recours au séquençage complet du génome et peut être appliqué sur base de séquences EST.

142 séquences EST associées à la résistance aux pathogènes chez la laitue ont été positionnées sur une carte inter-spécifique (MAP2JMR3) obtenue par croisement entre *L. serriola* et *L. sativa*. La première étape consiste à rapatrier manuellement les séquences de *L. serriola* correspondant aux 142 EST positionnés sur la carte de la laitue. Les séquences proviennent d'un croisement inter spécifique et nous ne savons pas si les séquences de la carte proviennent de *L. serriola* ou de *L. sativa*. Les séquences associées à *L. serriola* sur la carte seront retrouvées en blastant les séquences EST de la carte contre une banque d'EST de *L. serriola*.

Une banque locale d'EST contenant 15,057 unigènes de *Lactuca serriola* provenant du CGPDB est construite dans SeqTools 8.4.

Les 142 séquences ont été blastées contre la banque de données de *L. serriola* en TBLASTX. Les séquences montrant un minimum de 95% d'identité sur une longueur minimale de 100 pb avec une *e-value* proche de 0 ont été sélectionnées.

Une banque de données locale contenant 23,135 unigènes de chicorée provenant de trois banques EST est créée. Ces trois banques proviennent du projet PHYTOMOL, de Lille et de Davis et constituent un total de 54,000 séquences. Celles-ci ont été converties en 23,135 unigènes en utilisant le logiciel EGAssembler.

Les 142 séquences de *L. serriola* sont blastées contre la banque d'unigènes de chicorée. 60 séquences présentent un résultat de BLAST significatif. Pour l'analyse *reciprocal best BLAST hit*, les 60 séquences de chicorée sont blastées à leur tour contre la banque de données de *L. serriola*. Vingt-et-une séquences présentant une homologie sur un minimum de 100 pb répondent aux critères du *reciprocal best BLAST hit*. Ces vingt-et-une séquences sont des orthologues potentiels entre la laitue et la chicorée et sont associées à la résistance de la plante face aux pathogènes.

Les vingt-et-une séquences sont localisées à travers l'ensemble des 9 chromosomes de la laitue (voir table 2.1). La figure 2.1 est une représentation graphique de la position des orthologues trouvés sur la carte de la laitue MAP2JMR3

1.2 Recherche de gènes candidats de type NBS-LRR et mise en évidence de gènes associés à la pathogenèse.

1.2.1 Gènes de résistance contenant un domaine NBS chez la chicorée

Les gènes NBS-LRR constituent la plus grande famille des gènes de résistance chez les végétaux. Ces gènes sont caractérisés par un domaine appelé *Nucleotide Binding Site*, NBS ou NB-ARC. La recherche de gènes contenant un domaine NBS a permis de mettre en évidence un grand nombre de gènes candidats pour la résistance de la plante aux pathogènes et des études comparatives entre les gènes NBS de différentes espèces sont régulièrement réalisées.

En utilisant des amorces dégénérées pour le domaine NBS, Plocick *et al* amplifia en 2003 vingt-et-un gènes contenant un domaine NBS. Ces vingt-et-une séquences constituent actuellement les seuls gènes de résistance chez la chicorée présents dans les banques de données publiques. A l'exception de neuf séquences trouvées avec des paires d'amorces ciblant les gènes CC-NBS-LRR, les 12 autres séquences ne sont a priori associées à aucun domaine N- terminal.

Les vingt-et-une séquences partielles de chicorée associées aux gènes NBS-LRR ont été rapatriées de la banque de données NCBI. Plusieurs de ces séquences présentaient une grande homologie entre elles. Pour diminuer le risque d'amplification non spécifique, nous avons choisi de garder seulement les séquences divergentes. Nous souhaitons éviter au maximum les familles multigéniques. L'alignement des séquences montre que sept séquences ne forment pas de contig.

La présence du site NBS fut confirmée par des analyses Pfam. Six séquences montrent la présence d'un domaine NB-ARC caractéristique des gènes NBS-LRR chez les plantes. Ces six séquences sont donc sélectionnées comme gènes candidats pour notre travail.

Aucun domaine NB-ARC n'est trouvé pour la séquence AY153812.1. Cette séquence est annotée dans NCBI comme un gène candidat dans la résistance aux pathogènes. L'alignement des sept séquences ne montre pas que AY153812.1 est fondamentalement divergente des autres séquences associées au domaine NB-ARC (voir table 2.2). Suivant son homologie avec des séquences NB-ARC, nous avons décidé de l'intégrer à notre liste de gènes candidats.

La séquence partielle AY153810.1 présente un score élevé d'homologie avec la séquence 36078385.VIS provenant de la banque de données du CGPDB. Ces deux séquences sont fortement homologues (e-value = e-148; identité = 284/291 (97%)) et pourraient en fait correspondre à une seule et même séquence. La séquence 36078385.VIS est longue de 740 paires de bases (291pb pour AY153810.1) et montre un domaine NBS et TIR après recherche de domaines avec Pfam. Cette séquence est donc sélectionnée dans notre liste de gènes candidats à la place de la séquence AY153810.1.

1.2.2 Recherche de gènes NBS-LRR au départ de la banque d'*A. thaliana*

La plante modèle *A. thaliana* est totalement séquencée et un grand nombre de banques de données contenant des séquences annotées sont disponibles. Nous avons rapatrié toutes les séquences associées aux gènes NBS-LRR de la banque de données TAIR en évitant le plus possible de sélectionner des séquences annotées comme pseudogènes. 173 séquences furent ainsi sélectionnées.

Toutes les séquences d'*Arabidopsis* ont été blastées (TBLASTX) contre la banque locale de chicorée contenant 54,000 EST. L'algorithme TBLASTX compare des séquences nucléotidiques traduites en protéines contre une banque de séquences de nucléotides elle aussi traduite. Ce type de BLAST est la méthode la plus appropriée aux recherches d'homologies entre deux espèces différentes au départ de séquences nucléotidiques.

27 séquences de chicorée apparaissent de façon récurrente dans les résultats de BLAST pour les 173 séquences analysées. Pour ces 27 séquences, une recherche de domaines associés aux gènes NBS-LRR a été réalisée sur base d'analyse Pfam. La présence de motifs NBS-LRR a été confirmée dans 19 des 27 séquences de chicorée.

Comme illustré dans le tableau 2.3, les 19 séquences EST ont une taille comprise entre 503 et 867 pb. Sachant que la taille complète d'un gène TIR-NBS-LRR est proche de 4000 pb (exemple 4140 pb pour at5g41750 ADNc) les séquences EST trouvées ne peuvent contenir que des parties des gènes NBS-LRR. Les séquences contenant 2 domaines (TIR-NBS et NBS-LRR) présentent une haute probabilité d'appartenir aux gènes de résistance NBS-LRR. Les séquences EST contenant uniquement un domaine NB-ARC sont également de bons candidats. Les séquences présentant seulement des domaines TIR ou LRR ont une probabilité plus faible d'appartenir à la famille des gènes NBS-LRR.

Malgré le fait qu'elles puissent contenir des domaines identiques, aucune séquence ne présente un taux d'homologie élevé avec d'autres à l'exception de la séquence 31299385.VIS

présentant une homologie avec la séquence 22885385.VIS ainsi que la séquence 2880385.VIS présentant des homologies avec la séquence 30861385.VIS. Le taux de mutations dans les régions homologues est de 1 mutation toute les 12pb. Il s'agit vraisemblablement de paralogues. En effet, la banque de données Davis ne contient que des séquences d'un unique individu d'où proviennent les séquences ci-dessus. Les études de caractérisation du polymorphisme pour ces séquences pourront peut-être nous aider à confirmer s'il s'agit du polymorphisme d'une seule séquence ou s'il s'agit bien de paralogues.

1.2.3 Recherche de gènes candidats à partir de séquences de tournesol associées aux gènes NBS-LRR

La laitue, la chicorée et le tournesol sont des membres de la famille des *Asteraceae*. Les espèces phylogénétiquement proches présentent une homologie plus importante au niveau de leurs gènes. La banque de données NCBI contient 426 séquences d'ADN associées aux gènes de résistance NBS-LRR pour le tournesol, *Helianthus annuus*. Certaines d'entre elles sont décrites comme des pseudogènes.

Nous avons décidé d'utiliser toutes ces séquences pour rechercher des gènes présentant des homologies avec les gènes NBS-LRR chez la chicorée. Nous supposons qu'un pseudogène chez le tournesol ne l'est pas forcément chez la chicorée. Des pseudogènes récents chez le tournesol peuvent présenter une homologie avec la structure NBS-LRR. Un pseudogène peut ainsi être utilisé pour mettre en évidence des gènes fonctionnels dans notre banque de chicorée. Les gènes NBS-LRR sont annotés comme pseudogènes lorsqu'ils présentent des codons stop ou un décalage du cadre de lecture.

Une recherche TBLASTX entre les 426 séquences de tournesol et la banque d'unigènes de chicorée a été réalisée. Pour chaque séquence de tournesol testée, les séquences de chicorée avec la meilleure *e-value*, cut off de minimum 10^{-5} , ont été sélectionnées. Nous observons une redondance dans nos résultats de BLAST. 97 séquences de chicorée associées aux gènes NBS-LRR ont pu être détectées.

Les séquences de chicorée ont été analysées avec le logiciel Interproscan. Seize séquences ont été associées avec des domaines de résistance aux pathogènes. Huit séquences montrent un domaine NB-ARC.

Pour améliorer la probabilité d'être bien en présence d'un gène de résistance, ces séquences ont été blastées (TBLASTX) contre la banque de NCBI. Seulement 3 séquences ne présentent pas d'homologies avec des gènes NBS-LRR : CIC.EH690542 ; CONTIG4309 ; CONTIG6761. Cependant, les résultats de BLAST indiquent que ces séquences présentent des homologies avec des protéines de résistance.

L'analyse Interproscan apparaît comme une méthode efficace pour mettre en évidence des gènes associés à la pathogenèse. Des séquences associées à la pathogenèse ne présentant pas de domaine NBS lors de l'analyse Interproscan peuvent lors d'analyses BLAST présenter une homologie élevée avec des gènes NBS-LRR. Cette observation confirme l'importance d'utiliser des outils variés dans le cadre de l'annotation de séquences.

Les séquences trouvées ont été comparées avec la liste de gènes candidats déjà constituée pour éviter la redondance dans notre liste. Les résultats d'alignements de séquences révèlent six nouvelles séquences : EH690542, CIC.EH693729, CIC.EH678163, CONTIG4952, CONTIG6761, CIC.EH672777. Elles sont ajoutées dans la liste de gènes candidats sous le nom de référence CHICOSUNX soit respectivement de CHICOSUN1 à CHICOSUN6. Quatre des six séquences présentent une homologie élevée avec des gènes NBS-LRR dans la banque de données NCBI (TBLASTX, paramètre par défaut) voir table 2.4.

Suite à la mise en évidence de ces nouvelles séquences, nous avons pu observer, dans l'ensemble des séquences trouvées, que 4 séquences (CONTIG1413 ; 26173385.VIS ; CIC.EH694833 ; 1586035.VIS) pouvaient être alignées pour former un contig de 1786 pb. Cette grande séquence contient un domaine NB-ARC suivi d'un domaine LRR (voir figure 2.2). Les 4 séquences assemblées pourraient provenir d'un même gène. La création d'amorces de part et d'autre de la séquence pourrait montrer s'il s'agit bien d'une seule et même séquence.

1.2.4 Data mining : recherche de gènes associés à la pathogenèse dans la banque d'unigènes de chicorée à l'aide d'Interproscan

Comme détaillé dans l'introduction, une grande partie des protéines impliquées dans la reconnaissance des pathogènes sont les protéines NBS-LRR. Le domaine NB-ARC est un marqueur de la présence d'un gène NBS-LRR. Ce domaine peut être mis en évidence par des logiciels de prédiction de domaines protéiques comme le logiciel Pfam (Sonnhammer, Eddy et al. 1997; Finn, Tate et al. 2008). La banque d'unigènes de chicorée est susceptible de contenir une collection de gènes associés à la résistance aux pathogènes incluant les gènes NBS-LRR. En analysant la banque de données par un logiciel de prédiction de domaines, il devrait être possible de mettre en évidence des gènes de résistance chez la chicorée.

De nombreuses méthodes d'analyse et de prédiction de domaines sont disponibles. Il existe un grand nombre de banques de domaines : Pfam, Prodom, TIGRsfam, ... Chacune de ces méthodes et bases de données utilisent des paramètres d'entrées et de sorties différents et présentent chacune de plus ou moins grandes capacités à discriminer certains domaines. Interproscan permet de tester en une étape les banques de données suivantes : PROSITE, PRINTS, Pfam, ProDom, SMART, TIGRFAMs, PIR superfamily, SUPERFAMILY, Gene3D et PANTHER. En une unique opération, il est donc possible d'analyser une séquence d'acide nucléique par un grand nombre de méthodes différentes et par rapport à différentes bases de données. En plus des résultats obtenus par chaque méthode, Interproscan est capable de standardiser les résultats, de visualiser les domaines de manière graphique et de les exporter en fichier texte (Quevillon, Silventoinen et al. 2005).

Interproscan intègre un outil qui traduit la séquence nucléotidique dans les 6 cadres de lecture. Chacune de ces séquences est ensuite analysée suivant les paramètres choisis. Interproscan est capable d'analyser simultanément 50 séquences nucléotidiques. Une version locale d'Interproscan fonctionnant sous système Perl/Unix permet l'analyse simultanée de 100,000 séquences nucléotidiques. Nous n'avons pas réussi à utiliser ce logiciel sous une machine tournant sous Ubuntu 8.04.1. En attendant de pouvoir analyser les séquences en local, les 5,000 premières séquences de la banque d'unigènes de chicorées ont été analysées manuellement en utilisant l'interface web d'Interproscan (autorisant l'analyse simultanée de 50 séquences). Nous avons utilisé la version d'Interproscan 18.0 reprenant 16,549 domaines

standardisés (dont le domaine NB-ARC). Actuellement, 27.432 publications dans PUBMED font référence à Interproscan.

En analysant la banque d'unigènes de chicorée avec Interproscan, nous avons pour objectif de construire une banque de données annotées. Celle-ci devrait permettre la mise en évidence rapide de protéines associées à la pathogenèse chez la chicorée. Cette banque de données permettra la recherche d'EST et de leur description sur base de mots-clefs et permettra la mise en évidence de séquences dans des objectifs de recherche variés (métabolisme, floraison, récepteurs, ...).

Les 5.000 premières séquences de la banque d'unigènes de la chicorée ont été analysées avec succès. 3.361 ORFs ont ainsi été annotées. Une ORF peut contenir plusieurs descriptions : des domaines de fonctions différentes ou le même domaine trouvé dans plusieurs banques de données différentes. L'analyse complète des 5,000 séquences a donné 12,792 descriptions de domaines. La taille moyenne des ORF annotées est de 166 acides aminés. (table 2.5)

La mise en évidence et l'annotation des domaines s'est faite sur 13 banques de données différentes : HMMPanther, HMMPfam, superfamily, Gene3D, FPrintScan, ProfileScan, Seg, HMMSmart, ScanRegExp, BlastProDom, Coil, HMMTigr and HMMPPIR. Un nombre variable de domaines a été trouvé par l'analyse de ces banques de données. Le plus grand nombre de descriptions a été apporté par les analyses HMMPanther, et les analyse HMMPPIR ont donné le moins de résultats. La variation du nombre d'annotations obtenues devrait être normalisée en fonction du nombre de domaines disponibles dans chaque banque de données. (Table 2.5)

6,549 domaines ont été annotés par Interproscan. Le domaine Inteproscan le plus fréquent est "Calcium-binding EF-hand" qui apparaît 236 fois. Toutes annotations confondues, 406 séquences ont été associées à un domaine protéine kinase. 121 ORFs ont été associées avec des domaines LRR. 55 séquences sont associées avec un domaine "Pathogenesis-related transcriptional factor and ERF, DNA-binding". Les domaines interproscans les plus fréquents sont repris dans la table 2.6.

102 descriptions associées à la résistance ont été associées à 37 séquences EST. 31 séquences étaient annotées par des domaines caractéristiques de protéines impliquées dans la défense aux pathogènes. Pour augmenter la probabilité que ces gènes soient bien effectivement impliqués dans la défense de la plante, les séquences ont été blastées (TBLASTX) contre la banque de données NCBI. Nous n'avons conservé que les séquences dont le résultat de BLAST était directement associé à la pathogenèse ; parmi ces séquences, 3 séquences associées aux gènes NBS-LRR avaient déjà été trouvées précédemment.

Finalement, 18 séquences associées à la pathogenèse (suivant les résultats Interproscan et BLAST NCBI) ne se trouvant pas dans nos listes ont pu être mises en évidence et sont ajoutées à notre liste de gènes candidats. Les nouvelles séquences sont référencées sous le nom de code « CHICOSCAN » dans notre set de gènes candidats. Table 2.7

CHICOSCAN13 est un gène associé à la famille NBS-LRR. CHICOSCAN 1 et 15 sont associés aux facteurs de transcription WRKY. CHICOSCAN 3, 4, 9 et 16 sont homologues à des protéines impliquées dans la résistance au mildiou. CHICOSCAN1 montre une grande

homologie avec PR1, *pathogenesis related protein class1*. Cette protéine est activée dans les réponses de défense et notamment en réponse à l'activation de la voie de l'acide salicylique. Le criblage de la banque de données de chicorée par Interproscan est une méthode permettant la mise en évidence de séquences impliquées dans la réponse aux pathogènes et permet entre autres la détection des gènes NBS-LRR. (voir table 2.7)

1.2.5 Data mining : mise en évidence de gènes candidats à partir de la banque CGPDB

Comme nous l'avons mis en évidence dans le point précédent, l'annotation complète de banques de données peut s'avérer utile pour la mise en évidence de gènes associés à une fonction donnée. Le CGPDB a rendu public une banque d'unigènes de chicorée industrielle. Celle-ci contenant 22,038 unigènes est construite sur base de 41,704 séquences EST. Une version annotée de la banque est disponible.

L'annotation de chaque séquence consiste en la description du meilleur résultat de BLAST (BLASTX) contre la banque de donnée d'*A. thaliana*. 75,8% des séquences unigènes de chicorée ont un résultat de BLAST.

Nous sommes partis de cette base de données annotée pour mettre en évidence de nouveaux gènes candidats associés à la réponse aux pathogènes chez la chicorée industrielle.

Sur base de mots-clefs, 78 séquences dont le résultat de BLAST était associé à la pathogénèse ont été rapatriées. De nombreuses séquences présentent une homologie avec des gènes NBS-LRR chez *A. thaliana*.

Pour diminuer le risque d'amplification non spécifique, les gènes paralogues, présentant un taux trop important d'homologie entre eux, ont été supprimés. Nous avons également éliminé les séquences identiques ou très proches de celles précédemment trouvées.

44 séquences ont ainsi été sélectionnées. Nous avons analysé ces séquences avec Interproscan. Seulement trois séquences présentaient un domaine associé aux gènes de résistance (CI.EH684821, CI.EH672557, CI.CSA1.5614). CI.EH672557 est une séquence présentant une forte homologie avec les gènes NBS-LRR.

Les 44 séquences de chicorée ont été ensuite blastées (TBLASTX) contre la banque de données NCBI pour des recherches supplémentaires d'homologies de séquences. 37 séquences montrent une homologie avec des gènes de résistance (voir score de BLAST et e-value Table 2.8).

15 séquences présentent une forte homologie avec des gènes de résistance avec des résultats de BLAST ayant une e-value comprise entre 4^E-31 et 4^E-105 et des scores compris entre 100 et 232.

1.3 Recherche de gènes clefs dans les voies de défense de la plante : acide salicylique, jasmonate et éthylène

Nous avons sélectionné une liste de gènes ayant un rôle clef dans la réponse aux pathogènes chez *A. thaliana*. Les séquences ADNc des gènes suivants ont été rapatriées de la banque de données d'*A. thaliana* : NPR1 (AT1G64280), PDF1.2(AT5G44420) ; THI2.1(AT1G72260),

JAR1(AT2G46370), WRKY70 (AT3G56400), MAPK4 (AT4G01370), COI1(AT2G39940), ERF1(AT3G23240), JIN1(AT1G32640).

Les neuf séquences ont été blastées (TBLASTX) contre la banque d'unigènes de chicorée à l'aide du logiciel SeqTools 8.4 (paramètres par défaut).

Des résultats de BLAST significatifs (cut off : $e\text{-value} > E\text{-}50$) ont été obtenus pour NPR1, JAR1, MAPK4, COI1 et JIN1 (voir table 2.9). Pour confirmer la mise en évidence de gènes homologues, les cinq séquences de chicorée ont été blastées contre la banque de données NCBI (voir table 2.10).

CONTIG5419 présente une homologie presque parfaite avec la séquence MAPK4 d'*A. thaliana* et présente une homologie encore meilleure avec la MAPK4 de *N. tabacum*. Suivant les résultats de Blast, voir table 2.9, nous pouvons dire que nous avons trouvé des homologues sérieux des gènes NPR1, JAR1, MAPK4, COI1 et JIN1 chez la chicorée.

1.4 Recherche de gènes WAK (Wall Associated Kinase) chez la chicorée

1.4.1 Recherche de gènes WAK sur base de séquences *A. thaliana*

Il a été montré chez *A. thaliana* que les récepteurs de la famille WAK (sérine-thréonine kinase) pouvaient fixer la pectine. Ces récepteurs sont suspectés de jouer un rôle dans le système de défense de la plante. L'inhibition de WAK1 chez *A. thaliana* donne naissance à une plante naine, affaiblie et très sensible aux stress. Nous avons pour objectif de trouver des gènes homologues des WAK chez la chicorée industrielle. Une caractérisation du polymorphisme des WAK chez la chicorée pourra être utilisée lors d'études d'associations avec l'espoir de pouvoir augmenter nos connaissances sur cette famille de récepteurs.

Les séquences codantes des 5 WAK d'*A. thaliana* ont été rapatriées de la banque de données de NCBI. Les 5 séquences ont été blastées contre la banque d'unigènes de chicorée. Pour les 5 analyses BLAST, la séquence de chicorée présentant la plus grande homologie avec les récepteurs WAK était la séquence CONTIG5881. L' $e\text{-value}$ est comprise entre $1^E\text{-}113$ et $1^E\text{-}116$ et les scores de BLAST sont compris entre 259 et 273. Cette séquence a été sélectionnée comme un homologue de WAK.

Les autres séquences de chicorée présentant une homologie avec les WAK d'*A. thaliana* avec une $e\text{-value}$ de minimum $1^E\text{-}49$ ont été analysées dans Interproscan. La séquence CONTIG7645 est la seule séquence présentant un domaine WAK (figure 2.3). Cette séquence a été blastée contre la banque d'unigènes de chicorée. Parmi les résultats de BLAST, la séquence CIC.EH691923 est caractérisée par un domaine WAK dans Interproscan. La séquence CONTIG4966 est également sélectionnée car elle présente des structures homologues aux deux autres séquences.

Les recherches d'homologies entre les séquences WAK d'*A. thaliana* et la banque d'unigènes de chicorée ainsi que les analyses Interproscan ont permis de mettre en évidence 4 séquences de chicorée homologues des WAK. Ces 4 séquences ont été ajoutées à la liste des gènes candidats sous la référence CWR (*chicory wall associated kinase related*).

1.4.2 Recherche de WAK par data mining

La banque de données précédemment annotée pour les 5,000 premiers EST unigènes de chicorée a été utilisée pour mettre en évidence des domaines WAK. La séquence CI.EH691923, trouvée précédemment a pu être mise en évidence. L'annotation des 15,000 séquences EST restant devrait permettre la mise en évidence d'autres séquences présentant un domaine associé aux WAK.

Le criblage de la banque de données annotées de Davis par les résultats de BLAST contre *A. thaliana* a permis la mise en évidence de 4 séquences. L'analyse Interproscan de ces 4 séquences ne montre pas de domaine WAK bien qu'un domaine serine-thréonine kinase soit présent dans chacune des 4 séquences.

Les 4 séquences ont été blastées (TBLASTX, paramètres par défaut) contre la banque de données de NCBI. La séquence CI.EH691833 montre des résultats de BLAST particulièrement intéressants. Cette séquence présente une homologie avec WAK1 de *T. aestivum* avec un score de BLAST de 320 et une *e-value* de 10^{-84} (résultats table 2.11). Cette séquence est ajoutée à la liste des gènes candidats comme un homologue de WAK.

1.5 Synthèse de la recherche de gènes candidats impliqués dans les réponses de défenses.

Un gène candidat est un gène supposé être impliqué dans une fonction d'intérêt. Pour améliorer la résistance de la chicorée aux pathogènes, nous avons décidé d'analyser le polymorphisme de gènes candidats associés à la réponse aux pathogènes (Récepteurs, protéines clefs dans les voies de réponses, ...). La première étape de ce travail a été de dresser une liste de gènes candidats sur base des séquences EST de chicorée disponibles.

Notre recherche s'est principalement focalisée sur la recherche de gènes NBS-LRR. Sur base de recherches BLAST, 27 séquences de chicorée ont été associées aux gènes NBS-LRR. Cinq séquences de gènes homologues aux WAK ont également été trouvées. Une analyse *reciprocal best BLAST hit* a permis la mise en évidence de 21 gènes impliqués dans la résistance aux pathogènes potentiellement orthologues entre la laitue et la chicorée. 44 autres séquences ont été mises en évidence comme étant associées aux réponses de défense, comme par exemple des facteurs de transcriptions WRKY ou des gènes clefs des voies de défense chez la plante. 76 paires d'amorces ont été définies et 27 paires d'amorces ont déjà été testées positivement (après PCR, observation d'une seule bande sur gel de taille compatible avec les analyses EcoTILLING ainsi qu'avec le séquençage soit max 900 pb). (Table 2.12)

2 Caractérisation du polymorphisme chez *C. intybus*.

L'objectif de ce mémoire est de caractériser le polymorphisme de gènes candidats précédemment sélectionnés. Les applications de la mise en évidence du polymorphisme sont nombreuses. Des plantes peuvent être typées sur base de marqueurs SNP. A long terme, nous espérons pouvoir associer la présence de certains marqueurs avec un phénotype de sensibilité ou de résistance à un pathogène. Les marqueurs peuvent être bien sûr aussi positionnés sur une carte génétique de la chicorée.

La chicorée industrielle a été constituée en partie sur base de 29 cultivars de chicorée à café. Aujourd'hui, du polymorphisme présent à l'origine de la chicorée a sûrement disparu sous la

pression de sélection appliquée par les sélectionneurs de chicorée à inuline suivant des critères définis par ces derniers. Nous postulons que du polymorphisme associé à la défense des plantes a pu être perdu durant la sélection de la chicorée industrielle. Notre objectif est de retrouver le polymorphisme présent à l'origine de la sélection de la chicorée industrielle, de voir le polymorphisme qui a été sélectionné et celui qui a été perdu. De plus, à l'aide d'études d'association, nous espérons trouver des marqueurs moléculaires associés à la résistance de la chicorée. Ces marqueurs de résistance pourraient être ensuite réintroduits dans les plants actuels. Dans le cadre de ce mémoire, le polymorphisme ancien a été mis en évidence dans différents gènes candidats. Une comparaison entre le polymorphisme ancien et actuel a également été réalisée pour un gène.

2.1 Constitution de la « Core Collection »

Un pool d'ADN représentatif de la diversité présente à l'origine de la sélection de la chicorée industrielle a été constitué. Ce pool est formé de 64 extractions d'ADN réalisées avec le kit Nucleon phytopure au laboratoire. Une purification supplémentaire de ces ADNg a été réalisée avec le système Nucleospin Extract II (*Macherey-Nagel* #740 609.50). Les ADN de 21 autres échantillons dont l'extraction avait échoué ont été fournis par SA Warcoing.

Un superpool d'ADN a été constitué à partir de l'ADN extrait de ces 85 échantillons. Chaque échantillon y est représenté en concentration équivalente. La concentration finale du superpool primaire obtenu était de 5 ng d'ADN/ μ L. Des tests d'amplification PCR (50 μ L mix PCR et 1 μ L ADNg superpool) ont montré que l'amplification était réalisable jusqu'à une dilution 20X du superpool.

2.2 Etape de clonage des gènes candidats

Le polymorphisme de la *core collection* a été mis en évidence par la méthode de l'EcoTILLING. Pour permettre l'analyse automatisée sur séquenceur, les produits d'amplification doivent être marqués à leurs extrémités par un fluorochrome. Afin d'accélérer la procédure et de limiter le coût qu'engendrerait l'achat d'amorces spécifiques marquées pour chaque gène candidat, nous passerons par une étape de clonage. Tout gène candidat inséré de manière orientée dans le vecteur pBluescript KS +II peut ainsi être amplifié par la paire d'amorces universelles T3-T7 marquée.

Chaque gène candidat est amplifié par PCR sur l'ADN de la core collection avec des amorces spécifiques non marquées contenant à leurs extrémités des sites de restriction SalI HF et SpeI.

Les séquences amplifiées sont insérées de manière orientée dans le plasmide pBluescript II KS+ préalablement restreint par les enzymes XhoI et XbaI dans une étape de clonage par restriction-ligation simultanée. Après transformation de bactéries TOP10F' par choc thermique, l'étalement sur un milieu contenant de la tétracycline permet de sélectionner les bactéries transformées par le vecteur. Un test blanc/bleu permet la discrimination des clones ayant le vecteur avec insert des clones ayant le vecteur sans insert. Une PCR sur colonies avec les amorces marquées T3-T7 permet de confirmer la présence du gène candidat dans le vecteur et de préparer le matériel nécessaire pour l'étape d'EcoTILLING. Après mise au point du protocole, le taux de clones ayant été transformés par le plasmide et contenant l'insert est proche de 100%.

Au terme de ce mémoire, le clonage et le criblage des 15 gènes candidats suivant ont déjà été réalisés : CS58 ; CS60 ; CS62 ; CS76 ; CS78 ; CS90 ; CS92 ; CS94 ; CS112 ; CS126 ; CS160 ; CS230 ; CS242 ; CS284 ; CS286.

2.3 EcoTILLING

La technique de l'EcoTILLING permet la découverte rapide de SNP. Tout au long de ce mémoire, nous avons cherché à optimiser la mise en évidence de polymorphisme de type SNP dans une approche haut débit.

2.3.1 Approche en POOL

L'EcoTILLING permet la détection de SNP sur base de la comparaison entre deux séquences. L'analyse classique consiste à analyser le polymorphisme de différents individus contre une séquence de référence. Afin de réduire le nombre de manipulations et le coût d'analyse EcoTILLING sur sépareur de taille, nous avons développé une approche en pool.

Pour chaque gène candidat analysé, nous disposons de 64 clones différents. Les produits PCR marqués peuvent être organisés sur huit lignes et huit colonnes. Les séquences à analyser sont poolées par huit en ligne et en colonne. En croisant les résultats des lignes et des colonnes, nous espérons pouvoir déterminer le polymorphisme présent dans chaque clone, l'analyse en pool nous permettant de réduire considérablement le nombre d'analyses (16 en pool contre 64 en individuel).

2.3.2 Résultats des analyses EcoTILLING sur pool

Le polymorphisme de sept gènes candidats a été analysé sur POOL. L'ensemble des résultats est résumé dans le tableau 2.16.

La taille moyenne des séquences analysées est de 637 pb. La taille des séquences comprend la séquence du gène candidat amplifié bordée de part et d'autre d'une partie du vecteur contenant les sites T3 et T7 universels. CS76 est la séquence la plus grande analysée (965pb) et CS58 est la plus courte (467 pb). La fréquence moyenne de SNP sur l'ensemble des séquences analysées en pool est de un SNP toutes les 41 pb. La séquence CS230 présente un taux particulièrement élevé de polymorphisme avec un SNP prédit toutes les 18 pb. La séquence CS60 a une fréquence d'1SNP toutes les 57 pb pour une taille de 962 pb.

Sur base des résultats d'EcoTILLING obtenus durant ce mémoire, nous observons que le polymorphisme de type SNP entre les individus de la core collection est très important. La fréquence élevée des SNP s'explique par la diversité des individus utilisés pour la sélection de la chicorée industrielle. Cependant, cette seule explication ne justifie pas à nos yeux la présence de telles fréquences de SNP.

Les sorties EcoTILLING d'analyses sur pool ont, à plusieurs reprises, présentées un bruit de fond élevé rendant difficile la discrimination entre les signaux de SNP réels et le bruit de fond. L'analyse en pool sur la core collection est affectée par deux phénomènes : l'amplification soit non spécifique de gènes dont des familles multigéniques, et la présence de faux positifs.

Malgré l'attention portée à la sélection des gènes candidats et à la conception d'amorces spécifiques pour amplifier ces derniers, il est possible d'amplifier des gènes de familles

multigéniques ou de paralogues. Ces séquences sont homologues mais présentent plus de divergences que celles générées par le polymorphisme d'un même gène entre différents individus. Comme les séquences paralogues sont homologues, elles peuvent dans certaines situations être amplifiées par une même paire d'amorces. La présence d'une seule séquence paralogue dans un pool de huit suffit à générer un bruit important masquant le polymorphisme de la séquence d'intérêt.

Nous souhaitons extrapoler les résultats de polymorphismes pour chaque individu en croisant les résultats obtenus par lignes avec ceux obtenus par colonnes. La méthode donne potentiellement des faux positifs dans la mise en évidence de SNP fréquents dans les clones analysés. Pour illustrer le problème des faux positifs lors de l'analyse en pool, nous présentons ci-dessous une analyse fictive de SNP en pool de quatre. (illustration figure 2.4)

Seize clones positionnés sur une grille de quatre lignes et quatre colonnes sont analysés en pool par EcoTILLING (voir figure 2.4). Les clones 2, 4, 9 et 13 (fond vert) contiennent un SNP par rapport aux autres séquences. Si on poole les clones par ligne et par colonne, l'analyse EcoTILLING révèle : la présence d'un SNP dans les lignes A, B et D ainsi que la présence d'un SNP dans les colonnes a, c et d. En croisant les résultats EcoTILLING obtenus, on trouve que les clones 1, 2, 4, 9 et 13 ont un SNP. Le clone 1 apparaît être un faux positif par l'analyse en pool.

En résumé, l'analyse EcoTILLING en pool permet de montrer que la core collection contient beaucoup de polymorphisme. L'analyse en pool est sensible à la génération de bruit causé par l'amplification de paralogues et de familles multigéniques. L'analyse en pool donne des faux positifs pour les SNP fréquents mais permet la sélection de clones contenant des SNP rares.

2.3.3 Comparaison analyse individuelle et en pool : CS230 gène NBS-LRR

Suite aux problèmes rencontrés lors de l'analyse en pool, nous avons décidé de comparer les résultats obtenus sur pool et avec référence. La séquence CS230 a été analysée en EcoTILLING par pool de huit. Parmi les 64 clones analysés, il y avait 30 SNP détectés pour une fréquence d'un SNP toutes les 18 pb. Sur base des résultats EcoTILLING, les clones 8, 19, 27, 30, 49, 56 et 62 ont été sélectionnés pour être séquencés sur base de leur représentation du polymorphisme dans la core collection (Table 2.17).

Le clone 1 a été choisi arbitrairement comme référence dans le cadre des analyses un à un pour la séquence CS230. Six positions de SNP potentiels sont mises en évidence par cette analyse. Les 2 SNP les plus fréquents sont en position 100 pb et 189 pb (à partir de Table 2.18). Les résultats sont beaucoup plus nets et contrastés que lors de l'analyse en pool.

Sur base des résultats EcoTILLING avec référence, nous avons séquencé les clones 1, 6, et 42. Après édition des séquences, nous avons pu mettre en évidence sept SNP. La position de chaque SNP est reprise dans le tableau 2.19.

Trois SNP sont très proches les uns des autres en position 46, 48 et 51 pb. Leur proximité les rend difficiles à discriminer dans les résultats obtenus. Deux SNP sur trois avaient été trouvés par chacune des deux analyses EcoTILLING. Le SNP en position 79 n'a été mis en évidence que dans l'analyse en pool. Les SNP en position 121 et 140 ont été détectés dans les deux analyses. Le SNP en position 344 n'a été détecté par aucune des deux analyses.

La méthode d'EcoTILLING en pool a permis de mettre en évidence la majorité des SNP présents dans les clones séquencés. Cependant, les résultats apparaissent noyés dans le bruit de fond (cinq SNP réels sur trente SNP mis en évidence). L'analyse EcoTILLING avec référence n'a pas su mettre en évidence deux SNP qui ont été révélés par le séquençage de clones. Le bruit de fond est cependant beaucoup plus faible et le taux de faux positifs est de 33% (deux SNP prédits sur six se sont révélés être des faux positifs). Voir table 2.20.

Sur base de ces résultats, les analyses EcoTILLING suivantes ont été réalisées avec références. Cette approche permet une estimation plus fiable de la fréquence des allèles dans la population étudiée et est moins sujette au problème de bruit.

2.3.4 Analyse EcoTILLING avec référence

Les analyses sur pool ont mis en évidence la grande quantité de polymorphisme présent dans la core collection. L'analyse avec référence a pour objectif, outre de mettre en évidence de nouveaux SNP, de pouvoir estimer la fréquence de ces derniers dans la population. Nous testerons plus tard dans ce travail si le polymorphisme présent fréquent ou rare dans la core collection se retrouve dans les populations actuelles de chicorée.

L'analyse EcoTILLING avec référence a été réalisée sur trois autres gènes candidats : CS242, CS284 et CS286. (table 2.16)

2.4 Résultats de séquençage

Sur base des résultats EcoTILLING, nous avons sélectionné des clones à séquencer pour chaque gène candidat. Nous détaillerons ci-dessous les résultats obtenus pour chacun de ces derniers.

2.4.1 CS58, β 1,3 glucanase

Le gène CS58 est un orthologue potentiel entre la laitue et la chicorée pour un gène codant pour une β -1,3 glucanase. Les β -1,3 glucanases forment une famille de gènes étudiée dans de nombreuses plantes. Ce sont des protéines PR et elles jouent donc un rôle de défense contre les pathogènes. Elles sont impliquées dans les réarrangements de la paroi végétale et dans les processus de développement (Helleboid, Hendriks et al. 2000).

La taille totale de la séquence amplifiée et séquencée est de 358 pb. Des séquences exploitables ont été obtenues pour cinq clones différents et ont permis la mise en évidence de 18 nucléotides différents entre les cinq clones. On observe en moyenne une mutation toutes les 20 pb. Ce taux anormalement élevé de polymorphisme ne peut s'expliquer uniquement par la présence d'allèles différents. Le tableau 2.21 note la position des mutations dans les cinq clones. Ceux-ci peuvent être classés dans trois groupes différents. Il est probable que nous soyons en présence de 3 paralogues.

La séquence consensus (séquence obtenue sur base des cinq clones séquencés) de CS58 a été comparée à la séquence codante qui avait servi au design des amorces. Sur les 358 pb séquencée sur ADNg, il y a 87,5% d'identité avec la séquence EST ayant servi au design des amorces. Aucun intron n'est observé.

Le meilleur résultat de BLAST (TBLASTX) de la séquence consensus de CS058 contre la banque de données de NCBI est une glucanase de *Sambucus nigra* avec un score de blast de 176 et une e-value de 7^E-42 . Le deuxième résultat de BLAST a pour résultat une séquence EST de beta-1,3 glucanase chez un croisement entre *Cichorium intybus* et *Cichorium endivia* (score de 160 et e-value de 4^E-37). La séquence de laitue (1362 pb) orthologue au gène CS058 a pour premier résultat de BLAST la même séquence de chicorée codant pour une beta-1,3 glucanase (Score de 194, e-value de 2^E-126).

Quatre variations de nucléotides sont non silencieuses et induisent un changement fonctionnel de l'acide aminé. L'ensemble des mutations silencieuses et non silencieuses est repris dans le tableau 2.21.

En 2000, Helleboid et al ont montré que les beta-1,3 glucanases étaient encodées par une petite famille de 3 gènes chez *C. intybus* (Hybride 474) (Helleboid, Chapman et al. 2000). Nous pouvons émettre l'hypothèse suivante : la paire d'amorces dessinée pour CS058 a pu amplifier les 3 gènes de β 1,3 glucanase différents car ceux-ci présentent une forte homologie de séquences.

Pour tester cette hypothèse, nous avons aligné les trois séquences de β 1,3 glucanase disponibles sur NCBI avec la séquence EST de la database de chicorée associée à une β 1,3 glucanase sur base des résultats de *reciprocal best BLAST* hit et trois séquences représentatives des trois « paralogues » mis en évidence lors de ce travail. Nos trois paralogues apparaissent appartenir à un autre groupe que les séquences de NCBI. Elles sont caractérisées entre autres par une insertion de deux motifs TAA par rapport aux autres séquences voir illustration figure 2.5.

Les β 1,3 glucanases découvertes par Helleboid *et al.* ont été mises en évidence sur base de séquences cDNA exprimées dans des conditions spécifiques d'embryogenèse somatique. D'autres gènes codants pour des β 1,3 glucanases peuvent être exprimés dans d'autres situations. Nous avons potentiellement mis en évidence de nouvelles séquences de β 1,3 glucanases non apparentées à celles décrites précédemment par Helleboid *et al* (cfr figure 2.6)

2.4.2 CS60, homologue NBS-LRR

Le gène CS60 est un orthologue potentiel entre la laitue et la chicorée pour un gène CC-NBS-LRR. Le gène CS60 a été partiellement séquencé pour huit clones. La séquence amplifiée et séquencée est de 883pb. La taille élevée de la séquence rend difficile un recouvrement fiable des séquences obtenues par séquençage en T3 et T7.

Sur base des séquences disponibles, 29 mutations ont été mises en évidence. La fréquence de mutations est d'une mutation toutes les 30 pb.

L'alignement entre la séquence consensus génomique et la séquence EST, ayant servi à designer les amorces, montre trois exons. Huit mutations sont dans l'exon 1 (Sur base des résultats de séquençage ; l'« exon 1 » n'est pas forcément le premier exon du gène). Les 21 autres mutations sont localisées dans les introns. Les mutations introniques n'ont bien sûr pas d'impact direct sur la modification d'un acide aminé. Néanmoins, elles peuvent toucher des sites cryptiques et modifier entre autres l'épissage de l'ARN.

La traduction de l'exon 1 des séquences obtenues dans les six phases de lecture, n'a généré aucune séquence exempte de codon stop. La séquence EST ayant servi au design des amorces ne présente également aucune phase ouverte de lecture.

L'analyse BLAST de la séquence consensus de CS060 ne révèle pas d'homologie avec un gène CC-NBS-LRR attendu. La séquence EST ayant servi au design des amorces montre une homologie avec un gène CC-NBS-LRR chez le tournesol avec une faible valeur de BLAST de 43 et une e-value de 5^E-06 .

Nous sommes vraisemblablement en présence d'un pseudogène. L'amplification d'un pseudogène, soit un gène ayant perdu sa fonction, explique d'une part le grand nombre de mutations présentes dans le gène à travers les différents clones et explique également le résultat médiocre de BLAST. Un pseudogène n'est plus soumis à la pression de sélection, il accumule donc plus de mutations et sa séquence n'est donc pas conservée.

2.4.3 CS62, homologue de *Avr9/Cf-9 rapidly elicited protein 75*

Le gène CS62 de la chicorée est un orthologue potentiel de la laitue. Ces deux gènes orthologues présentent une homologie avec le gène *Avr9/Cf-9 rapidly elicited protein 75* (ACRE75) du tabac. CS62 a été séquencé chez huit clones et a une taille de 363 pb. L'alignement des séquences des huit clones montre la présence de 6 SNP soit une fréquence d'un SNP toutes les 60 pb et décrit 6 allèles différents.

La taille de la séquence amplifiée sur ADNg étant celle prédite par amplification sur ADNc, il n'y a pas d'intron dans la séquence du gène amplifiée. Sur les 363 pb de base séquencées sur ADNg, il y a 99,2% d'homologie avec la séquence d'ADNc ayant été utilisée pour le design des amorces. Deux SNP sont des mutations silencieuses. Les 4 autres SNP induisent un changement d'acides aminés mais deux d'entre eux seulement changent les propriétés de la chaîne latérale (cfr table 2.22).

L'analyse BLAST de la séquence consensus de CS62 montre que celle-ci présente une homologie avec un gène de résistance ACRE75 (*Avr9/Cf-9 rapidly elicited protein 75*) du tabac avec une e-value de 1^E-07 et un score de BLAST (TBLASTX) de 57,6.

Nous sommes potentiellement en présence d'un gène orthologue de la laitue codant pour un gène homologue de *Avr9/Cf-9 rapidly elicited protein 75* (ACRE75) du tabac. Chez la tomate, le gène *Cf-9* confère la résistance au champignon *C. fulvum*. Le mode d'action de la protéine *Cf-9* n'est pas encore connu à l'heure actuelle, mais elle semble être nécessaire à l'activation des mécanismes de défense chez les plantes et une augmentation de la transcription du gène *Cf-9* est observée dans plusieurs interactions hôte/pathogène. Certains gènes peuvent être dépendants de l'expression *Cf-9* comme des β -1,3-glucanases et des chitinases (Durrant, Rowland et al. 2000).

2.4.4 CS76, homologue *Cf-5 disease resistance protein*

Le gène CS76 est un orthologue potentiel de la laitue pour un gène présentant une forte homologie avec le gène de résistance *Cf-5 disease resistance protein-like* d'*A. thaliana*. Nous avons séquencé huit clones du gène CS76. Les amorces *forward* et *reverse* contiennent par erreur chacune un même site de restriction SpeI. L'insertion d'un produit PCR, contenant de part et d'autre un site de restriction SpeI, dans un vecteur restreint XbaI et XhoI montre que

cette étape d'insertion ligation est très robuste. Cependant, le clonage n'est pas orienté. Les amorces ont été dessinées sur une séquence EST de 431 pb. La séquence correspondante sur ADN génomique est d'environ 930 pb. La séquence analysée doit contenir au moins un intron. La taille élevée de la séquence à amplifier rend délicat le recouvrement des séquences obtenues par séquençage par les deux extrémités du gène. A partir de l'amorce *forward*, 457 pb ont pu être séquencées proprement. 401 pb ont été séquencées à partir de l'amorce *reverse*. Ces séquences s'arrêtent chacune sur de longues séquences de nucléotides répétés qui détériorent l'efficacité de la réaction de séquençage.

La comparaison entre la séquence ADN_g et la séquence EST ayant servi au design des amorces révèle la présence de deux introns dans la partie séquencée du *forward* et deux introns dans la partie séquencée *reverse*. Sur 11 SNP trouvés et une délétion d'un trinuéclotide C, cinq SNP sont localisés dans les exons. Deux d'entre eux induisent des mutations non synonymes (voir Table 2.23 et figure 2.7 et 2.8).

Sur un total de 858 pb séquencées il y a un SNP chaque 71pb en moyenne. Sur les huit clones séquencés, nous observons 6 allèles. L'analyse TBLASTX des séquences obtenues pour CS076 montre un résultat contre une protéine de résistance Cf-5 d'*A.thaliana* avec une e-value de 5^E-23 et un score de BLAST de 60. La séquence de laitue (971 pb), orthologue du gène CS76, présente une forte homologie avec le gène Cf-5 *disease resistance protein-like* d'*A. thaliana* (score de 462 et e-value de 7^E-128). L'analyse des séquences avec le logiciel Interproscan met en évidence un domaine LRR caractéristique des gènes type Cf-5 (Dixon, Hatzixanthi et al. 1998). Le gène Cf-5 chez la tomate confère la résistance au champignon *C. fulvum*. La séquence de laitue positionnée sur la carte génétique MAP2JMR3 est annotée comme similaire à HcrVf2, une protéine homologue des protéines de résistance au champignon *C. fulvum*. HcrVf2 confère la résistance à l'ascomycète *Venturia inaequalis* chez la pomme (Belfanti, Silfverberg-Dilworth et al. 2004). Ces familles de gènes sont caractérisées par un nombre variable de répétition LRR (Dixon, Hatzixanthi et al. 1998). Un homologue de ce gène chez la laitue et la chicorée peut être impliqué dans d'autres types de résistance via une reconnaissance spécifique de pathogènes à la laitue et à la chicorée.

2.4.5 CS78, homologue *Glycine max putative Hs1pro-1-like receptor*

Le gène CS78 est un orthologue potentiel d'un gène de laitue codant pour une séquence homologue *Glycine max putative Hs1pro-1-like receptor* du soja. Neuf clones du gène CS78 ont été séquencés. La taille totale du fragment amplifié est de 450pb. Il n'y a pas d'intron et la séquence consensus des neuf clones présente 98,9% d'homologie avec la séquence EST ayant servi à la conception des amorces. Sur base des clones séquencés, nous observons cinq allèles pour un total de cinq SNP sur l'ensemble de la séquence, soit une fréquence de un SNP toutes les 90 pb. (table 2.23)

Quel que soit l'allèle séquencé, la phase de lecture la plus probable contient un stop contre plusieurs dans le cas des cinq autres phases. Ce codon stop peut être le signe que nous sommes en présence d'un pseudogène. Ce codon stop peut également marquer la fin de la séquence codante et l'entrée dans la partie 3'UTR. Le séquençage d'une plus grande partie du gène devrait permettre de confirmer ou d'infirmer cette hypothèse. L'analyse de la séquence avec le logiciel Interproscan met en évidence un domaine Hs1pro-1 C-term (score 323, e-value $3,7^E-94$) caractéristique de la partie C-terminale de protéines de résistance aux nématodes. Le domaine C-terminal correspondant à la fin de la protéine, la présence d'un

codon stop est normale. Trois SNP induisent un changement d'acides aminés. Deux d'entre eux donnent un acide aminé non polaire à la place d'un acide aminé polaire. (table 2.23)

L'analyse TBLASTX contre la banque de données de NCBI montre une homologie de la séquence consensus des neuf clones avec la protéine de résistance de soja *Glycine max putative Hs1pro-1-like receptor* avec un score de 217 et une e-value de 4^E-54 . La séquence orthologue de laitue (1440pb) « BLAST » contre une protéine de résistance aux nématodes avec un score de 270 et une e-value de 6^E-129 . Elle présente également des résultats de blast pour la séquence de soja *Glycine max putative Hs1pro-1-like receptor* avec un score de 169 et une e-value de 4^E-120 . L'homologie de séquence entre CS78 et son orthologue chez la laitue avec une protéine de résistance aux nématodes Hs1pro est différente de l'annotation donnée sur la carte de la laitue (TIR-NBS-LRR). L'annotation de la carte n'est pas optimale et doit être abordée avec précaution.

2.4.6 CS160, homologue *ethylene responsive element binding factor 4*

Le gène CS160 est un orthologue potentiel entre la laitue et la chicorée pour un gène homologue d'une protéine *ethylene responsive element binding factor 4* (ERF) d'*A. thaliana* (résultat de BLAST de la séquence EST de laitue cartographiée (915pb), score= 114 et e-value= 2^E-22).

Sur base des résultats EcoTILLING, nous avons séquencé six clones. Après séquençage, il est apparu que deux gènes différents avaient été amplifiés. Les amorces utilisées ne sont pas spécifiques.

Une séquence de 361 pb a été amplifiée pour trois clones (Table 2.24). Cette séquence s'aligne parfaitement (100% d'identité) avec la séquence EST ayant servi au design des amorces. Il s'agit donc de la séquence potentiellement orthologue de la laitue codant pour une protéine ERF. Les résultats TBLASTX contre la banque de données de NCBI montrent une faible homologie de la séquence avec une protéine ERF du riz (score de 40 et e-value de 0,47). Deux SNP sont trouvés pour trois clones séquencés et trois allèles sont mis en évidence. Aucune phase de lecture n'est exempte de codon stop. Une phase de lecture ne contient qu'un unique codon stop. Il s'agit peut être du codon stop de la protéine. Seul un séquençage d'une plus grande partie du gène pourrait permettre de confirmer s'il s'agit bien du codon stop de la protéine ou si la séquence en contient d'autres et serait en fait un pseudogène. Aucun des deux SNP présents n'a d'effet sur ce codon stop.

Une séquence de 382 pb a été séquencée chez trois autres clones. La séquence s'aligne avec la séquence EST de chicorée uniquement au niveau des amorces (amorce *forward* 91% d'identité et 100% d'identité sur l'amorce *reverse*). La région comprise entre les amorces est totalement différente. Cette séquence montre une forte homologie avec des séquences de la sous unité ribosomique 26S de différentes espèces végétales avec un score de minimum de 304 et une e-value de 2^E-80 . Il pourrait être intéressant d'utiliser cette séquence pour des études de phylogénie en réalisant un arbre sur base des 100 séquences homologues d'espèces différentes trouvées dans NCBI homologues.

2.4.7 CS230, gène NBS-LRR

Le gène CS230 code potentiellement pour un gène NBS-LRR et a été trouvé dans la banque de données de NCBI. Nous disposons de la séquence CS230 pour dix clones sélectionnés lors

de l'analyse EcoTILLING en pool et individuellement (voir résultats précédents). La longueur du fragment analysé est de 424 pb. Sept SNP sont présents entre les dix clones pour une fréquence finale d'un SNP toutes les 42 pb. Nous observons cinq allèles (cfr table 2.25). La séquence consensus des dix clones présente une homologie de 99,3% avec la séquence EST ayant servi au design des amorces. Aucun intron n'est observé.

Une analyse BLAST (TBLASTX) contre la banque de données de NCBI montre que la séquence consensus CS230 est homologue à la séquence de chicorée AY153825.1 (score de 327 et e-value de 3^E-87). C'est précisément cette séquence de gène NBS-LRR qui avait servi au design des amorces. Les autres résultats de BLAST confirment l'homologie de la séquence avec les gènes NBS-LRR et ce domaine est également confirmé par une analyse Interproscan.

Trois SNP n'ont aucun impact sur la traduction. Trois SNP modifient les acides aminés correspondants sans induire un changement radical des propriétés de ces derniers. Un SNP induit un codon stop. Ce SNP présente donc un intérêt particulier : en fonction de l'allèle porté par la plante, la protéine sera tronquée ou non. L'impact d'une telle mutation sur un gène de résistance pourrait générer une sensibilité si le gène NBS-LRR CS230 est important dans la résistance à un pathogène. Cfr table 2.25.

2.4.8 CS242, gène NBS-LRR

Le gène CS242 est associé à la famille des gènes NBS-LRR et a été trouvé sur base d'homologie avec des séquences NBS-LRR d'*A. thaliana*. Huit clones ont été séquencés sur base des résultats EcoTILLING précédents. Après édition des séquences (vérification de la séquence en T3 et en T7, vérification des mutations à l'aide du chromatogramme, retrait des séquences du vecteur), les séquences des différents clones avaient des tailles comprises entre 415 et 422 nucléotides. Nous avons amplifié des paralogues. Ces séquences se sont alignées en trois groupes différents. (table 2.26)

Le premier groupe reprend les clones 30, 52, 27 et 60. La séquence consensus de ces quatre séquences a une taille de 422 pb. Le clone 27 est caractérisé par une délétion de six nucléotides successifs et le clone 52 par une délétion de trois nucléotides et d'une autre délétion d'un seul nucléotide. Ces délétions n'ont aucun impact sur le cadre de lecture sauf dans le cas du clone 52 où la délétion d'un T entraîne la génération de codons stop. Les séquences 30, 27 et 60 sont codantes. L'analyse BLAST (TBLASTX) de la séquence consensus du premier groupe contre la banque de données NCBI montre une forte homologie avec un gène candidat de résistance chez *L. sativa* (BLAST score = 203 ; e-value= 4^E-51) ainsi que des homologies significatives avec des gènes CC-NBS-LRR chez le tournesol (BLAST score = 140 ; e-value= min 4^E-32).

Le second groupe contient les clones 1, 28 et 35. Les clones 1 et 28 sont caractérisés par une délétion d'un T générant des codons stop. Seul le clone 35 contient une phase ouverte de lecture sans codon stop. L'analyse BLAST (TBLASTX) de la séquence du clone 35 montre une homologie avec un gène candidat de résistance chez *L. sativa* (BLAST score = 200 ; e-value= 4^E-49).

Le dernier groupe est composé de la séquence du clone 64. Aucune phase de lecture n'est trouvée dans cette séquence qui apparaît donc comme non codante.

Les résultats de séquençage des clones de CS242 montrent l'amplification d'une grande diversité de paralogues codants ou non. La paire d'amorces pour le gène CS242 ne semble pas être assez spécifique. L'alignement entre les séquences génomiques et la séquence d'ADNc n'a pas mis en évidence de présence d'intron.

2.4.9 CS284, gène *WAK-related*

Le gène CS284 est un gène candidat homologue des gènes WAK chez *A. thaliana*. Huit clones de CS284 ont été séquencés sur base des résultats EcoTILLING obtenus précédemment. 494 pb ont été séquencés et 26 mutations ont été mises en évidence. La fréquence de mutation est d'une mutation toutes les 19 pb. Suivant les mutations de chaque clone, ceux-ci peuvent être classés en trois groupes distincts. D'après la fréquence élevée de mutation, nous sommes vraisemblablement en présence de trois « paralogues ». (voir table 2.27)

L'alignement de la séquence consensus des huit clones avec la séquence d'ADNc montre une homologie de 98,4% entre les deux séquences. Il n'y a pas d'intron.

L'analyse BLAST (TBLASTX) contre la banque de données de NCBI montre que la séquence consensus CS284 est homologue d'une protéine *WAK-like* suivant une e-value de 3^E-19 et un score de 87,2. D'autre part une analyse Interproscan confirme la présence d'un domaine associé aux WAK avec une e-value de 3^E-36 .

La traduction dans les six phases de lecture n'a pas révélé de séquence sans codon stop. Une des phases de lecture contient seulement deux codons stop. La figure 2.9 montre l'alignement des séquences protéiques de huit clones ainsi que la position des deux codons stop. A droite de ces deux codons stop, on observe une séquence de 147 acides aminés sans interruption. L'analyse BLASTP de cette séquence dans NCBI révèle une homologie avec une séquence de WAK d'*A. thaliana* avec un score de 129 et une e-value de 5^E-29 . Une analyse BLASTX, moins sensible aux codons stop de la séquence CS284 n'a montré aucun alignement significatif. Suivant les résultats obtenus dans Spidey, les codons stop ne se trouvent pas dans un exon. Nous sommes vraisemblablement en présence d'un pseudogène. L'analyse de l'impact des mutations sur la séquence « codante » de CS284 n'a pas d'utilité si l'on venait à confirmer que nous sommes bien en présence d'une séquence non codante.

2.4.10 CS286, gène *WAK-related*

Le gène CS286 est un gène homologue des WAK. Sept clones ont été séquencés sur base de résultats d'EcoTILLING obtenus précédemment. La longueur du fragment amplifié est de 507 pb et huit SNP ont été mis en évidence après alignement des séquences. Sept des huit SNP se trouvent dans un individu qui paraît différent des autres clones. La fréquence de SNP est de 1 SNP toutes les 63 pb et on observe trois allèles sur les huit clones séquencés. (table 2.28)

L'alignement entre la séquence consensus de CS286 et la séquence EST ayant servi au design des amorces ne montre pas d'intron. Sur les 507 pb séquencés, l'ADNc et l'ADNg présentent 98,4% d'homologie.

L'analyse TBLASTX de la séquence consensus contre la banque de données de NCBI montre que la séquence présente une forte homologie avec une séquence *WAK-like* chez *A. thaliana* suivant un score de blast de 280 et une *e-value* de 6^E-73 . Une analyse interproscan a confirmé la présence d'un domaine *WAK-related*.

Six SNP sur huit ne causent aucun changement d'acide aminé. Un des allèles, défini par un SNP, est caractérisé par le passage d'un acide aminé polaire en acide aminé avec chaîne latérale ionisée basique (voir table 2.28). L'absence de codon stop dans la phase de lecture nous laisse espérer que nous sommes en présence d'un gène fonctionnel.

2.5 Etude du polymorphisme actuel chez la chicorée industrielle, comparaison avec les résultats obtenus sur la core collection.

Les analyses EcoTILLING ont permis de montrer qu'il existait une grande diversité de polymorphisme dans la core collection représentant la diversité allélique à l'origine de la sélection de la chicorée industrielle.

Des graines des individus de départ ont été conservées et il est possible de ré-obtenir les plants à l'origine de la sélection. La caractérisation phénotypique de ces individus ainsi que des études d'association complémentaires utilisant le polymorphisme, mis en évidence lors de ce mémoire, pourraient servir à mettre en évidence des allèles de résistance perdus lors de la sélection.

Peut-on retrouver dans les populations actuelles du polymorphisme mis en évidence sur la core collection ? Certains marqueurs ont-ils été sélectionnés ? En utilisant un clone de référence de la core collection, l'EcoTILLING peut être utilisé pour mettre rapidement en évidence le polymorphisme présent dans des cultivars actuels et le comparer au polymorphisme précédemment mis en évidence.

Nous avons à notre disposition l'ADN extrait de plus de 200 cultivars actuels. Ceux-ci ont été phénotypés et décrits comme présentant une sensibilité ou une résistance à la rouille, l'oïdium et à *Pseudomonas syringae*. La comparaison entre le polymorphisme des gènes candidats d'individus sensibles et résistants devrait mettre en évidence des marqueurs de résistance ou de sensibilité.

Une analyse EcoTILLING du gène candidat CS230 a été réalisée sur 24 cultivars actuels présentant un phénotype de résistance ou de sensibilité. Le clone CS230-1, ayant préalablement servi à l'analyse EcoTILLING individuelle de la core collection, a été utilisé comme référence.

Les cultivars actuels ont été analysés premièrement contre eux même afin de mettre en évidence des individus homozygotes et hétérozygotes pour un ou plusieurs SNP. Quatre individus sur 25 présentent un profil hétérozygote pour le SNP en position 228 pb.

L'analyse EcoTILLING contre référence a mis en évidence trois positions de SNP localisées à 228 pb, 285 pb et 379 pb. L'intensité des pics observés est faible. L'étape de digestion ne s'est peut être pas déroulée correctement. Sur base de ces résultats, à considérer avec précaution, il y a potentiellement trois positions de SNP possible entre la séquence de référence et les cultivars actuels pour le gène CS230.

Le SNP en positionné à 228 pb à partir de l'amorce *forward* n'avait pas été mis en évidence lors de la caractérisation du polymorphisme de la core collection. Les SNP situés à 285 pb et 379 pb de l'amorce *forward* avaient déjà été mis en évidence. Le SNP à 285pb n'induisait pas de changement d'acide aminé. Le clone CS230-1 contient un SNP à une position de 376 pb induisant la création d'un codon stop. C'est peut-être ce même SNP qui pourrait être présent dans les variétés actuelles.

L'analyse EcoTILLING devra être recommencée pour le gène CS230. Le séquençage ultérieur d'individus actuels pourra permettre de confirmer ou non qu'une partie du polymorphisme trouvé dans la core collection est toujours présent aujourd'hui.

Conclusions et perspectives

La culture de la chicorée industrielle peut être affectée principalement par deux maladies, la rouille et l'oïdium, causées respectivement par les champignons *Puccinia cichorii* et *Erysiphe cichoracearum*. Plusieurs approches sont envisageables dans la lutte contre les pathogènes. L'approche la plus utilisée à l'heure actuelle consiste à recourir aux fongicides. L'utilisation de ces derniers est cependant controversée, à cause de leur impact sur l'environnement et pour la sécurité du consommateur. La sélection conventionnelle ou assistée par marqueurs moléculaires permet de sélectionner les individus naturellement résistants aux maladies et par la même occasion de diminuer le recours aux fongicides.

Ce mémoire s'inscrit dans une démarche de sélection assistée par marqueurs moléculaires.

Nous avons opté pour l'approche « gène candidat ». Une liste de gènes associés à la résistance de la chicorée a été constituée sur base de recherches bioinformatiques. La recherche a été principalement focalisée sur la mise en évidence de gènes de résistance NBS-LRR. On retrouve également parmi nos gènes candidats : des homologues de gènes clefs dans les voies de défense d'*A. thaliana* ; des séquences associées aux gènes WAK ; diverses séquences présentant des homologies avec des gènes de résistance de la banque de données de NCBI. Vingt-et-un orthologues potentiels entre la laitue et la chicorée ont été proposés en utilisant la méthode du *reciprocal best BLAST hit*. Cette méthode, basée sur des analyses BLAST, permet la mise en évidence d'orthologues entre deux espèces à partir d'EST. Des amorces ont été conçues pour la majorité des gènes candidats mis en évidence, et 27 paires d'amorces ont déjà été testées avec succès par PCR.

La mise en évidence de gènes candidats a été réalisée sur une banque d'EST de chicorée. Des gènes de résistance peuvent ne pas être exprimés en conditions non pathologiques et ne se trouvent donc pas dans nos banques d'EST.

La recherche d'orthologues entre la laitue et la chicorée s'est basée sur des banques d'EST de ces deux espèces. Il est probable que les résultats obtenus auraient été différents si nous avions eu à notre disposition les génomes complets de la laitue et de la chicorée. Une banque EST ne contient pas l'ensemble des gènes présents dans le génome concerné, mais seulement ceux exprimés au moment de l'échantillonnage. D'autre part, seules des études fonctionnelles et la cartographie des gènes candidats pourront confirmer que nous sommes bien en présence d'orthologues.

Excepté pour les séquences de chicorée associées à la résistance et déjà disponibles dans la banque de données de NCBI, tous les gènes candidats ont été mis en évidence à l'aide d'homologies de séquences ou de recherche de domaines. La recherche d'homologies a été principalement réalisée suivant l'algorithme TBLASTX. Celui-ci apparaît adapté à la mise en évidence d'homologies inter-spécifiques. Il est important de préciser qu'il ne s'agit que de prédictions auxquelles nous avons attribué un degré de confiance (score de BLAST et *e-value*). Seule l'analyse fonctionnelle des gènes candidats proposés durant ce travail pourra

valider ou infirmer la fonction attribuée aux gènes. En absence de cette analyse, il nous paraît indispensable de diversifier les approches utilisées afin de maximiser la détection de gènes candidats et d'évaluer au mieux la confiance accordée à ces derniers.

L'annotation des bases de données permet la mise en évidence facile et rapide de séquences potentiellement impliquées dans une fonction recherchée. C'est dans ce but que nous avons annoté les 5,000 premières séquences de la banque d'unigènes de chicorée à l'aide du logiciel Interproscan. L'annotation du quart de la banque d'unigènes a permis de mettre en évidence des gènes associés à la résistance aux pathogènes et de nombreuses autres fonctions, facilement accessibles sur base de mots-clefs. L'analyse manuelle, basée sur une version en ligne du logiciel, est contraignante en temps et en manipulation. L'installation du logiciel en local sur une machine UNIX/Perl adaptée (voir matériel et méthodes) devra permettre l'analyse rapide et automatisée des 17,200 dernières séquences non encore annotées. L'automatisation de l'analyse Interproscan pourrait également être utilisée pour l'annotation de la banque de chicorée contenant 54,000 EST ayant servi à la création de la banque d'unigènes.

Des méthodes alternatives à l'approche *in silico* pourraient permettre de découvrir des gènes candidats. L'approche de mutagenèse dirigée, bien que théoriquement réalisable, nécessite un investissement beaucoup trop lourd pour être appliquée à l'étude de la chicorée industrielle.

Une approche alternative à la mise en évidence de gènes potentiellement impliqués dans la résistance aux pathogènes est basée sur des études de transcriptomique comparative. En comparant le profil d'expression d'individus résistants et sensibles soumis à l'attaque d'un pathogène, il devrait être possible de trouver les gènes spécifiquement impliqués dans la résistance ou la sensibilité au pathogène donné. L'analyse du polymorphisme de ces séquences devrait très certainement mener à la mise en évidence de marqueurs de résistance et de sensibilité.

Au terme de ce mémoire, nous avons développé une approche robuste permettant le clonage orienté de nos gènes candidats dans des bactéries thermocompétentes TOP10F', permettant ainsi l'amplification de chaque gène candidat à l'aide d'amorces universelles T3-T7 marquées. L'analyse EcoTILLING avec référence a permis la mise en évidence du polymorphisme de nos gènes candidats et la sélection de clones représentatifs de la diversité allélique qui seront sélectionnés pour confirmer et caractériser la nature du polymorphisme. L'analyse en pool s'est montrée très sensible à la présence de paralogues. Elle génère des faux positifs lorsqu'il y a trop de polymorphisme présent dans une population. Néanmoins, cette approche devrait pouvoir être utilisée pour la mise en évidence de mutations rares dans une population peu polymorphe comme par exemple la comparaison d'individus sensibles et résistants ayant un background génétique très proche.

L'analyse EcoTILLING avec référence génère des résultats plus propres que l'approche en pool. Elle est cependant plus lourde à mettre en place pour l'analyse d'une population d'ADN. Son utilisation a été validée pour permettre une évaluation de la fréquence des SNP dans une population testée.

L'étape d'EcoTILLING doit être optimisée. Nous avons encore une trop grande variabilité dans les résultats et un taux de faux positifs/négatifs trop important. L'amplification de

chaque gène candidat avec la même paire d'amorces (T3-T7) peut, suite à des contaminations de pipettes et du matériel utilisé, générer un bruit de fond et altérer les résultats.

Le coût en baisse des réactions de séquençage pourrait être une alternative à la technique de l'EcoTILLING. Il est aujourd'hui possible d'obtenir d'excellents résultats de séquençage sur une longueur de 800 pb pour environ 4€ (séquençage Macrogen). Il serait intéressant de quantifier le coût d'une analyse EcoTILLING (telle que réalisée dans ce mémoire) et de le comparer au coût du séquençage d'un nombre de clones sélectionnés au hasard, représentatifs de la diversité du polymorphisme.

Si cette solution apparaissait financièrement réalisable, elle présenterait de nombreux avantages dont la simplicité d'utilisation et la caractérisation immédiate du polymorphisme. Durant ce travail, nous avons étudié le polymorphisme de 10 gènes candidats associés à la pathogenèse. La mise en évidence d'un grand nombre de codons stop trouvés dans certains gènes candidats peut être due à la présence de pseudogènes. Un pseudogène n'est plus fonctionnel et est a priori inutile pour la résistance aux pathogènes. Pour maximiser la mise en évidence de gènes codants, chaque séquence EST de gène candidat doit être préalablement traduite dans les six phases de lecture. Ce test permettrait d'éviter l'analyse complète de séquences non codantes.

L'étude du polymorphisme des gènes candidats a mis en évidence la présence de paralogues dans un certain nombre de gènes étudiés. L'amplification de gènes paralogues est due généralement à des amorces non spécifiques ou à une grande conservation entre les paralogues. Afin d'éviter ces amplifications, il serait utile de tester préalablement les amorces, à défaut d'un génome complet, *in silico* sur la banque de données d'EST. A cette fin, le logiciel FastPCR multiplateforme de l'université d'Helsinki pourrait être utilisé. Les amorces devraient être optimisées pour une séquence spécifique.

Au cours de ce mémoire, nous avons choisi de rechercher les gènes NBS-LRR. Le choix de cette famille de gènes augmente la probabilité d'amplification non spécifique car ces gènes contiennent un domaine conservé NBS. Le domaine LRR est également très fréquent. Les gènes NBS-LRR apparaissent sous forme de clusters et se diversifient sur base de recombinaisons homologues : un gène NBS-LRR récemment apparu est donc très proche de son ancêtre. La probabilité d'amplifier ces deux gènes avec une seule paire d'amorces est alors élevée. Des tests PCR *in silico* pourraient réduire également ce risque d'amplification non spécifique.

Les analyses du polymorphisme faites par EcoTILLING et confirmées par les résultats de séquençage, ont mis en évidence une grande diversité de SNP dans la population de chicorées présente à l'origine de la sélection.

Le polymorphisme trouvé dans la core collection pourra être utilisé dans des études d'associations. Pour cela, une étape de phénotypage de la core collection devra être réalisée. Des allèles de résistance perdus lors de la sélection de la chicorée pourraient alors être ré-introduits dans les variétés actuelles par croisement, et ainsi permettre la création de nouvelles variétés naturellement résistantes.

Comme nous l'avons vu précédemment, les études d'association nécessitent le phénotypage des individus génotypés pour les SNP mis en évidence. Cette étape de phénotypage de la core collection est lourde à mettre en place. Cependant, nous disposons depuis peu de l'ADN extrait de plus de 200 individus actuels phénotypés pour leur résistance ou leur sensibilité à un pathogène. Il sera intéressant de vérifier si du polymorphisme présent dans la core collection est encore présent dans les variétés actuelles. Ce polymorphisme pourra alors être utilisé pour des études d'associations sur les individus actuels.

Dans ce travail, nous avons confirmé l'existence d'une grande diversité allélique à l'origine de la sélection de la chicorée industrielle. Il serait intéressant d'étudier le polymorphisme des populations de chicorées actuelles. La diversité allélique est-elle toujours aussi importante aujourd'hui ? S'est-elle réduite avec la sélection effectuée durant les trente dernières années ? Le polymorphisme actuel devra également être caractérisé et, après des études d'association, pourrait permettre une sélection assistée par marqueurs moléculaires chez les cultivars actuels.

Les études d'association se réalisant sur base d'analyses statistiques, le génotypage d'un grand nombre d'individus sera nécessaire. Le typage d'une population entière d'individus sensibles et résistants doit être abordé par une approche à haut débit. La création d'un microarray de type Illumina Golden Gate assay, défini sur base des SNP caractérisés, pourrait permettre le typage d'une population à haut débit. Afin de réduire le coût d'une telle technique, il serait utile d'envisager l'adaptation du principe de la technique Illumina à un microarray mis au point et conçu au laboratoire ou d'envisager une autre technique de typage.

Bibliographie

- Afzal, A. J., A. J. Wood and D. A. Lightfoot (2008). "Plant receptor-like serine threonine kinases: Roles in signaling and plant defense." Molecular Plant-Microbe Interactions 21(5): 507-517.
- Anderson, C. M., T. A. Wagner, M. Perret, Z. H. He, D. Z. He and B. D. Kohorn (2001). "WAKs: cell wall-associated kinases linking the cytoplasm to the extracellular matrix." Plant Molecular Biology 47(1-2): 197-206.
- Andreasson, E., T. Jenkins, P. Brodersen, S. Thorgrimsen, N. H. Petersen, S. Zhu, J. L. Qiu, P. Micheelsen, A. Rocher, M. Petersen, M. A. Newman, H. Bjorn Nielsen, H. Hirt, I. Somssich, O. Mattsson and J. Mundy (2005). "The MAP kinase substrate MKS1 is a regulator of plant defense responses." Embo J 24(14): 2579-89.
- Ayliffe, M. A. and E. S. Lagudah (2004). "Molecular genetics of disease resistance in cereals." Ann Bot (Lond) 94(6): 765-73.
- Barkley, N. A., M. L. Wang, A. G. Gillaspie, R. E. Dean, G. A. Pederson and T. M. Jenkins (2008). "Discovering and verifying DNA polymorphisms in a mung bean [*V. radiata* (L.) R. Wilczek] collection by EcoTILLING and sequencing." BMC Res Notes 1: 28.
- Becraft, P. W. (2002). "Receptor kinase signaling in plant development." Annual Review of Cell and Developmental Biology 18: 163-192.
- Belfanti, E., E. Silfverberg-Dilworth, S. Tartarini, A. Patocchi, M. Barbieri, J. Zhu, B. A. Vinatzer, L. Gianfranceschi, C. Gessler and S. Sansavini (2004). "The HcrVf2 gene from a wild apple confers scab resistance to a transgenic cultivated variety." Proc Natl Acad Sci U S A 101(3): 886-90.
- Birch, P. R., A. P. Rehmany, L. Pritchard, S. Kamoun and J. L. Beynon (2006). "Trafficking arms: oomycete effectors enter host plant cells." Trends Microbiol 14(1): 8-11.
- Bittel, P. and S. Robatzek (2007). "Microbe-associated molecular patterns (MAMPs) probe plant immunity." Current Opinion in Plant Biology 10(4): 335-341.
- Botstein D, W. R., Skolnick M Davies RW (1980). "Construction of a genetic linkage map in man using restriction fragment length polymorphisms." Am J Hum Genet.
- Brookes, A. J. (1999). "The essence of SNPs." Gene 234(2): 177-86.
- Caplan, J., M. Padmanabhan and S. P. Dinesh-Kumar (2008). "Plant NB-LRR immune receptors: from recognition to transcriptional reprogramming." Cell Host Microbe 3(3): 126-35.
- Castiglioni, B., E. Rizzi, A. Frosini, K. Sivonen, P. Rajaniemi, A. Rantala, M. A. Mugnai, S. Ventura, A. Wilmotte, C. Boutte, S. Grubisic, P. Balthasart, C. Consolandi, R. Bordoni, A. Mezzelani, C. Battaglia and G. De Bellis (2004). "Development of a universal microarray based on the ligation detection reaction and 16S rRNA gene polymorphism to target diversity of cyanobacteria." Appl Environ Microbiol 70(12): 7161-72.
- Chini, A. and G. J. Loake (2005). "Motifs specific for the ADR1 NBS-LRR protein family in Arabidopsis are conserved among NBS-LRR sequences from both dicotyledonous and monocotyledonous plants." Planta 221(4): 597-601.
- Chisholm, S. T., G. Coaker, B. Day and B. J. Staskawicz (2006). "Host-microbe interactions: Shaping the evolution of the plant immune response." Cell 124(4): 803-814.

- Christine KEYSER, E. P. (2006). "Utilisation des SNP pour l'identification humaine." Spectra Analyse.
- Colbert (2001). "High-Throughput Screening for Induced Point Mutations " Plant Physiol Vol. 126: pp. 480-484.
- Comai, L., K. Young, B. J. Till, S. H. Reynolds, E. A. Greene, C. A. Codomo, L. C. Enns, J. E. Johnson, C. Burtner, A. R. Odden and S. Henikoff (2004). "Efficient discovery of DNA polymorphisms in natural populations by Ecotilling." Plant J 37(5): 778-86.
- Cornelis (2003). "Guide de Bonnes Pratiques en Culture de Chicorée Industrielle." IRBAB.
- Dangl, J. L. and J. D. Jones (2001). "Plant pathogens and integrated defence responses to infection." Nature 411(6839): 826-33.
- Dangl, J. L. and J. M. McDowell (2006). "Two modes of pathogen recognition by plants." Proc Natl Acad Sci U S A 103(23): 8575-6.
- Dauchot, N. (2007). EST derived molecular markers applied to polyfructan metabolism in *Chichorium intybus*. Presse universitaire de Namur.
- de Wit, P. J. G. M. (2007). "Visions & reflections (minireview) - How plants recognize pathogens and defend themselves." Cellular and Molecular Life Sciences 64(21): 2726-2732.
- Decreux, A., A. Thomas, B. Spies, R. Brasseur, P. Van Cutsem and J. Messiaen (2006). "In vitro characterization of the homogalacturonan-binding domain of the wall-associated kinase WAK1 using site-directed mutagenesis." Phytochemistry 67(11): 1068-79.
- DeYoung, B. J. and R. W. Innes (2006). "Plant NBS-LRR proteins in pathogen sensing and host defense." Nat Immunol 7(12): 1243-9.
- Dixon, M. S., K. Hatzixanthis, D. A. Jones, K. Harrison and J. D. Jones (1998). "The tomato Cf-5 disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number." Plant Cell 10(11): 1915-25.
- Dong, X. (1998). "SA, JA, ethylene, and disease resistance in plants." Curr Opin Plant Biol 1(4): 316-23.
- Druart, N., J. De Roover, W. Van den Ende, P. Goupil, A. Van Laere and S. Rambour (2001). "Sucrose assimilation during early developmental stages of chicory (*Cichorium intybus* L.) plants." Planta 212(3): 436-43.
- Durrant, W. E., O. Rowland, P. Piedras, K. E. Hammond-Kosack and J. D. Jones (2000). "cDNA-AFLP reveals a striking overlap in race-specific resistance and wound response gene expression profiles." Plant Cell 12(6): 963-77.
- Ellis, J. G., P. N. Dodds and G. J. Lawrence (2007). "Flax rust resistance gene specificity is based on direct resistance-avirulence protein interactions." Annu Rev Phytopathol 45: 289-306.
- Eulgem, T. and I. E. Somssich (2007). "Networks of WRKY transcription factors in defense signaling." Curr Opin Plant Biol 10(4): 366-71.
- Fan, J. B., M. S. Chee and K. L. Gunderson (2006). "Highly parallel genomic assays." Nat Rev Genet 7(8): 632-44.
- Finn, R. D., J. Tate, J. Mistry, P. C. Coggill, S. J. Sammut, H. R. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. Sonnhammer and A. Bateman (2008). "The Pfam protein families database." Nucleic Acids Res 36(Database issue): D281-8.
- Flor (1971). "Current status of the gene-for-gene concept." Annu Rev Phytopath.
- Gibson, M. (2004). Précis de génomique. Bruxelles.
- Gilchrist, E. J., G. W. Haughn, C. C. Ying, S. P. Otto, J. Zhuang, D. Cheung, B. Hamberger, F. Aboutorabi, T. Kalynyak, L. Johnson, J. Bohlmann, B. E. Ellis, C. J. Douglas and

- Q. C. Cronk (2006). "Use of Ecotilling as an efficient SNP discovery tool to survey genetic variation in wild populations of *Populus trichocarpa*." Mol Ecol 15(5): 1367-78.
- Glazebrook, J. (2001). "Genes controlling expression of defense responses in Arabidopsis--2001 status." Curr Opin Plant Biol 4(4): 301-8.
- Glazebrook, J. (2005). "Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens." Annual Review of Phytopathology 43: 205-227.
- Gut, I. G. (2001). "Automation in genotyping of single nucleotide polymorphisms." Hum Mutat 17(6): 475-92.
- Hartl, D. L. (2004). "Genetics Analysis of genes and genomes." Jones and Bartlett Publishers: 73.
- He, Z. H., I. Cheeseman, D. He and B. D. Kohorn (1999). "A cluster of five cell wall-associated receptor kinase genes, Wak1-5, are expressed in specific organs of Arabidopsis." Plant Mol Biol 39(6): 1189-96.
- Helleboid, S., A. Chapman, T. Hendriks, D. Inze, J. Vasseur and J. L. Hilbert (2000). "Cloning of beta-1,3-glucanases expressed during Cichorium somatic embryogenesis." Plant Mol Biol 42(2): 377-86.
- Helleboid, S., T. Hendriks, G. Bauw, D. Inze, J. Vasseur and J. L. Hilbert (2000). "Three major somatic embryogenesis related proteins in Cichorium identified as PR proteins." J Exp Bot 51(348): 1189-200.
- Hermann (2006). "Les maladies foliaires en chicorée industrielle: Recommandations pour une protection fongicide raisonnée." IRBAB: 429.
- Hopkins, Ed. (2003). Physiologie Végétale, de boek.
- Humphrey, T. V., D. T. Bonetta and D. R. Goring (2007). "Sentinels at the wall: cell wall receptors and sensors." New Phytol 176(1): 7-21.
- Jones, J. D. and J. L. Dangl (2006). "The plant immune system." Nature 444(7117): 323-9.
- Kiers, A. M., T. H. Mes, R. van der Meijden and K. Bachmann (2000). "A search for diagnostic AFLP markers in Cichorium species with emphasis on endive and chicory cultivar groups." Genome 43(3): 470-6.
- Kohler, A., C. Rinaldi, S. Duplessis, M. Baucher, D. Geelen, F. Duchaussoy, B. C. Meyers, W. Boerjan and F. Martin (2008). "Genome-wide identification of NBS resistance genes in *Populus trichocarpa*." Plant Mol Biol 66(6): 619-36.
- Koornneef, A. and C. M. Pieterse (2008). "Cross talk in defense signaling." Plant Physiol 146(3): 839-44.
- Liu, J. (2004). "DNA marker technologies and their applications in aquaculture genetics." Aquaculture 238: 1-37.
- Liu, J. and G. Coaker (2008). "Nuclear trafficking during plant innate immunity." Molecular Plant 1(3): 411-422.
- Liu, J., X. Liu, L. Dai and G. Wang (2007). "Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants." J Genet Genomics 34(9): 765-76.
- Loake, G. and M. Grant (2007). "Salicylic acid in plant defence--the players and protagonists." Curr Opin Plant Biol 10(5): 466-72.
- Lorenzo, O. and R. Solano (2005). "Molecular players regulating the jasmonate signalling network." Curr Opin Plant Biol 8(5): 532-40.
- Mackey, D. and A. J. McFall (2006). "MAMPs and MIMPs: proposed classifications for inducers of innate immunity." Mol Microbiol 61(6): 1365-71.

- McCallum (2000). "Targeting Induced Local Lesions IN Genomes (TILLING) for Plant Functional Genomics " Plant Physiol Vol. 123: pp. 439-442.
- McCann, H. C. and D. S. Guttman (2008). "Evolution of the type III secretion system and its effectors in plant-microbe interactions." New Phytol 177(1): 33-47.
- McDowell, J. M. and J. L. Dangl (2000). "Signal transduction in the plant immune response." Trends Biochem Sci 25(2): 79-82.
- McDowell, J. M. and S. A. Simon (2008). "Molecular diversity at the plant-pathogen interface." Developmental and Comparative Immunology 32(7): 736-744.
- McDowell, J. M. and B. J. Woffenden (2003). "Plant disease resistance genes: recent insights and potential applications." Trends Biotechnol 21(4): 178-83.
- McHale, L., X. Tan, P. Koehl and R. W. Michelmore (2006). "Plant NBS-LRR proteins: adaptable guards." Genome Biol 7(4): 212.
- Meyers, B. C., A. Kozik, A. Griego, H. Kuang and R. W. Michelmore (2003). "Genome-wide analysis of NBS-LRR-encoding genes in Arabidopsis." Plant Cell 15(4): 809-34.
- Meyers, B. C., M. Morgante and R. W. Michelmore (2002). "TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in Arabidopsis and other plant genomes." Plant J 32(1): 77-92.
- Michelmore, R. (2000). "Genomic approaches to plant disease resistance." Curr Opin Plant Biol 3(2): 125-31.
- Mondragon-Palomino, M., B. C. Meyers, R. W. Michelmore and B. S. Gaut (2002). "Patterns of positive selection in the complete NBS-LRR gene family of Arabidopsis thaliana." Genome Res 12(9): 1305-15.
- Montesano, M., G. Brader and E. T. Palva (2003). "Pathogen derived elicitors: searching for receptors in plants." Molecular Plant Pathology 4(1): 73-79.
- Moreira, R. F. and C. J. Noren (1995). "Minimum duplex requirements for restriction enzyme cleavage near the termini of linear DNA fragments." Biotechniques 19(1): 56, 58-9.
- Morgante, M. and F. Salamini (2003). "From plant genomics to breeding practice." Curr Opin Biotechnol 14(2): 214-9.
- Morin, A. (2004). "SNPs in ecology, evolution and conservation." Trends in Ecology & Evolution.
- Najimi, B. (2003). "Applications des marqueurs moléculaires dans l'amélioration du blé tendre pour la résistance aux maladies et aux insectes." Biotechnol. Agron. Soc. Environ.
- Nurnberger, T., F. Brunner, B. Kemmerling and L. Piater (2004). "Innate immunity in plants and animals: striking similarities and obvious differences." Immunol Rev 198: 249-66.
- Pflieger, S. (2001). "The candidate gene approach in plant genetics: a review." Molecular Breeding: 275-291.
- Pieterse, C. (2001). "Cross-talk between plant defence signalling pathways: boost or burdent?" AgBiotechNet 3.
- Quenon (2000). "NON-DESTRUCTIVE METHOD FOR INTERNAL QUALITY DETERMINATION OF BELGIAN ENDIVE (CICHORIUM INTYBUS L.)." Int. Agrophysics.
- Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder, R. Apweiler and R. Lopez (2005). "InterProScan: protein domains identifier." Nucleic Acids Res 33(Web Server issue): W116-20.
- Radwan, O., S. Gandhi, A. Heesacker, B. Whitaker, C. Taylor, A. Plocik, R. Kesseli, A. Kozik, R. W. Michelmore and S. J. Knapp (2008). "Genetic diversity and genomic

- distribution of homologs encoding NBS-LRR disease resistance proteins in sunflower." Mol Genet Genomics 280(2): 111-25.
- Rairdan, G. J. and P. Moffett (2006). "Distinct domains in the ARC region of the potato resistance protein Rx mediate LRR binding and inhibition of activation." Plant Cell 18(8): 2082-93.
- Rani, N. a. M. K. (2003). "Cardioprotective effects of Cichorium intybus in ageing myocardium of albino rats." Current Science 84(7): p. 941-943.
- Raven, Ed. (2007). Biologie végétale. Bruxelles, de boek.
- Schwessinger, B. and C. Zipfel (2008). "News from the frontline: recent insights into PAMP-triggered immunity in plants." Curr Opin Plant Biol 11(4): 389-95.
- Shah, J. (2003). "The salicylic acid loop in plant defense." Current Opinion in Plant Biology 6(4): 365-371.
- Shiu, S. H. and A. B. Bleeker (2001). "Plant receptor-like kinase gene family: diversity, function, and signaling." Sci STKE 2001(113): RE22.
- Shiu, S. H. and A. B. Bleeker (2001). "Receptor-like kinases from Arabidopsis form a monophyletic gene family related to animal receptor kinases." Proceedings of the National Academy of Sciences of the United States of America 98(19): 10763-10768.
- Sobrinho, B., M. Brion and A. Carracedo (2005). "SNPs in forensic genetics: a review on SNP typing methodologies." Forensic Sci Int 154(2-3): 181-94.
- Sonnhammer, E. L., S. R. Eddy and R. Durbin (1997). "Pfam: a comprehensive database of protein domain families based on seed alignments." Proteins 28(3): 405-20.
- Sorensen, B. S., L. P. Kiaer, R. B. Jorgensen and T. P. Hauser (2007). "The temporal development in a hybridizing population of wild and cultivated chicory (Cichorium intybus L.)." Mol Ecol 16(16): 3292-8.
- Stoddart, H. (2003). CHICORY. Interactive European Network for Industrial Crops and their Applications.
- Sulas, L. (2004). "Forage chicory: A valuable crop for Mediterranean environments." CIHEAM-IAMZ Cahiers Options Méditerranéennes v. 62.
- Takken, F. L., M. Albrecht and W. I. Tameling (2006). "Resistance proteins: molecular switches of plant defence." Curr Opin Plant Biol 9(4): 383-90.
- Thaler, J. S. (2002). "Cross-talk between jasmonate and salicylate plant defense pathways: effects on several plant parasites." Oecologia: 227-235.
- Till, B. J., T. Zerr, E. Bowers, E. A. Greene, L. Comai and S. Henikoff (2006). "High-throughput discovery of rare human nucleotide polymorphisms by Ecotilling." Nucleic Acids Res 34(13): e99.
- Trdan, S., N. Valic, J. Jerman, D. Ban and D. Znidarcic (2004). "Efficacy of three natural chemicals to reduce the damage of Erysiphe cichoracearum on chicory in two meteorologically different growing seasons." Journal of Phytopathology 152(10): 567-574.
- Triques (2008). "Mutation detection using ENDO1: Application to disease diagnostics in humans and TILLING and Eco-TILLING in plants." BMC Molecular Biology.
- Turner, J. G., C. Ellis and A. Devoto (2002). "The jasmonate signal pathway." Plant Cell 14 Suppl: S153-64.
- van der Biezen, E. A. and J. D. Jones (1998). "The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals." Curr Biol 8(7): R226-7.

- van Ooijen, G., G. Mayr, M. M. Kasiem, M. Albrecht, B. J. Cornelissen and F. L. Takken (2008). "Structure-function analysis of the NB-ARC domain of plant disease resistance proteins." J Exp Bot 59(6): 1383-97.
- Varshney, R. K., A. Graner and M. E. Sorrells (2005). "Genomics-assisted breeding for crop improvement." Trends Plant Sci 10(12): 621-30.
- Vasyukova, N. I. and O. L. Ozeretskovskaya (2007). "Induced plant resistance and salicylic acid: A review." Applied Biochemistry and Microbiology 43(4): 367-373.
- Vienne, D. (1998). "Les marqueurs moléculaires en génétique et biotechnologies végétales." INRA.
- Vignal, A., D. Milan, M. SanCristobal and A. Eggen (2002). "A review on SNP and other types of molecular markers and their use in animal genetics." Genet Sel Evol 34(3): 275-305.
- Vos P, H. R., Bleeker M, Reijans M, van der Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995). "AFLP: a new technique for DNA fingerprinting." Nuc Ac Res.
- Wang, X. (2006). "Analysis of single nucleotide polymorphisms and its application in a disputed paternity case." International Congress Series 1288: 76-78.
- Wiermer, M., B. J. Feys and J. E. Parker (2005). "Plant immunity: the EDS1 regulatory node." Curr Opin Plant Biol 8(4): 383-9.
- Zhu, C. (2008). "Status and Prospect of Association Mapping in Plants." The plant genome 1(NO. 1): 5-18.
- Zhu, M. and S. Zhao (2007). "Candidate gene identification approach: progress and challenges." Int J Biol Sci 3(7): 420-7.
- Zipfel, C. (2008). "Pattern-recognition receptors in plant innate immunity." Curr Opin Immunol 20(1): 10-6.
- Zipfel, C. and G. Felix (2005). "Plants and animals: a different taste for microbes?" Curr Opin Plant Biol 8(4): 353-60.



Figure 1.3: Plant de chicorée industrielle attaqué par le champignon *Erysiphe cichoracearum*. La feuille de chicorée illustrée ci dessus est recouverte d'un duvet blanchâtre caractéristique d'une plante atteinte d'oïdium. Photo (Cornelis 2003)



Figure 1.4: Plant de chicorée industrielle attaqué par le champignon *Puccinia cichorii*. La feuille de chicorée illustrée ci dessus est recouverte de taches brunâtres caractéristiques d'une plante atteinte de rouille. Photo (Cornelis 2003)

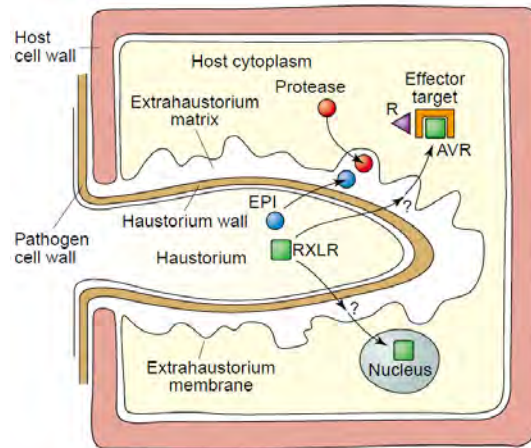


Figure 1.5 : L'haustorium, structure d'injection d'effecteurs dans la cellule végétale chez les champignons et les oomycètes. L'haustorium est une structure du champignon qui dégrade la paroi cellulaire végétale et s'invagine dans la cellule végétale. La cellule végétale sécrète des protéases pour lutter contre l'haustorium. Celui-ci dispose d'inhibiteurs de protéases qu'il sécrète à l'interface haustorium/ membrane cytoplasmique. Par des mécanismes encore inconnus, des effecteurs vont dans le cytoplasme de la cellule végétale ou dans le noyau de cette dernière si ils contiennent une séquence de localisation nucléaire. Illustration de l'haustorium de *Phytophthora infestans* d'une cellule de pomme de terre. (Birch, Rehmany et al. 2006)

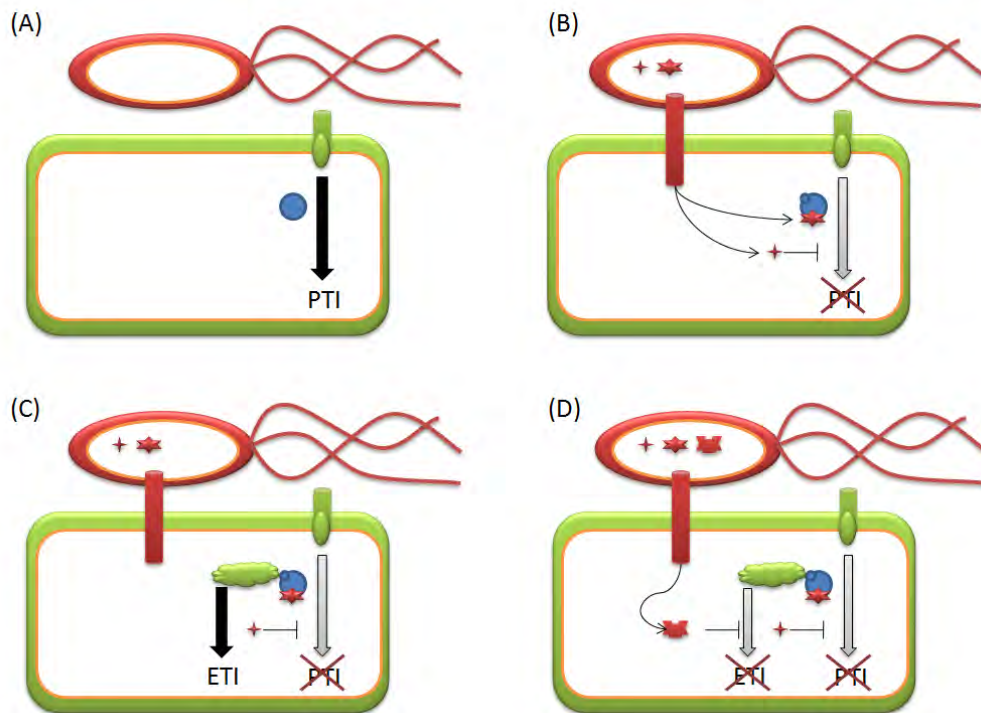


Figure 1.6: Interaction hôte- pathogène, co evolution hôte- pathogène. (A) Défense de type PTI (*PAMP Triggered Immunity*) de la cellule végétale. La cellule végétale perçoit la présence d'un pathogène en détectant des PAMP (*Pathogens Associated Molecular Patterns*) ou des MAMP (*Microbes Associated Molecular Patterns*) via des récepteurs de type sérine thréonine kinase. La détection du pathogène entraîne l'activation de voies de défense. (B) Evolution du pathogène et suppression du système PTI grâce à l'injection d'effecteurs dans la cellule végétale à l'aide du système de sécrétion de type 3 (TTSS) (chez les bactéries). Des effecteurs peuvent d'une part inhiber les réponses de défense (Etoile à 4 branches rouge illustrée) et peuvent d'autre part affecter d'autres voies de la cellule végétale (Etoile à 6 branches rouges illustrée). (C) Détection directe ou indirecte des effecteurs bactériens. Ici, la plante dispose d'une protéine (nuage vert illustré) qui détecte la modification d'une protéine due à un effecteur bactérien (Etoile à 6 branches rouges illustrée). La reconnaissance directe ou indirecte (cas illustré) induit une réponse de défense spécifique contre le pathogène appelée ETI (*Effectors Triggered Immunity*). (D) Adaptation de la bactérie. Suppression de la réponse ETI grâce à l'injection via le TTSS d'un effecteur inhibant la réponse ETI. Illustrations inspirées de (McCann and Guttman 2008)

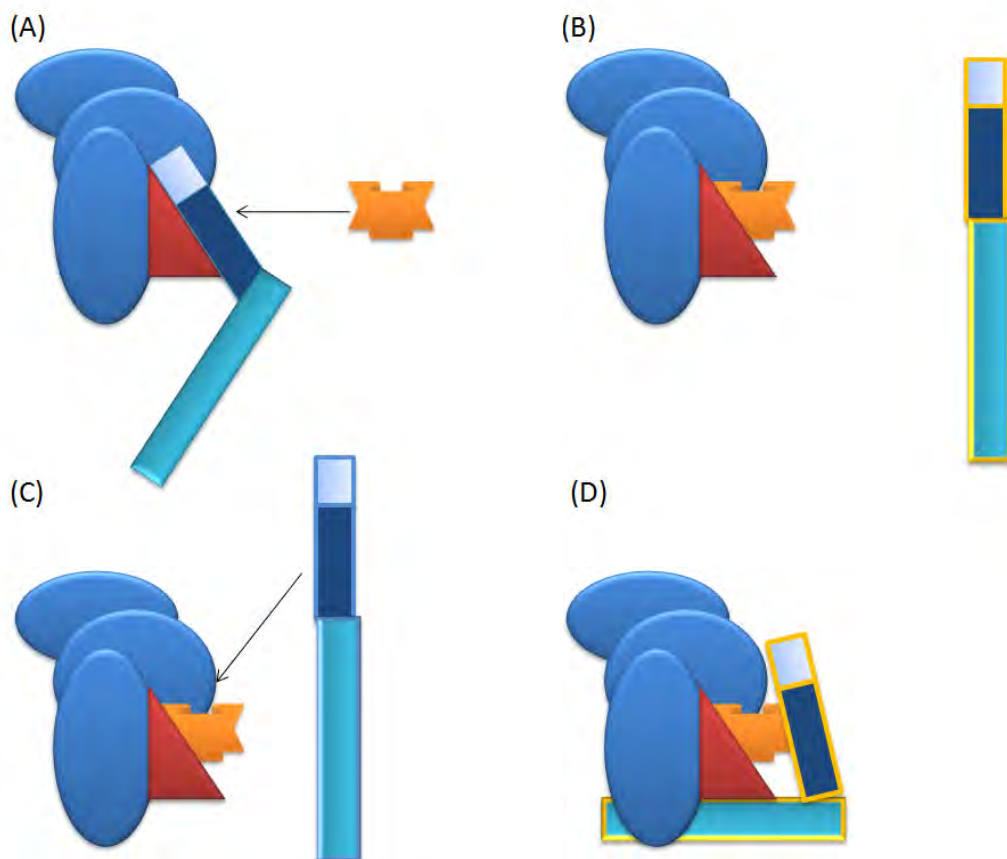


Figure 1.7: L'hypothèse de garde (*Guard Hypothesis*). Contrairement au modèle antérieur, *gene for gene* de FLOR, dans la grande majorité des cas, la détection de l'effecteur est indirecte. (A) Un complexe de protéines, cible potentielle d'un effecteur pathogène, est surveillé par une molécule de garde à laquelle vient se fixer un gène de défense NBS-LRR (inactif). (B) La fixation de l'effecteur est détectée par la molécule de garde. La détection de celui-ci induit la libération du gène NBS-LRR et l'activation de celui-ci (symbolisé par l'encadrement jaune) qui va activer des réponses spécifiques au pathogènes. (C) Autre fonctionnement possible : l'éliciteur interagit avec sa cible (complexe bleu) qui contient une protéine de garde. (D) La présence de l'éliciteur sur sa cible contenant une protéine de garde induit la reconnaissance de ce nouveau complexe par un gène de défense NBS-LRR qui est activé et active à son tour les défenses spécifiques contre le pathogène. Adaptation de (Dangl and Jones 2001)

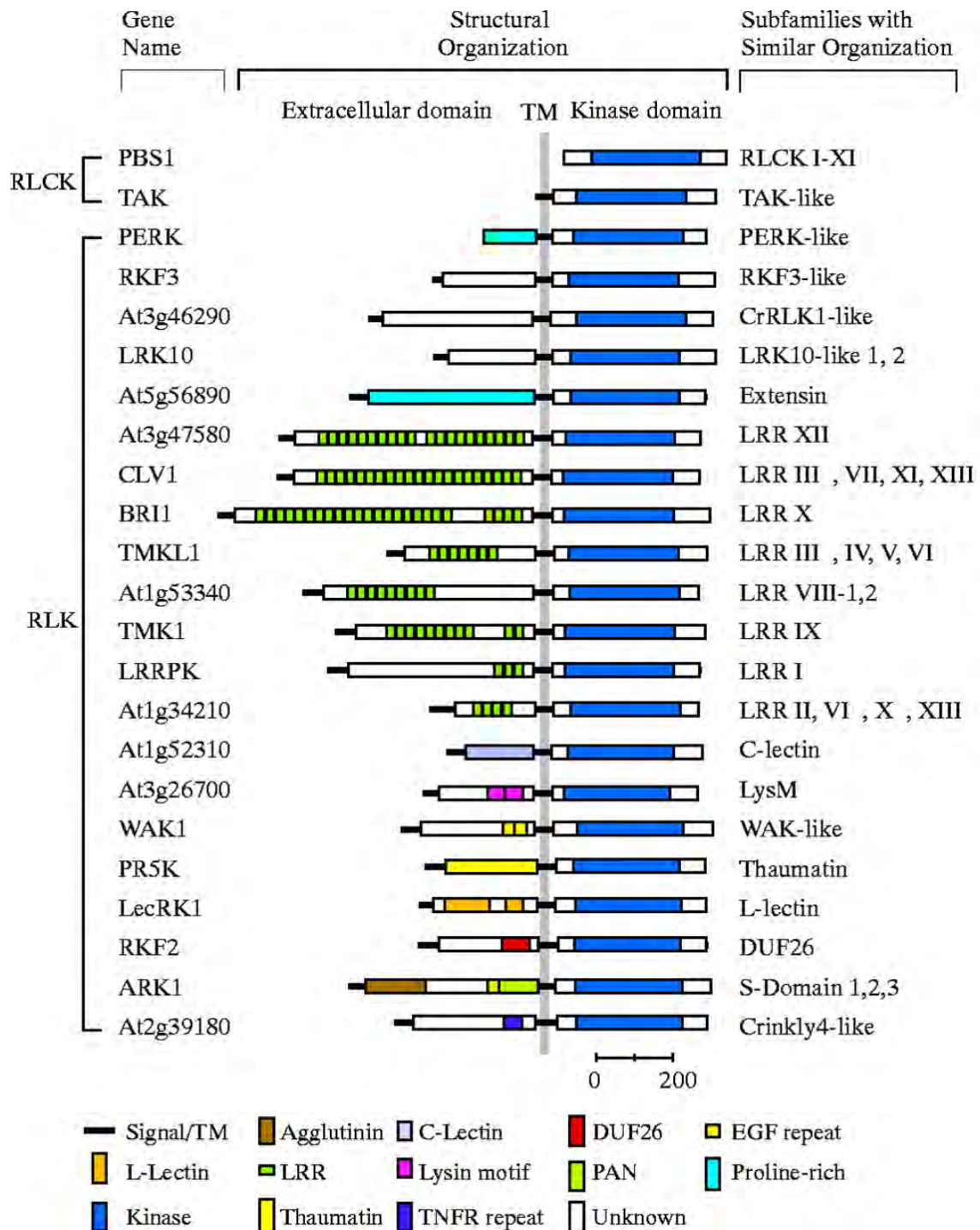


Figure 1.8 : Diversité des récepteurs like kinase chez les plantes. Les récepteurs contenant un domaine extracellulaire sont définis comme des récepteurs like kinase (RLK). Les récepteurs like kinase sans domaine extracellulaire sont appelés RKCK (récepteurs like kinase cytoplasmique). La ligne verticale grise représente la membrane cellulaire. Le signal peptide est présent chez les RLK mais n'est normalement pas présent dans les protéines matures. Un nom de gène modèle est donné à chaque sous famille de gène RLK. Les domaines ont été attribués à l'aide des banques de données SMART et Pfam. *DUF*, domain of unknown function; *EGF*, epidermal growth factor; *C lectin*, C-type lectin; *L-lectin*, legume lectin; *PAN*, plasminogen/apple/nematode protein domain; *TM*, transmembrane region; *TNFR*, tumor necrosis factor receptor. (Shiu and Bleecker 2001)

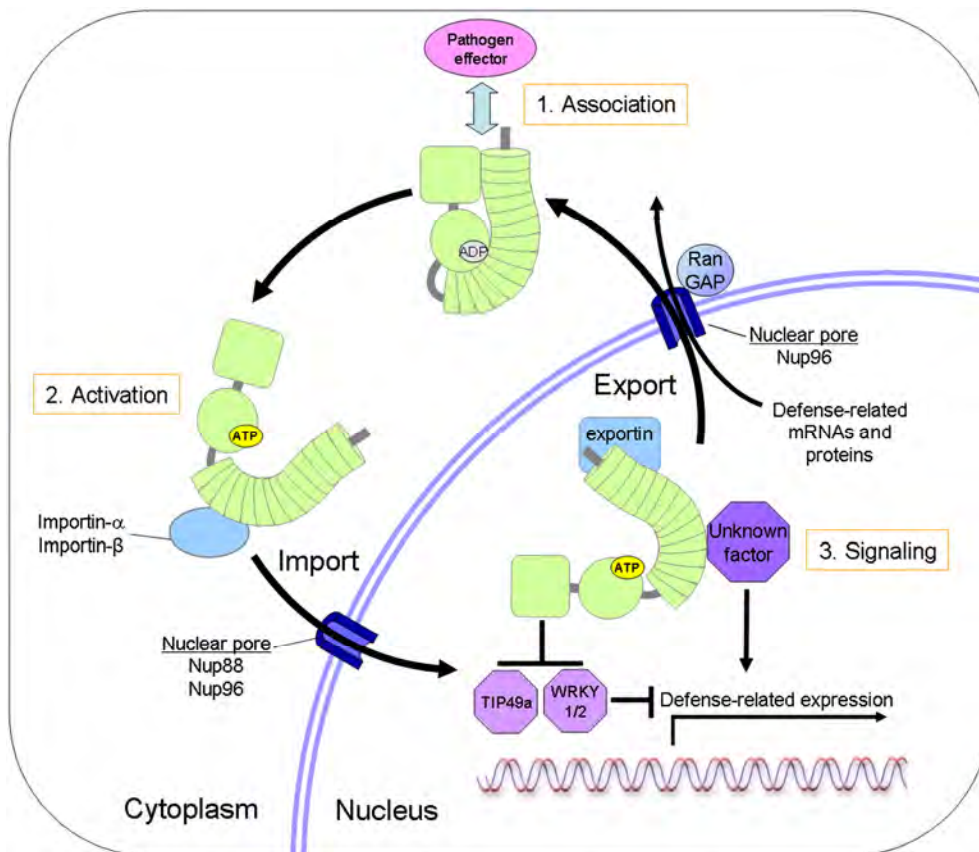


Figure 1.9: Modèle d'induction de défense par les protéines NBS-LRR. 1. Reconnaissance de l'effecteur pathogène (rose) par le gène NBS-LRR (vert). 2. La reconnaissance d'un effecteur induit un changement de conformation et la fixation d'une molécule d'ATP (jaune). La protéine NBS-LRR est activée. 3. L'activation de la protéine de défense est suivie de sa localisation du cytoplasme vers le noyau via des importines permettant le passage de pores nucléaires spécifiques. Dans l'illustration ci-dessus, la protéine NBS-LRR lève l'inhibition de gènes de défense causée par les protéines TIP49a et WRKY1/2 (mauve clair). L'hydrolyse de l'ATP marque l'inactivation de la protéine qui est ensuite exportée vers le cytoplasme (fixation d'exportine (bleu clair)) par un pore nucléaire. Des gènes NBS-LRR sont exclusivement localisés dans le noyau où ils peuvent être directement activés pour activer à leur tour l'expression de gènes de défense. L'activation dans le noyau peut être due à la présence d'un effecteur ou à d'autres facteurs (illustration ci-dessus d'un facteur inconnu en mauve). (Illustration : Caplan, Padmanabhan et al. 2008)

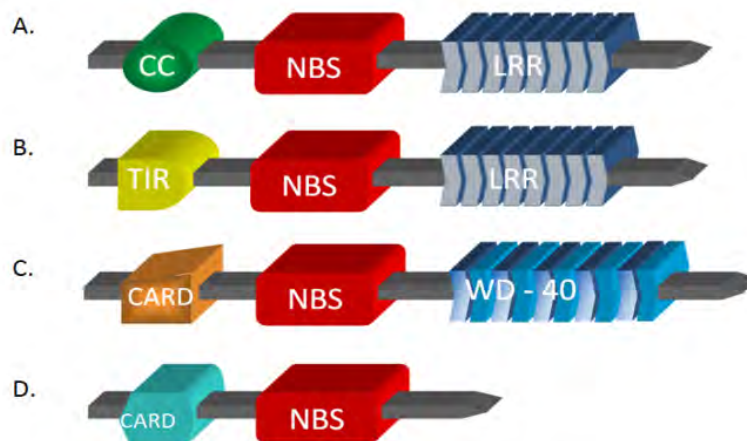


Figure 1.10: Conservation du domaine NBS entre protéines de défense végétales NBS-LRR (A et B) et effecteurs de mort cellulaire animale (C et D). A., CC-NBS-LRR (coiled coil – nucleotide binding site- leucine rich repeat); B., TIR-NBS-LRR (Toll/like interleukin receptor – nucleotide binding site- leucine rich repeat); C., Apaf-1; D., Ced4. (adapté de Dangl et Jones 2001)

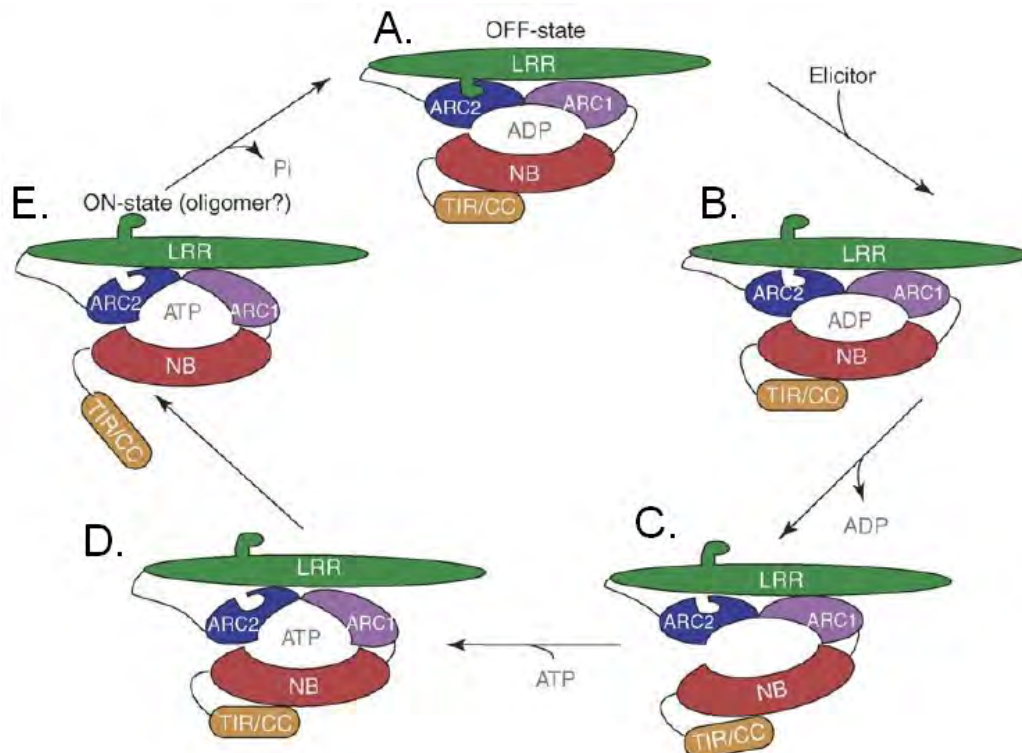


Figure 1.11: Modèle d'activation des gènes NBS-LRR suite à la détection d'un éliciteur. A. A l'état inactif, la protéine de défense NBS-LRR fixe une molécule d'ADP. B. Un éliciteur est reconnu de manière directe ou indirecte et induit un changement de conformation dans le domaine LRR (en vert). C. Libération de la molécule d'ADP. D. Fixation d'une molécule d'ATP. E. La fixation de la molécule d'ATP induit un changement de conformation au niveau du domaine N-terminal. La protéine est activée et va induire des réponses de défense spécifique au pathogène. Après hydrolyse de la molécule d'ATP en ADP, la protéine NBS-LRR retourne à son état inactif. (Takken, Albrecht et al. 2006)

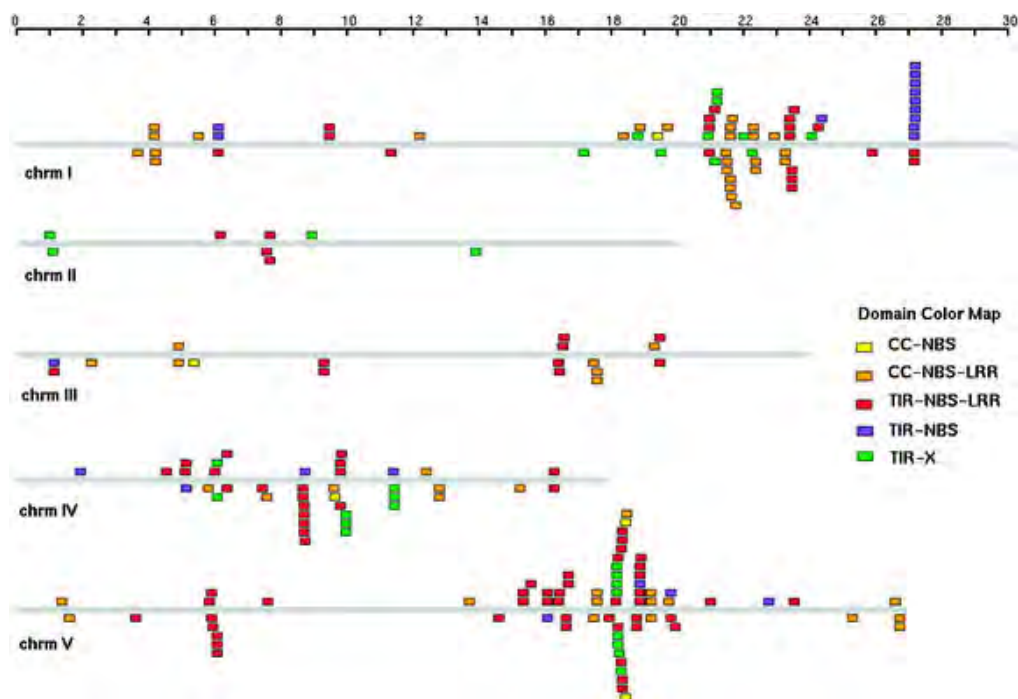


Figure 1.12: localisation majoritairement en clusters des gènes NBS-LRR et associés sur une carte génétique d'*A. thaliana*. Les 5 chromosomes d'*A. thaliana* (chrn I à V) apparaissent en gris. L'échelle située au dessus du schéma donne la position des gènes NBS-LRR et associés sur les différents chromosomes (en mégabases). Légende des codes couleurs directement sur le schéma. (Meyers, Kozik et al. 2003)

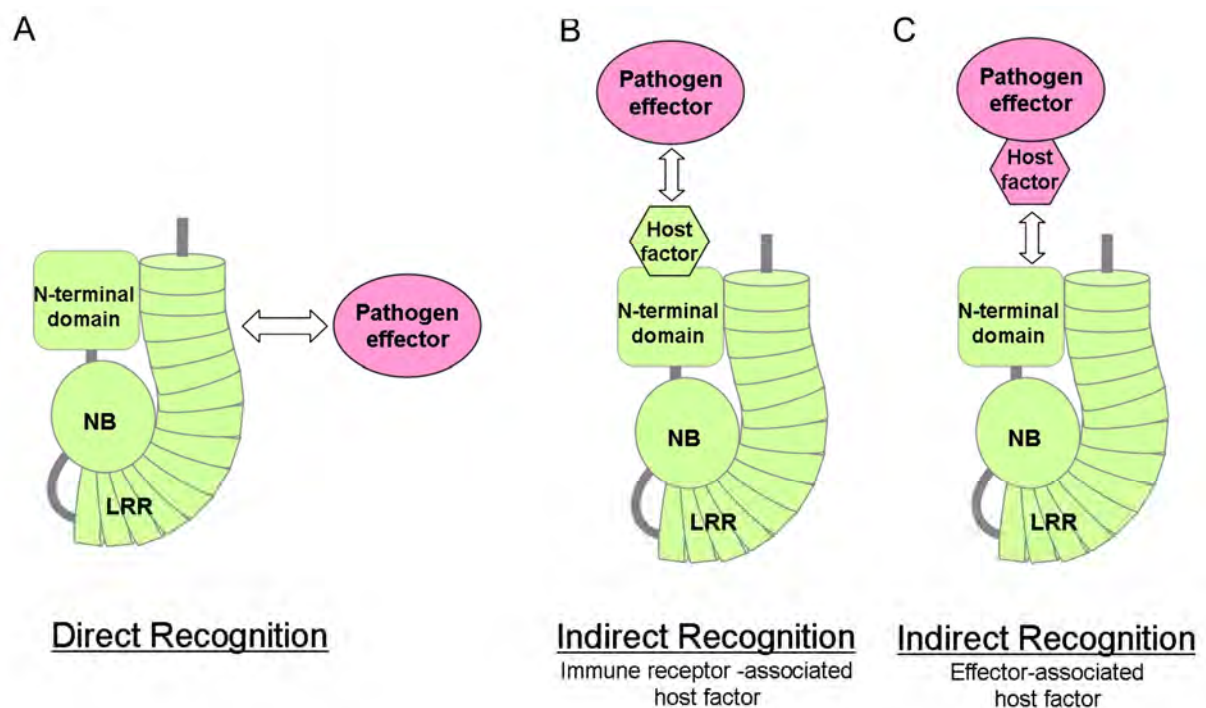


Figure 1.13: 3 modes de reconnaissance de l'effecteur pathogène par les protéines de défense NBS-LRR. A. Reconnaissance directe d'un effecteur du pathogène le plus souvent par le domaine LRR de la protéine. B. Reconnaissance indirecte d'un effecteur du pathogène, une protéine adaptatrice (Host factor) se lie au domaine N-terminal de la protéine (X, CC ou TIR) et permet la reconnaissance indirecte de l'effecteur (Pathogen effector). C. Reconnaissance d'un effecteur du pathogène sur lequel une protéine adaptatrice de la cellule végétale s'est fixée. (Caplan, Padmanabhan et al. 2008)

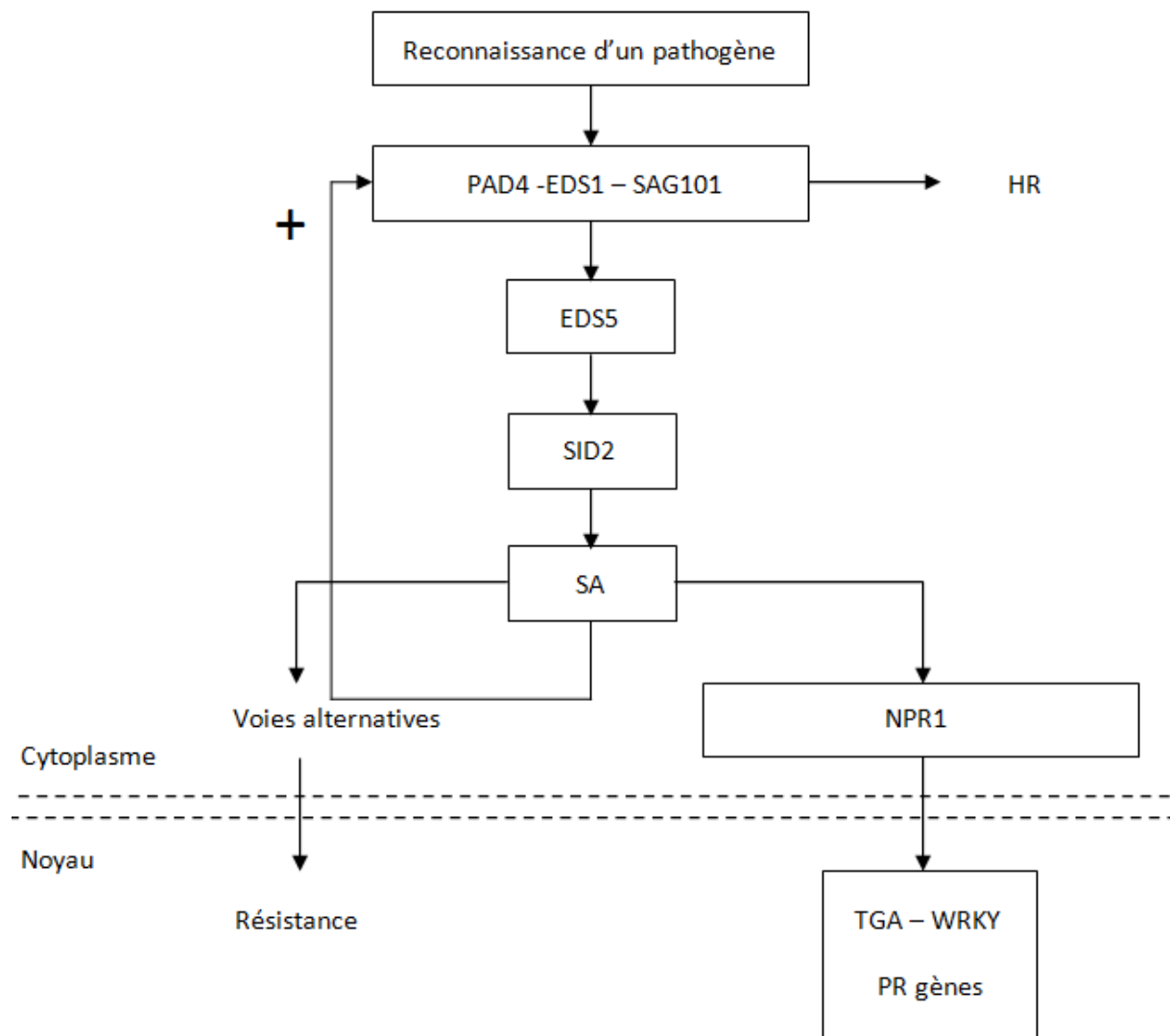


Figure 1.14 : Illustration des principaux acteurs de la voie de réponse aux pathogènes impliquant l'acide salicylique. Cette illustration ne reprend en aucun cas tous les acteurs de la réponse de défense impliquant l'acide salicylique. Les protéines encadrées sont explicitement citées dans le paragraphe relatif à la réponse à l'acide salicylique. En pointillé, la membrane nucléaire. Le « + » indique la boucle de rétro activation de l'acide salicylique sur l'expression des gènes PAD4 et EDS1. Par voies alternatives, nous entendons voies de réponses à l'acide salicylique NPR1 indépendantes. La reconnaissance du pathogène peut se faire au niveau extra et intracellulaire.

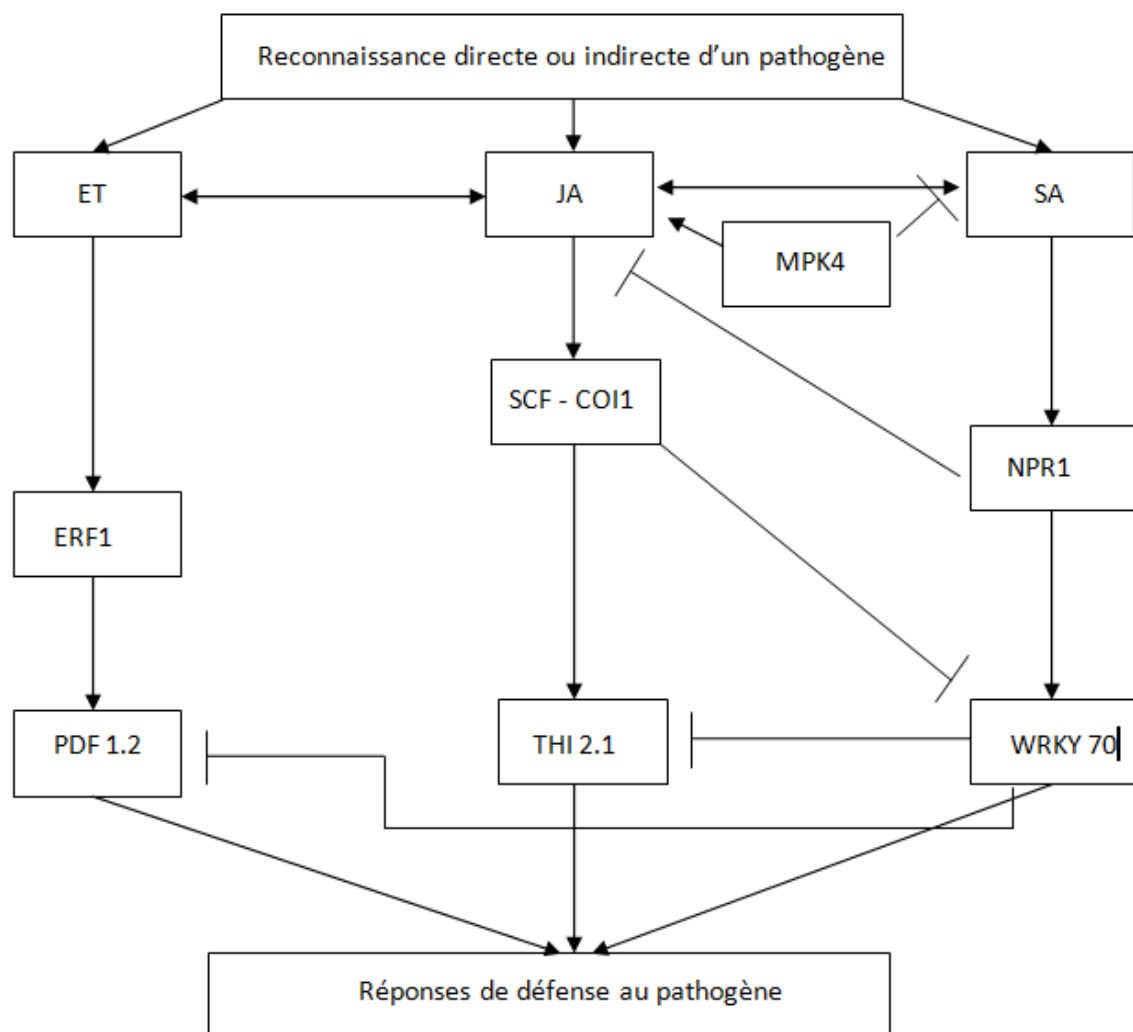


Figure 1.15 : Figure représentant les interactions entre les voies de l'acide salicylique (SA), de l'éthylène (ET) et du jasmonate (JA) suite à la détection d'un pathogène. Cette figure n'est pas représentative de l'ensemble des régulations entre les différents acteurs des différentes voies. Elle permet la visualisation graphique des interactions définies dans l'introduction. Une flèche simple = active positivement. Flèche avec barre : inhibe la voie ou la protéine ciblée. Les liens reliant entre eux les voies de l'acide salicylique (SA), de l'éthylène (ET) et du jasmonate (JA) doivent être interprétés de la manière suivante : les voies de l'éthylène et du jasmonate peuvent se stimuler mutuellement. Une des deux voies peut également inhiber l'autre. Les interactions entre la voie du jasmonate et de l'acide salicylique sont majoritairement antagonistes. Lorsqu'une voie est activée, elle inhibe l'autre.

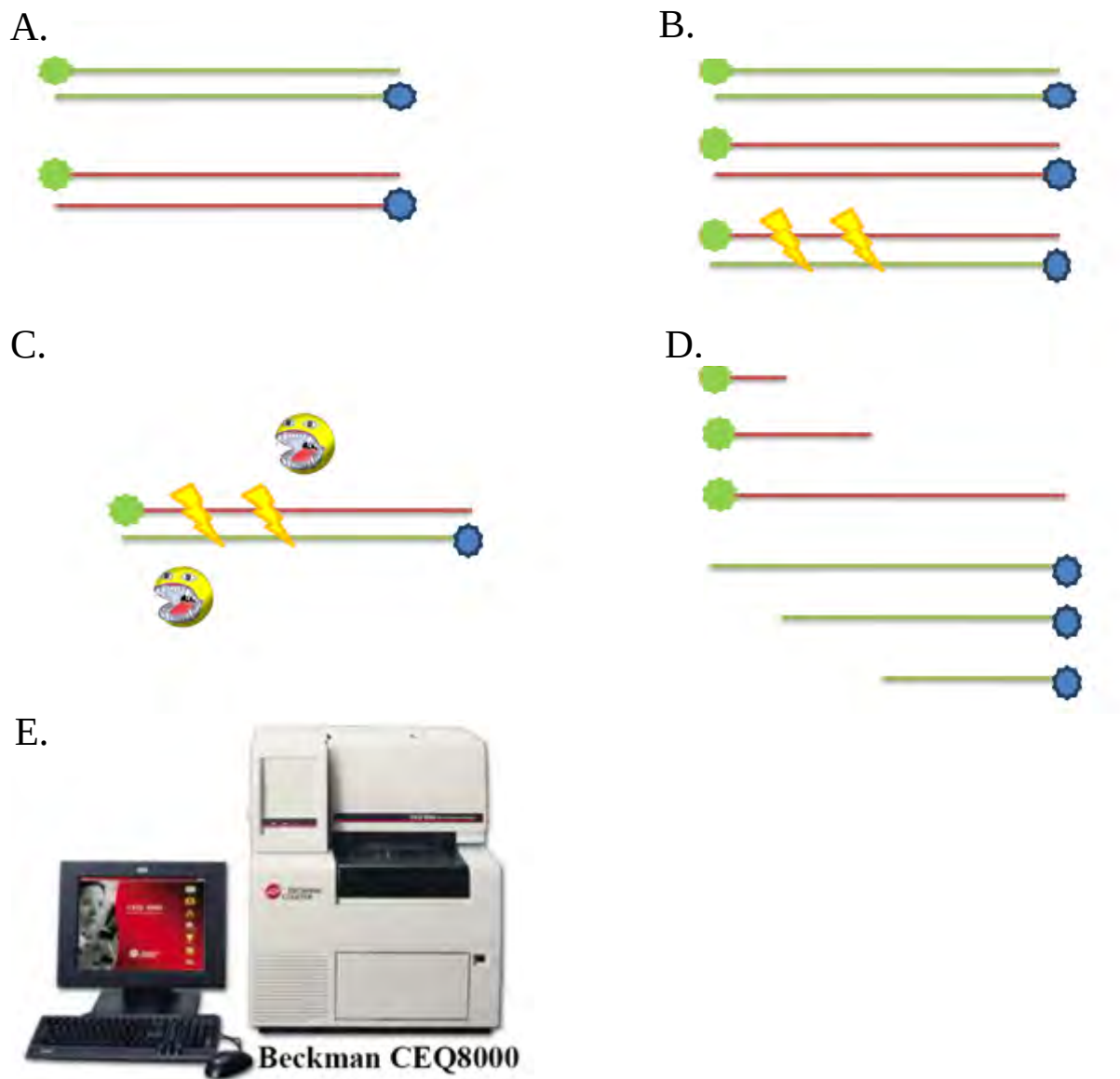
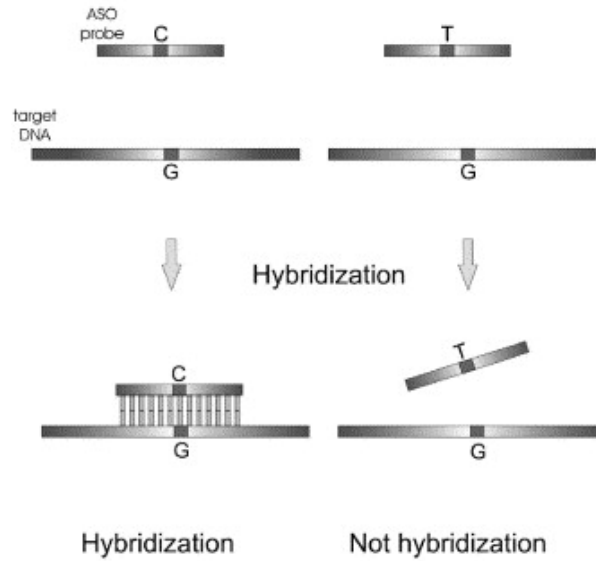
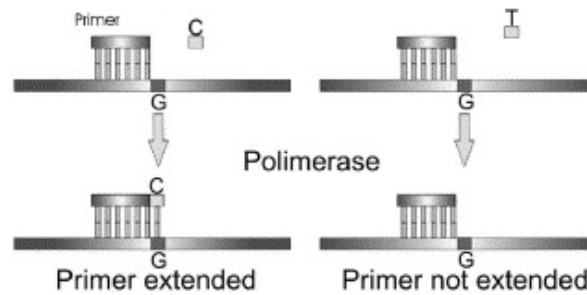


Figure 1.16: Principe de l'EcoTILLING illustré en 5 étapes. A. Amplification PCR d'un individu de référence et d'un individu à tester avec une paire d'amorces marquées (Vert et Bleu). B. Dénaturation du mélange des produits PCR marqués de l'individu de référence et de l'individu à tester puis re- hybridation. Création d'hybrides. Les SNP apparaissent sous la forme de mismatch (Eclairs). C. Ajout d'endo1, une endonucléase *single strand specific* (SSS nuclease) qui coupe à chaque mismatch soit à chaque position de SNP. D. Après digestion, nous avons plusieurs fragments de tailles différentes marqués en fonction de la position des SNP. Chaque mismatch est détecté par deux fragments de couleur différente et de taille complémentaire. E. Séparation des fragments en fonction de la taille de ces derniers à l'aide du séquenceur CEQ8000 (Beckman Coulter) en mode séparateur de fragments.

A.



B.



C.

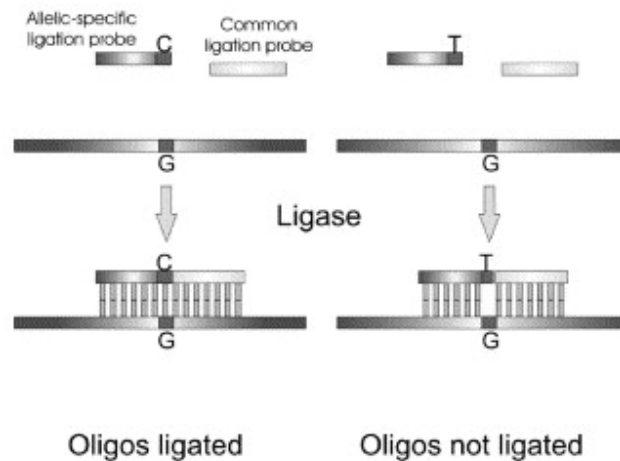
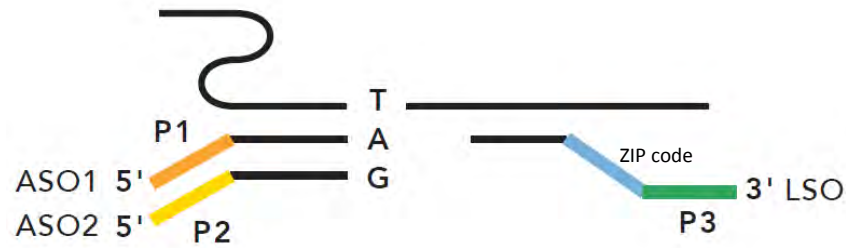
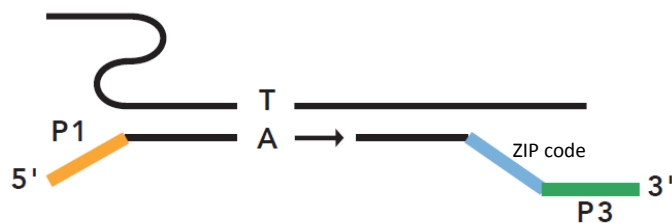


Figure 1.17 : Méthodes de base de typage de SNP : Allèle spécifique hybridation (A.) ; Extension de primers (B) ; Oligonucléotides ligations (C.). A. La sonde spécifique portant l'allèle C s'hybride parfaitement avec l'ADN cible où le SNP à typer est un G. La sonde portant l'autre allèle (T) ne s'hybride pas avec la séquence d'ADN cible. B. A gauche, l'ajout d'un nucléotide C est possible car il est complémentaire du SNP G à typer. A droite, il n'y a pas d'extension d'amorce avec l'ajout d'un nucléotide T qui n'est pas complémentaire du SNP G à typer. C. A gauche, la sonde allèle spécifique et la sonde locus spécifique s'hybride côte à côte sans mismatch. Les deux sondes sont liguées entre elles par une ligase. A droite, la sonde allèle spécifique et la sonde locus spécifique s'hybride côte à côte avec un mismatch au niveau du SNP à typer. Il n'y a pas de ligation des deux sondes. (Sobrino, Brion et al. 2005)

A.



B.



C.



D.

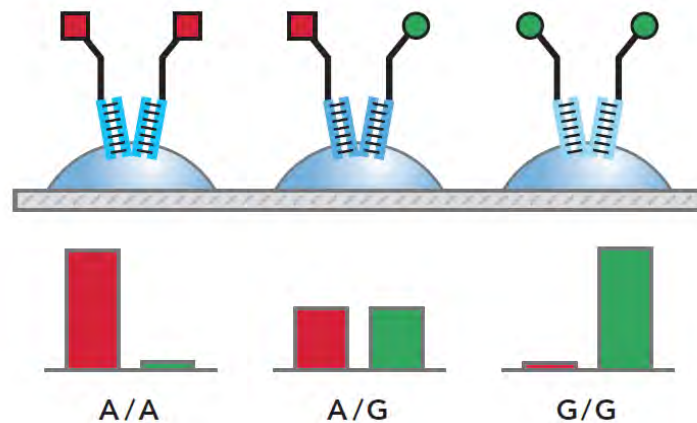


Figure 1.18: Principe technique Illumina GoldenGate Assay. A. Mix contenant la séquence à typer pour un SNP, 2 sondes allèles spécifiques (ASO) (spécifique de l'allèle T ou C) contenant à leur extrémité 5' un site d'amplification P1 ou P2, une sonde locus spécifique (LSO) contenant une séquence d'adressage pour se fixer à un endroit donné du microarray et un site d'amplification P3. B. La sonde allèle spécifique du SNP T s'hybride. Il y a extension d'amorces sur quelques nucléotides et ligation avec la sonde locus spécifique hybridée quelques nucléotides après la position du SNP à typer. C. En utilisant des amorces marquées spécifiques P1 ou P2 (marquées différemment) et l'amorce P3 non marquée, on amplifie le ou les couples de sondes amplifiées précédemment. D. Grâce à la séquence d'adressage (ZIP code) de la sonde locus spécifique, les produits PCR de chaque SNP vont se localiser à un endroit défini du microarray. La nature du SNP typé peut être observée suivant le signal. Il est possible de déterminer si nous sommes en présence d'un individu homo ou hétérozygote pour le SNP typé. Illustrations Illumina.

Original Sequence	master_id	Chr n°	Position Cm	Name	Accession unigène C.I	Expect	Pct idnt	Ratio idnt	Name unigène C.I	Accession L.S	Expect	Pct idnt	Ratio idnt	RBH
QGF3e10.yg.ab1	LE0239	1	4,571	QGF3E10.YG.AB1	CONTIG595	8,00E-90	89	259/290	CONTIG595	QGF3E10.YG.AB1	4,00E-90	89	259/290	Y
CLX_S3_Contig2780	LE0093	1	46,499	QGF11F18.YG.AB1	CONTIG9036	1,00E-146	84	572/676	CONTIG9036	QGF11F18.YG.AB1	1,00E-146	84	572/676	Y
QGF20G21.yg.ab1	LE0200	1	86,868	QGF20G21.YG.AB1	CONTIG8191	6,00E-63	88	213/240	CONTIG8191	QGF20G21.YG.AB1	3,00E-63	88	213/240	Y
QG_CA_Contig8085	LE0090	2	75,617	XLR1_CONTIG4677	CONTIG9396	0	90	778/863	CONTIG9396	XLR1_CONTIG4677	0	90	778/863	Y
AF017751.1	LR0010	2	15,381	QGJ8I16.YG.AB1	CONTIG154	1,00E-169	89	461/515	CONTIG154	QGJ8I16.YG.AB1	1,00E-169	89	461/515	Y
CLX_S3_Contig4550	LE0393	2	69,959	QGF17F10.YG.AB1	CONTIG6268	0	93	606/649	CONTIG6268	QGF17F10.YG.AB1	0	93	606/649	Y
CLX_S3_Contig7847	LE0064	2	23,674	QGF22M18.YG.AB1	RSL006_F01.GB	1,00E-110	90	316/351	RSL006_F01.GB	QGF22M18.YG.AB1	1,00E-110	90	316/351	Y
CLX_S3_Contig535	LE0317	3	82,449	XLR1_CONTIG1428	CONTIG9692	1,00E-132	93	321/345	CONTIG9692	XLR1_CONTIG1428	1,00E-132	93	321/345	Y
CLX_S3_Contig15048	LE0092	3	106,647	XLR1_CONTIG3056	CONTIG8136	2,00E-99	91	254/278	CONTIG8136	XLR1_CONTIG3056	2,00E-99	91	254/278	Y
QG_CA_Contig5295	LE0149	3	28,919	XLR1_CONTIG2508	CONTIG5419	0	86	784/907	CONTIG5419	XLR1_CONTIG2508	0	86	784/907	Y
CLX_S3_Contig7798	LE0362	3	90,096	QGE13O05.YG.AB1	CONTIG9190	1,00E-130	93	299/321	CONTIG9190	QGE13O05.YG.AB1	1,00E-130	93	299/321	Y
CLX_S3_Contig12210	LE0402	4	130,391	XLR1_CONTIG3010	CONTIG6761	5,00E-42	87	156/178	CONTIG6761	XLR1_CONTIG3010	5,00E-42	87	156/178	Y
QG_CA_Contig6979	LE0279	4	58,109	XLR1_CONTIG1452	CONTIG4582	6,00E-26	82	161/195	CONTIG4582	XLR1_CONTIG1452	8,00E-26	82	161/195	Y
QG_CA_Contig5752	M5752	5	12,231	XLR1_CONTIG2908	CONTIG4366	1,00E-137	89	368/410	CONTIG4366	XLR1_CONTIG2908	1,00E-137	89	368/410	Y
QG_CA_Contig2949	LE0091	5	66,845	QGG26K09.YG.AB1	CONTIG3680	4,00E-05	80	93/116	CONTIG3680	QGG26K09.YG.AB1	3,00E-05	80	93/116	Y
QGF5E16.yg.ab1	LE0114	5	105,959	QGF5E16.YG.AB1	CONTIG5089	7,00E-53	87	186/212	CONTIG5089	QGF5E16.YG.AB1	7,00E-53	87	186/212	Y
QG_CA_Contig6982	M6982	7	107,249	XLR1_CONTIG4058	CONTIG7246	8,00E-66	88	213/242	CONTIG7246	XLR1_CONTIG4058	2,00E-66	88	213/242	Y
CLX_S3_Contig6420	LE0310	9	68,730	XLR1_CONTIG2156	CONTIG7494	1,00E-72	91	191/209	CONTIG7494	XLR1_CONTIG2156	1,00E-72	91	191/209	Y
CLX_S3_Contig8436	M4990	9	55,464	XLR1_CONTIG2213	CONTIG6958	1,00E-147	88	433/487	CONTIG6958	XLR1_CONTIG2213	1,00E-148	88	433/487	Y
QGG19L12.yg.ab1	LE0456	9	1,789	QGG19L12.YG.AB1	CONTIG5759	1,00E-180	92	423/457	CONTIG5759	QGG19L12.YG.AB1	1,00E-180	92	423/457	Y
CLX_S3_Contig790	LE0329	9	68,876	XLR1_CONTIG3097	CONTIG2630	1,00E-110	94	240/253	CONTIG2630	XLR1_CONTIG3097	1,00E-110	94	240/253	Y

Table 2.1: 21 orthologues potentiels entre la laitue et la chicorée associés à la pathogenèse, données complètes des analyses *reciprocal best BLAST hit*. Colonnes 1 à 4: Séquences d'origine de la carte de laitue MAP2JMR3. Contenu des colonnes : original Sequence, nom d'accension de la séquence EST dans la carte de la laitue ; master_Id, nom de référence du marqueur sur la carte MAP2JMR3 ; Chr n°, numéro du chromosome où est positionné le marqueur sur la carte de laitue ; Position Cm, localisation du marqueur sur le chromosome en centimorgan ; Expect, *e-value* ; Pct idnt, pourcentage d'identité entre la séquence de laitue et la séquence de chicorée ; Ratio idnt, ratio d'identité. Colonnes 5 à 9 : BLAST séquences EST de la carte MAP2JMR3 contre banque de données unigènes chicorée. Contenu des colonnes : Name, nom de la séquence de *L. serriola* ; Accession unigènes C.I, meilleur résultat de BLAST dans la banque de données d'unigènes de *C. intybus* ; Expect, *e-value* ; Pct idnt, pourcentage d'identité entre la séquence de laitue et la séquence de chicorée ; Ratio idnt, ratio d'identité. Colonnes 10 à 14 : BLAST séquences EST chicorée contre banque d'EST de *L. serriola*. Contenu des colonnes : Name unigènes C.I, nom de la séquence dans la banque d'unigènes de *C. intybus* ; Accession L.S, meilleur résultat de BLAST dans la banque de données EST de *L. serriola*. Colonne 15, confirmation du RBH, Y= oui

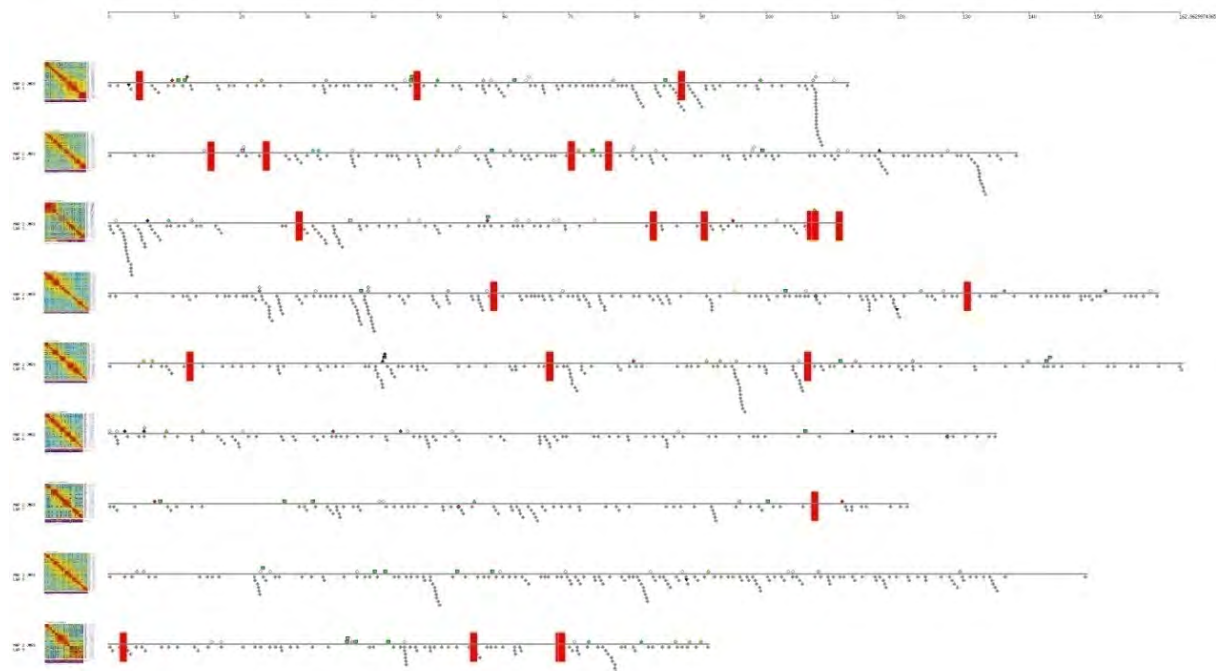


Figure 2.1 : Localisation des orthologues potentiels entre la laitue et la chicorée sur la carte MAP2JMR3. Chaque orthologue est défini par un rectangle rouge. Les 21 orthologues potentiels sont positionnés sur l'ensemble des chromosomes de la laitue à l'exception du chromosome 8.

NBS identity	AY153825.1	AY153822.1	AY153817.1	AY153815.1	AY153812.1	AY153810.1	AY153807.1
AY153825.1	100	52	49	48	49	47	49
AY153822.1	52	100	71	48	50	46	47
AY153817.1	49	71	100	48	49	48	47
AY153815.1	48	48	48	100	58	56	57
AY153812.1	49	50	49	58	100	75	82
AY153810.1	47	46	48	56	75	100	74
AY153807.1	49	47	47	57	82	74	100

Table 2.2: Identité des 7 séquences de chicorée associées à un domaine NBS après alignement. En vert, identité maximale observée entre 2 séquences différentes ; En rouge, identité minimale entre 2 séquences ; en vert clair, identité supérieure à 50% entre 2 séquences ; en rosé, identité inférieure à 50% entre 2 séquences.

Arabidopsis	eVal	Pct Idnt	Ratio Idnt	Chicory Best Hit	TIR	NB ARC	LRR	Lenght
AT3G14460.1	2,00E-29	50	37/73	08467385.VIS		yes	yes	596
AT1G27170.1	2,00E-33	55	30/54	13096385.VIS	yes	yes		823
AT5G17680.1	2,00E-44	61	27/44	15860385.VIS		yes		819
AT5G17970.1	1,00E-33	54	44/81	17666385.VIS	yes			791
AT4G04110.1	3,00E-18	41	28/67	22885385.VIS	yes			805
AT1G58400.1	5,00E-19	39	20/51	25510385.VIS		yes		832
AT4G16880.1	5,00E-06	36	27/74	26173385.VIS			yes	819
AT4G14370.1	8,00E-45	48	50/104	28440385.VIS	yes	yes		768
AT1G63350.1	4,00E-05	35	20/56	28763385.VIS			yes	842
AT1G33560.1	3,00E-39	34	55/160	28802385.VIS			yes	867
AT5G45230.1	8,00E-19	36	35/96	30486385.VIS		yes	yes	765
AT5G66910.1	7,00E-60	47	61/128	30861385.VIS			yes	765
AT5G36930.1	3,00E-47	53	37/69	31129385.VIS			yes	775
AT4G16900.1	5,00E-29	46	43/92	31299385.VIS	yes			766
AT3G46530.1	1,00E-13	30	18/59	31352385.VIS		yes		677
AT1G62630.1	2,00E-05	41	26/63	32538385.VIS			yes	751
AT5G41750.1	2,00E-32	47	32/68	36078385.VIS	yes	yes		740
AT3G46710.1	1,00E-15	42	26/61	FCL007E03.GB		yes		691
AT1G61310.1	8,00E-07	42	25/59	RLN006A01.FASTA			yes	503
				AT5G41750 ref	yes	yes	yes	4140

Table 2.3 : 19 EST de chicorée présentant un domaine associé aux protéines TIR-NBS-LRR. Contenu des colonnes : *Arabidopsis*, numéro d'accension TAIR; eVal, e-valeur du BLAST séquences *A. thaliana* contre banque unigènes de chicorée; Pct Idnt, pourcentage d'identité entre la séquence d'*Arabidopsis* et la séquence de chicorée ; Ratio idnt, ratio d'identité; Chicory Best Hit, référence de la séquence de chicorée ayant donné les meilleurs résultats de BLAST; TIR/NB ARC/LRR, présence (yes) ou absence du domaine spécifié, le domaine NB ARC est équivalent au domaine NBS (nucleotide binding site) ; Length, taille de la séquence EST de chicorée. Codes couleurs : Vert, EST contenant un domaine NB-ARC et un domaine LRR; Jaune, EST contenant un domaine TIR et un domaine NB-ARC ; Rouge, EST contenant uniquement un domaine NB-ARC ; Bleu, EST contenant un domaine TIR uniquement ; Gris, EST contenant seulement un domaine LRR ; Orange, séquence TIR-NB-ARC-LRR d'*A. thaliana* utilisée comme référence (NCBI référence: AT5G41750).

Name	Chicory accession	ORF length	Analyze mode	Ref database	Description	IPRSCAN eval	Patho Blast Related	NBS Blast result
CHICOSUN1	CIC.EH690542_6_ORF1	238	Gene3D	G3D.3.80.10.1	no description	1.1e-22	OK	Null
	CIC.EH690542_6_ORF1	238	superfamily	SSF52047	RNI-like	3.2e-26	OK	
	CIC.EH690542_6_ORF1	238	HMMPanther	PTHR11017.5	DISEASE RESISTANCE PROTEIN	8.3e-16	OK	
	CIC.EH690542_6_ORF1	238	HMMPanther	PTHR11017	LEUCINE RICH REPEAT PROTEIN	8.3e-16	OK	
CHICOSUN2	CIC.EH693729_3_ORF1	222	superfamily	SSF52058	L domain-like	2.2e-09	OK	NBS
	CIC.EH693729_3_ORF1	222	HMMPanther	PTHR11017.5	DISEASE RESISTANCE PROTEIN	7.8e-05	OK	
	CIC.EH693729_3_ORF1	222	HMMPanther	PTHR11017	LEUCINE RICH REPEAT PROTEIN	7.8e-05	OK	
CHICOSUN3	CIC.EH678163_3_ORF2	249	superfamily	SSF46785	Winged helix DNA-binding domain	8.4e-26	OK	NBS
	CIC.EH678163_3_ORF2	249	FPrintScan	PR00364	DISEASERSIST	1.1e-05	OK	
	CIC.EH678163_3_ORF2	249	FPrintScan	PR00364	DISEASERSIST	1.1e-05	OK	
CHICOSUN4	CONTIG4952_1_ORF2	368	superfamily	SSF52058	L domain-like	1.8e-27	OK	NBS
	CONTIG4952_1_ORF2	368	HMMPanther	PTHR11017.5	DISEASE RESISTANCE PROTEIN	4.9e-06	OK	
	CONTIG4952_1_ORF2	368	HMMPanther	PTHR11017	LEUCINE RICH REPEAT PROTEIN	4.9e-06	OK	
CHICOSUN5	CONTIG6761_2_ORF1	337	Gene3D	G3D.3.80.10.1	no description	1.6e-31	OK	Null
	CONTIG6761_2_ORF1	337	HMMPanther	PTHR11017.5	DISEASE RESISTANCE PROTEIN	5.5e-30	OK	
	CONTIG6761_2_ORF1	337	HMMPanther	PTHR11017	LEUCINE RICH REPEAT PROTEIN	5.5e-30	OK	
CHICOSUN6	CIC.EH672777_1_ORF1	276	HMMPFam	PF00931	NB-ARC	4.8e-09	OK	NBS
	CIC.EH672777_1_ORF1	276	FPrintScan	PR00364	DISEASERSIST	6.6e-09	OK	
	CIC.EH672777_1_ORF1	276	FPrintScan	PR00364	DISEASERSIST	6.6e-09	OK	

Table 2.4: 6 gènes de résistance mis en évidence sur base de séquences NBS-LRR du tournesol, résultat Interproscan et BLAST NCBI. Contenu des colonnes : Name, nom de référence du gène candidat dans la liste de gènes candidats ; Chicory accession, référence de la séquence de chicorée associé au gène de résistance. L'ORF où se trouve le domaine trouvé par l'analyse interproscan est ajoutée à la fin de la référence ; ORF length, taille de l'ORF où se trouve le domaine de résistance ; Analyze mode, base de données où un domaine a été mis en évidence ; Ref database, référence du domaine trouvé dans la séquence de chicorée ; Description, description détaillée du domaine de résistance trouvé ; IPRscan e-value, *e-value* pour le domaine trouvé dans la séquence de chicorée ; Patho Blast related, confirmation de la présence d'une homologie avec une protéine de résistance lorsque la séquence de chicorée est blastée (TBLASTX) contre la banque de données de NCBI ; NBS BLAST result, montre si un domaine NBS est trouvé parmi les meilleurs résultats de BLAST.

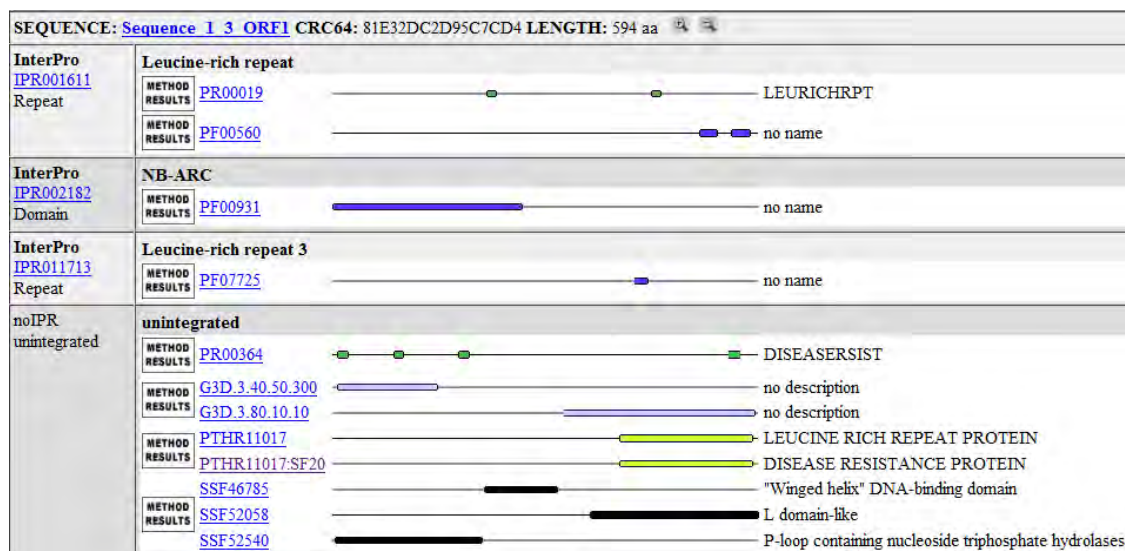


Figure 2.2: Résultats de l'analyse Interproscan du contig de 1786 pb formé par l'alignement des séquences CONTIG1413, 26173385.VIS, CIC.EH694833 et 1586035.VIS. Mise en évidence du domaine NB-ARC (ou NBS-LRR) suivi du domaine LRR ainsi que la présence de domaines associés à des résistances.

Global view: Results of EST chicory unigenes database datamining

Analyzed sequences	Databases	Databases results
5000	HMMPanther	2938
ORF total with description	HMMPFam	2062
3361	superfamily	1746
Total descriptions	Gene3D	1415
12792	FPrintScan	1150
Average length of analyzed ORF (aa)	ProfileScan	781
166	Seg	683
Candidate genes pathogenesis founded	HMMSmart	604
37	ScanRegExp	574
	BlastProDom	410
	Coil	294
	HMMTigr	110
	HMMPIR	25
	Number of databases	Total descriptions
	13	12792

Table 2.5 : Résultats globaux de l'analyse Interproscan des 5000 premières séquences EST de la banque d'unigènes de *C. intybus*. Contenu des colonnes : Colonne n°1 : total des séquences analysées, nombre d'ORF ayant été décrites, nombre total de descriptions obtenues, longueur moyenne des ORF annotés, nombre total de nouveaux gènes candidats trouvés par data mining. Colonne n° 2: Banques de données analysées. Colonne n°3 : nombre de descriptions obtenues par banques de données interrogées.

Most frequent Interproscan descriptions

IPS descriptions	Fréquence
Calcium-binding EF-hand	236
WD40 repeat	163
Cytochrome P450	129
Protein kinase-like	112
Pentatricopeptide repeat	103
Leucine-rich repeat	101
Protein kinase, core	86
Tyrosine protein kinase	80
Transcription factor, MADS-box	63
Pathogenesis-related transcriptional factor and ERF, DNA-binding	55
Zinc finger, RING-type	53
Cytochrome P450, E-class, group I	53
Haem peroxidase, plant/fungal/bacterial	49
Heat shock protein DnaJ, N-terminal	49
Ras GTPase	47
RNA recognition motif, RNP-1	46
Major intrinsic protein	46
Serine/threonine protein kinase-related	45
Aldo/keto reductase	44
Ankyrin	44
Zinc finger, CCHC-type	43
Plant peroxidase	42
Serine/threonine protein kinase	42
Serine/threonine protein kinase, active site	41
Histone H3	40
Zinc finger, C2H2-type	39
Ubiquitin	38
NAD(P)-binding	37
Homeodomain-like	37
Mitochondrial substrate carrier	35
Short-chain dehydrogenase/reductase SDR	34
C2 calcium-dependent membrane targeting	34
Cyclin-like F-box	33
Glucose/ribitol dehydrogenase	29
Myb, DNA-binding	28
Nucleoside diphosphate kinase, core	28
Basic helix-loop-helix dimerisation region bHLH	27
Pectate lyase/Amb allergen	27
Thioredoxin-like fold	26
Peptidase S10, serine carboxypeptidase	26
Plant lipid transfer protein/Par allergen	24
Armadiillo-type fold	24
Tetratricopeptide region	24
Thioredoxin-like subdomain	23
Mitochondrial transcription termination factor-related	22
Plant lipid transfer protein/seed storage/trypsin-alpha amylase inhibitor	21
Cytochrome b5	21
Armadiillo	21
MYB-like	20
Leucine-rich repeat, typical subtype	20
Synaptobrevin	20
Autres	4019

Table 2.6 : Domaines les plus fréquents trouvés dans l'annotation systématique des 5000 premières séquences EST de la banque d'unigènes de chicorée. Première colonne : Descriptions Interproscan. Deuxième colonne : fréquence de description du domaine Inteproscan. Couleur : Jaune = facteur de transcription, Olive = liens avec Calcium, Bleu= liens protéines kinase, Rouge= facteur de transcription associé à la pathogenèse, Gris= domaines LRR

Reference Name	Unigene accession	Methods	Accession	e-value	Interpro ref	Interpro description	NCBI blast description
CHICOSCAN1	>Cich_inty.EH709933	HMMPfam	PF03106	5.8e-39	IPR003657	DNA-binding WRKY	WRKY
CHICOSCAN2	>Cich_inty.EH709264	HMMPfam	PF03018	1.3e-50	IPR004265	Plant disease resistance response protein	DISEASE RESISTANCE
CHICOSCAN3	>Cich_inty.EH709073	HMMPfam	PF03094	2.1e-11	IPR004326	Mlo-related protein	MILDEW RESISTANCE
CHICOSCAN4	>Cich_inty.EH708104	HMMPfam	PF03094	1.2e-23	IPR004326	Mlo-related protein	MILDEW RESISTANCE
CHICOSCAN5	>Cich_inty.EH706551	HMMSmart	SM00380	2.4e-28	IPR001471	Pathogenesis-related transcriptional factor and ERF, DNA-binding	"AVR-9"
CHICOSCAN6	>Cich_inty.EH703640	HMMPanther	PTHR10264.S	7.00E-32	NULL	NULL	HR
CHICOSCAN7	>Cich_inty.EH703550	HMMPanther	PTHR10334.S	7.8e-50	NULL	NULL	PR1
CHICOSCAN8	>Cich_inty.EH701710	HMMSmart	SM00380	4.7e-26	IPR001471	Pathogenesis-related transcriptional factor and ERF, DNA-binding	TRANSCRIPTION FACTOR AP2 AVR-9
CHICOSCAN9	>Cich_inty.EH701125	HMMPfam	PF03094	2.3e-13	IPR004326	Mlo-related protein	MLO MILDEW RESISTANCE
CHICOSCAN10	>Cich_inty.EH700709	HMMPfam	PF03018	5.7e-65	IPR004265	Plant disease resistance response protein	DISEASE RESISTANCE
CHICOSCAN11	>Cich_inty.EH700393	HMMPfam	PF00847	6.7e-05	IPR001471	Pathogenesis-related transcriptional factor and ERF, DNA-binding	ET RESPONSE
CHICOSCAN12	>Cich_inty.EH699232	HMMPanther	PTHR11017.S	4.9e-08	NULL	NULL	DISEASE RESISTANCE
CHICOSCAN13	>Cich_inty.EH695658	HMMPfam	PF00931	4.2e-10	IPR002182	NB-ARC	NBS-LRR
CHICOSCAN14	>Cich_inty.EH695155	HMMSmart	SM00380	3.3e-36	IPR001471	Pathogenesis-related transcriptional factor and ERF, DNA-binding	ET RESPONSE
CHICOSCAN15	>Cich_inty.EH693800	HMMPfam	PF03106	4.6e-22	IPR003657	DNA-binding WRKY	WRKY
CHICOSCAN16	>Cich_inty.EH693796	HMMPfam	PF03094	3.1e-29	IPR004326	Mlo-related protein	MILDEW RESISTANCE
CHICOSCAN17	>Cich_inty.EH692411	HMMPfam	PF00847	7.1e-41	IPR001471	Pathogenesis-related transcriptional factor and ERF, DNA-binding	ET RESPONSE
CHICOSCAN18	>Cich_inty.EH691945	HMMPfam	PF00847	2.6e-39	IPR001471	Pathogenesis-related transcriptional factor and ERF, DNA-binding	WOUND-RESPONSE, AP2 like factor

Table 2.7 : 18 gènes candidats associés à la résistance aux pathogènes trouvés par *datamining* de la banque d'EST unigènes de chicorée annotée pour les 5000 premières séquences. Contenu des colonnes : Reference Name, référence du gène candidat ; Unigene accession, référence de la séquence dans la banque d'unigènes de chicorée ; Methods, banque de données ayant permis la mise en évidence du domaine associé à la pathogenèse ; Accession, référence du domaine protéique trouvé dans la banque de données spécifiée précédemment ; e-value, e-value de la prédiction du domaine dans la séquence de chicorée analysée ; Interpro ref, référence du domaine standardisé Interproscan ; Interpro description, description du domaine Interproscan ; NCBI blast description, description de la séquence de chicorée donné par analyse TBLASTX contre la banque de données NCBI.

Davis accession	Hit A. thaliana	Blast description	e-value	NCBI best blast result associated with pathogene	Blast score	e-value
CI.CSA1.4062	AT5G51700	RAR1 disease resistance protein (RAR1), Contains a 3 nt micro-exon at exon 2	1,00E-75	Nicotiana tabacum Rar1 (Rar1) mRNA, comp	232	4,00E-105
CI.EH679530	AT1G33590	disease resistance protein-related / LRR protein-related, contains leucine rich-repeat domains Pfam:	1,00E-68	Arabidopsis thaliana disease resistanc	174	4,00E-64
CI.EH675505	AT2G34930	disease resistance family protein, contains leucine rich-repeat domains Pfam:	1,00E-34	Malus floribunda hcrVf2 gene	131	5,00E-53
CI.EH689532	AT3G14460	disease resistance protein (NBS-LRR class), putative, domain signature NBS-LR	1,00E-13	Helianthus annuus clone Ha-NTIR11g CC-NB	117	5,00E-53
CI.EH704725	AT4G04220	disease resistance family protein, contains leucine rich-repeat (LRR) domains	1,00E-55	Arabidopsis thaliana disease resistanc	166	2,00E-47
CI.EH672964	AT3G14460	disease resistance protein (NBS-LRR class), putative, domain signature NBS-LR	1,00E-11	Helianthus annuus clone Ha-NTIR3B CC-NBS	100	4,00E-47
CI.EH678027	AT4G33000	disease resistance protein (CC-NBS-LRR class), putative, domain signature CC-L	1,00E-51	Arabidopsis thaliana ADR1-L1 (ADR1-LIK	182	4,00E-43
CI.EH692382	AT3G11010	disease resistance family protein / LRR family protein, contains leucine rich-re	1,00E-46	Arabidopsis thaliana leucine-rich repe	133	2,00E-42
CI.EH678903	AT1G61100	disease resistance protein (TIR class), putative, domain signature TIR exists, su	1,00E-36	Arabidopsis thaliana disease resistanc	116	3,00E-37
CI.DT211429	AT3G48090	disease resistance protein (EDS1), identical to disease resistance protein/lip	1,00E-26	Solanum tuberosum enhanced disease susce	158	1,00E-36
CI.EH700984	AT3G24900	disease resistance family protein / LRR family protein, contains leucine rich-re	1,00E-02	Glycine max RFLP clone A 45-10 s	102	1,00E-36
CI.EH706298	AT5G42510	disease resistance-responsive family protein, similar to disease resistance resp	1,00E-32	Arachis hypogaea disease resistance-resp	157	1,00E-35
CI.EH676855	AT5G48770	disease resistance protein (TIR-NBS-LRR class), putative, domain signature TIR	1,00E-36	Populus trichocarpa clone 124140 TIR-NBS	155	4,00E-35
CI.EH698573	AT3G25020	disease resistance family protein, contains leucine rich-repeat (LRR) domains	1,00E-22	Glycine max disease resistance-r	69	8,00E-35
CI.EH694197	AT2G34930	disease resistance family protein, contains leucine rich-repeat domains Pfam:	1,00E-28	Malus floribunda hcrVf3 gene	126	2,00E-31
CI.CSA1.1307	AT3G48090	disease resistance protein (EDS1), identical to disease resistance protein/lip	1,00E-31	Nicotiana benthamiana EDS1-like protein	128	4,00E-31
CI.EH684821	AT5G36930	disease resistance protein (TIR-NBS-LRR class), putative, domain signature TIR	1,00E-23	Helianthus annuus cultivar SMF3 NBS-LRR	115	1,00E-29
CI.EH677776	AT3G48090	disease resistance protein (EDS1), identical to disease resistance protein/lip	1,00E-26	Solanum tuberosum enhanced disease susce	82	2,00E-27
CI.EH698988	AT5G49140	disease resistance protein (TIR-NBS-LRR class), putative, domain signature TIR	1,00E-17	Solanum tuberosum nematode resistance-li	120	1,00E-25
CI.EH706131	AT5G27060	disease resistance family protein, contains leucine rich-repeat (LRR) domains	1,00E-26	Lycopersicon pimpinellifolium Cf-2.3 gen	94	2,00E-25
CI.EH677081	AT5G36930	disease resistance protein (TIR-NBS-LRR class), putative, domain signature TIR	1,00E-25	Populus trichocarpa clone 65744 TIR-NBS	73	2,00E-24
CI.EH679541	AT5G36930	disease resistance protein (TIR-NBS-LRR class), putative, domain signature TIR	1,00E-31	Populus trichocarpa isolate S-117-1 TIR-NBS	97	7,00E-24
CI.EH688475	AT3G14470	disease resistance protein (NBS-LRR class), putative, domain signature NBS-LR	1,00E-18	Solanum demissum isolate 752C15a disease	79	1,00E-20
CI.EH673977	AT2G34930	disease resistance family protein, contains leucine rich-repeat domains Pfam:	1,00E-17	Arabidopsis thaliana putative disease	57	3,00E-18
CI.EH703858	AT3G08040	MATE efflux family protein, low similarity to enhanced disease susceptibility 5	1,00E-26	Arabidopsis thaliana MATE efflux famil	97	1,00E-17
CI.EH672557	AT5G66900	disease resistance protein (CC-NBS-LRR class), putative, domain signature CC-L	1,00E-21	Medicago sativa disease resistance prote	95	1,00E-16
CI.EH687353	AT4G09430	disease resistance protein (TIR-NBS-LRR class), putative, domain signature TIR	1,00E-13	Solanum tuberosum nematode resistance-li	65	4,00E-15
CI.EH682281	AT1G27170	disease resistance protein (TIR-NBS-LRR class), putative, domain signature TIR	1,00E-13	Arabidopsis lyrata disease resistance pr	60	4,00E-14
CI.EH701619	AT5G17680	disease resistance protein (TIR-NBS-LRR class), putative, domain signature TIR	1,00E-14	Medicago sativa cultivar Arc TIR-NBS-LRR	75	1,00E-13
CI.EH676704	AT5G66910	disease resistance protein (CC-NBS-LRR class), putative, domain signature CC-L	1,00E-27	Arabidopsis thaliana disease resistanc	46	2,00E-12
CI.CSA1.2020	AT5G66910	disease resistance protein (CC-NBS-LRR class), putative, domain signature CC-L	1,00E-13	Arabidopsis thaliana disease resistanc	46	6,00E-12
CI.EH697087	AT3G14470	disease resistance protein (NBS-LRR class), putative, domain signature NBS-LR	1,00E-11	Solanum tuberosum potato resistance-like	41	5,00E-11
CI.EH681460	AT1G47890	disease resistance family protein, contains leucine rich-repeat domains Pfam:	1,00E-13	Arabidopsis thaliana disease resistanc	68	3,00E-09
CI.EH678070	AT4G19500	disease resistance protein (TIR-NBS-LRR class), putative, domain signature TIR	1,00E-15	Arabidopsis thaliana disease resistanc	40	1,00E-07
CI.EH681763	AT1G27180	disease resistance protein (TIR-NBS-LRR class), putative, domain signature TIR	1,00E-24	Arabidopsis lyrata disease resistance pr	44	5,00E-06
CI.EH706408	AT5G45050	disease resistance protein-related, similar to NL27 (Solanum tuberosum) GI:33	1,00E-11	Helianthus annuus cultivar SMF3 NBS-LRR	51	1,00E-04
CI.EH699035	AT1G27170	disease resistance protein (TIR-NBS-LRR class), putative, domain signature TIR	1,00E-13	Arabidopsis lyrata disease resistance pr	45	4,00E-03

Table 2.8: Liste de 37 gènes candidats sur base de la banque d'EST du CGPDB annotée. Contenu des colonnes : Davis accession, référence de la séquence EST de *C. intybus* dans la banque d'unigènes de Davis ; Hit A. thaliana, meilleur de résultat de BLAST de la séquence EST de chicorée contre la banque de donnée d'*A. thaliana* ; Blast description, description de la séquence d'*A. thaliana* ; e-value ; NCBI best BLAST result associated with pathogenesis, meilleur résultat de BLAST associé à la résistance aux pathogènes lors de l'analyse TBLASTX de la séquence de chicorée de Davis contre la banque de données de NCBI ; BLAST score, score de BLAST ; e-value.

Key gene	A. thaliana ref	Chicory sequence	BLAST score	e-value
JAR1	AT2G46370	CONTIG7082	252	1,00E-110
MAPK4	AT4G01370	CONTIG5419	804	0
COI1	AT2G39940	CONTIG3063	347	1,00E-100
JIN1	AT1G32640	CONTIG9038	183	3,00E-80
NPR1	AT1G64280	CIC.EH689598	265	3,00E-89

Table 2.9 : 5 séquences de *C. Intybus* présentant une forte homologie avec un gène clef des voies de résistance d'*A. thaliana*. Contenu des colonnes : key gene, nom du gène clef recherché ; *A. thaliana* ref, numéro d'accèsion du gène clef d'*A. thaliana*. Chicory sequence, nom de la séquence de chicorée présentant la meilleure homologie avec les séquences d'*A. thaliana* testées (TBLASTX) ; BLAST score ; e-value.

Key gene	Chicory sequence	Best BLAST result description	BLAST score	e-value
JAR1	CONTIG7082	Arabidopsis thaliana JAR1 (JASMONATE RESISTA	252	7,00E-99
MAPK4	CONTIG5419	Petroselinum crispum mitogen-activated protein kinase 4 (MPK4)	852	0
COI1	CONTIG3063	Nicotiana tabacum coronatine-insensitive 1	358	7,00E-103
JIN1	CONTIG9038	Solanum tuberosum partial mRNA for MYC transcription factor	184	2,00E-102
NPR1	CIC.EH689598	Nicotiana tabacum NPR1 mRNA, complete cds	473	1,00E-130

Table 2.10 : 5 séquences de *C. intybus* présentant une forte homologie avec un gène clef des voies de résistance, TBLASTX NCBI. Contenu des colonnes : Key gène, gène clef recherché ; Chicory sequence, référence de la séquence candidate pour être homologue d'un gène clef, Best BLAST result description, meilleur résultat de TBLASTX de la séquence de chicorée contre la banque de données de NCBI ; BLAST score ; e-value.

SEQUENCE: CONTIG6745_2_ORF3 CRC64: AA996CC6DB5CC8D9 LENGTH: 361 aa			
InterPro IPR000719 Domain	Protein kinase, core		
	METHOD RESULTS	PD000001	Q943E4_ORYSA_Q943E4;
	METHOD RESULTS	PS50011	PROTEIN_KINASE_DOM
InterPro IPR001245 Domain	Tyrosine protein kinase		
	METHOD RESULTS	PR00109	TYRKINASE
InterPro IPR002290 Domain	Serine/threonine protein kinase		
	METHOD RESULTS	SM00220	no description
InterPro IPR008271 Active_site	Serine/threonine protein kinase, active site		
	METHOD RESULTS	PS00108	PROTEIN_KINASE_ST
InterPro IPR011009 Domain	Protein kinase-like		
	METHOD RESULTS	SSF56112	Protein kinase-like (PK-like)
InterPro IPR017442 Domain	Serine/threonine protein kinase-related		
	METHOD RESULTS	PF00069	no name
noIPR unintegrated	unintegrated		
	METHOD RESULTS	G3D.1.10.510.10	no description
	METHOD RESULTS	PTHR11795	PROTEIN KINASE
	METHOD RESULTS	PTHR11795:SF116	WALL-ASSOCIATED KINASE-RELATED

Figure 2.3 : Analyse Interproscan de la séquence CONTIG674, mise en évidence d'un domaine WAK-related.

References Davis	Ara best hit	Ara best hit description	eval ara	NCBI Best Wak Hit	Eval	Score
Cich_inty.EH710675	AT1G21270	wall-associated kinase 2 (W	0.31	Arabidopsis thaliana RFO1 (RESISTANCE TO FUSARIUM OXYSPORUM 1), similar to WAKL	8,00E-31	141
Cich_inty.EH691833	AT1G16130	wall-associated kinase, put	0.76	Triticum aestivum wall-associated kinase 1 (WAK1)	1,00E-84	320
Cich_inty.DT211509	AT1G21210	wall-associated kinase 4 [A	0.11	Arabidopsis thaliana WAK4	1,00E-07	62,2
Cich_inty.EH685165	AT1G21230	wall-associated kinase, put	0.22	No	No	No

Table 2.11: 4 séquences de chicorée annotées présentant une homologie avec *A. thaliana*, résultat de TBLASTX contre la banque de données de NCBI. Contenu des colonnes : References Davis, référence de l'EST de chicorée dans la banque de Davis ; Ara best hit, meilleur résultat de BLAST de la séquence de chicorée contre la banque d'*A. thaliana* ; Ara best hit description, description du meilleur résultat de BLAST ; e-val ara, e-value du résultat de BLAST Davis contre *A. thaliana* ; NCBI Best Wak Hit, meilleur résultat de TBLASTX contre la banque de données de NCBI associé avec une protéine WAK ; e-val ; Score, score de BLAST.

(Table insérée)

Table 2.12 : Aperçu des gènes candidats mis en évidence sur base des recherches bioinformatiques. Contenu des colonnes : Ref, nom de référence de l'amorce pour amplifier le gène candidat. Les gènes ne disposant pas encore d'amorces sont désignés par « WAIT » ; Name, nom de l'amorce ou du gène candidat. L'amorce contient le nom de référence du gène candidat suivi du site de restriction ajouté à celle-ci. Dans le cas d'un gène où les amorces n'ont pas encore été dessinées, seul le nom de référence du gène candidat est indiqué ; Target, séquence EST ciblée par l'amorce suivie dans la majorité des cas par une brève description ; ADNc, taille amplifiée par la paire d'amorces sur ADNc ; ADNg, taille du produit d'amplification sur ADNg ; PCR conditions, conditions PCR validées ; PCR, PCR réalisée avec succès avec les amorces dessinées. Note : le tableau ci-dessus ne reprend pas les 44 gènes candidats mis en évidence par *data-mining* de la banque de DAVIS mais peuvent être trouvés dans le tableau 2.8 (aucune amorce n'a encore été dessinée pour ces séquences).

Gène	Taille	Analyse POOL	Nombre de SNP prédits	Fréquence	Analyse 1 Contre 1	Nombre de SNP prédits	Fréquence
CS058	467	OUI	11	42	NON		
CS060	962	OUI	17	57	NON		
CS062	477	OUI	15	32	NON		
CS076	965	OUI	19	51	NON		
CS078	558	OUI	13	43	NON		
CS160	495	OUI	11	45	NON		
CS230	537	OUI	30	18	OUI	13	41
CS242	525	NON			OUI	4	131
CS284	603	NON			OUI	13	46
CS286	617	NON			OUI	7	88

Table 2.16 : Tableau récapitulatif des analyses EcoTilling en pool et avec référence pour les 10 gènes candidats analysés. Contenu des colonnes : Gènes, référence du gène (amorce *forward*) ; Taille, taille observée incluant l'insert bordé des régions T3 et T7 ; Analyse POOL, analyse de la séquence en pool ; Nombre de SNP prédits, nombre de SNP prédits par l'analyse EcoTilling en POOL ; Fréquence, fréquence de SNP observés (Taille/Nb de SNP prédits) ; Analyse 1 Contre 1, EcoTILLING avec référence ; Nombre de SNP prédits, nombre de SNP détectés lors de l'analyse 1 contre 1 ; Fréquence, fréquence de SNP observés (Taille/Nb de SNP prédits).

	a	b	c	d
A	1	5	9	13
B	2	6	10	14
C	3	7	11	15
D	4	8	12	16

Figure 2.4 : Illustration du problème de faux positifs lors de l'analyse en POOL. Dans l'exemple illustré, les colonnes et les lignes sont poolées par 4. Les clones sont représentés par un chiffre (de 1 à 16), les pools de ligne (contenant 4 clones) sont représentés par des lettres majuscules, les pools de colonne (contenant 4 clones) sont représentés par des lettres minuscules. Les clones 2, 4, 9 et 15 sont des SNP « fréquents ». Lors de l'analyse en pool, le croisement de la ligne « A » et de la colonne « a » indique chacun la présence d'un SNP. On en déduit donc la présence d'un SNP en position « aA » soit pour le clone 1. Il s'agit d'un faux positif car le clone 1 ne contient pas le SNP porté par les clones 2, 4, 9 et 13.

CS230 Pool								
Clone n°	8	19	27	30	38	49	56	62
Clone pos	a8	c3	d3	d6	e6	g1	g8	h6
SNP								
57						57		
70			70	70		70	70	
78								
88			88	88		88		
95		95						
100	100	100	100	100	100	100	100	100
113	113	113	113	113	113			
129		129				129		
133	133							
144				144				
152								152
166								
172	172	172	172	172	172	172	172	172
178		178	178	178				
186								
189								
200				200				
218				218				218
228								
245							245	
274								
290								290
299								
323					323			
329								
349								
361							361	
401								
448								
455							455	

Table 2.17 : SNP détectés par analyse sur POOL pour le gène CS230 (NBS-LRR), sélection de 8 clones représentant le polymorphisme de la core collection. Contenu du tableau : Clone n°, clone sélectionné pour le séquençage sur base de son polymorphisme dans la core collection ; Clone pos, position du clone dans la plaque de screening ; SNP, position des SNP sur la séquence à partir de l'amorce T3 marquée. La première colonne reprend l'ensemble des SNP détectés lors de l'analyse en POOL. En bleu clair, les SNP présents dans les clones sélectionnés pour le séquençage. En bleu foncé, position des SNP prédits dans les clones sélectionnés pour le séquençage.

CS230 ref												
Clone n°	CS230 1-1	CS230 1-6	CS230 1-8	CS230 1-19	CS230 1-27	CS230 1-29	CS230 1-30	CS230 1-42	CS230 1-49	CS230 1-56	CS230 1-62	Fréquence
94						94		94	94			7,8
100		100			100						100	36
171		171										3,5
178						178						3,5
189						189	189	189	189			13
245						245						6,25

Table 2.18 : Sélection de résultat EcoTILLING avec référence (CS230 clone 1) pour 11 clones, fréquence de SNP (CS230, NBS-LRR). Description du tableau : Clone n°, ligne reprenant le gène analysé (CS230, le clone de référence (CS230-1) et le clone analysé ; En bleu clair, position des SNP prédits pour l'ensemble des clones sélectionnés. La position du SNP contient la région amplifiée T3 (50pb) ; En bleu foncé, position des SNP prédits par clone ; Fréquence, fréquence du SNP prédit dans la core collection sur base de l'analyse individuelle de 64 clones.

CS230 Seq									
CS230-1	CS230-6	CS230-8	CS230-19	CS230-27	CS230-30	CS230-42	CS230-49	CS230-56	CS230-62
			46						
48									
	51			51					51
			79						
	121			121					121
140									
							343		

Table 2.19 : Résultats de séquençage pour 10 clones séquencés sur base des résultats d'EcoTILLING en pool et avec référence (CS230, NBS-LRR). Description du tableau : En jaune, numéro du clone séquencé ; En bleu, position du SNP détecté par séquençage.

SNP POOL	7	20	28	38	45	50		63	79	83	94	102	116	122	128	136	139	150	168	178	195	224	240	249	273	279	299	311		351	398	405	
SNP REF					44	50								121	128		139				195												
SNP SEQ					46	48	51		79					121			140													343			

Table 2.20 : Tableau comparatif des SNP prédits par EcoTilling en pool et avec référence (clone CS230-1) par rapport aux SNP réels. Première ligne, Position des SNP prédits à partir de l'amorce reverse (Les SNP confirmés par séquençage sont en vert foncé) ; Deuxième ligne, position des SNP prédits à partir de l'amorce reverse (Les SNP confirmés par séquençage sont en vert) ; Position des SNP mis en évidence par séquençage (Les SNP retrouvés par les deux analyses EcoTilling sont en vert clair).

CS058					
Clone 1	Clone 15	Clone 25	Clone 41	Clone 45	Changement aa
				46	
	54			54	A/V
		70	70		
		103	103		I/M
	142			142	
			148		N/K
		166	166		
172					
	192			192	G/D
	194	194		194	D/H
		199			
		202	202		
		205	205		
		208	208		D/E
		211	211		
		226	226		
		231	231		T/K
241					

Table 2.21 : Résultats de séquençage de 5 clones pour le gène CS058 (β 1,3 glucanase), modifications d'acides aminés. La position de chaque SNP est calculée à partir de l'amorce *forward*. En Jaune, n° du clone séquencé ; En vert, SNP caractérisant le paralogue 1 ; En Bleu, SNP caractérisant le paralogue 2 ; En orange, SNP caractérisant le paralogue 3. La dernière colonne indique les modifications d'acides aminés induites par les SNP. En rouge, changement radical des propriétés de l'acide aminé modifié par le SNP.

CS058-25 bet	AAATCTCCGTACGCGTTGTTTACGGCACCGGGACCGTTGCGACTGATAATAATACGGTCTCAATACTCTAACCTTTTGACGCCATGTTAGATGCTCACTATGCACTCAAGCAGCTCTTGGTGGCAAAATTTGGA
CS058-1 bet	AAATCTCCGTACGCGTTGTTTACGGCACCGGGACCGTTGCGACTGATAATAATATGCTCTCAATACTCTAACCTTTTGACGCCATATTAGATGCTCACTATGCACTCAAGCAGCTCTTGGTGGCAAAATTTGGA
CS058-15 bet	AAATCTCCGTACGCGTTGTTTACGGCACCGGGACCGTTGCGACTGATAATAATATGCTCTCAATACTCTAACCTTTTGACGCCATATTAGATGCTCACTATGCACTCAAGCAGCTCTTGGTGGCAAAATTTGGA
1-Pat-Ci-LE0093	AAATCTCCGTATGCAATTTGTTTACGGCACCGGGACCGTTGCGACTGATAATAATATGCTCTCAATACTCTAACCTTTTGACGCCATTTAGATGCTCACTATGCACTCAAGCAGCTCTTGGTGGCAAAATTTGGA
AJ011769.1	AAATCTGTCGTATGCAATTTGTTTACGGCACCGGGACCGTTGCGACTGATAATAATATGCTCTCAATACTCTAACCTTTTGACGCCATTTAGATGCTCACTATGCACTCAAGCAGCTCTTGGTGGCAAAATTTGGA
AJ249292.1 cg2	AAATCTCCATATGCGTTGTTTACGGCACCGGGACCGTTGCGACTGATAATAATATGCTCTCAATACTCTAACCTTTTGACGCCATTTAGATGCTCACTATGCACTCAAGCAGCTCTTGGTGGCAAAATTTGGA
AJ249293.1 cg3	AAATCTCCATATGCGTTGTTTACGGCACCGGGACCGTTGCGACTGATAATAATATGCTCTCAATACTCTAACCTTTTGACGCCATTTAGATGCTCACTATGCACTCAAGCAGCTCTTGGTGGCAAAATTTGGA
Consensus	AAATCTCCGTATGCGTTGTTTACGGCACCGGGACCGTTGCGACTGATAATAATATGCTCTCAATACTCTAACCTTTTGACGCCATTTAGATGCTCACTATGCACTCAAGCAGCTCTTGGTGGCAAAATTTGGA

Figure 2.5 : Vue partielle de l'alignement entre les différentes séquences de β 1,3 glucanases. Les trois premières séquences nucléotidiques représentent les trois groupes mis en évidence par séquençage dans ce travail, la quatrième séquence alignée est la séquence EST de la banque d'unigènes, les trois dernières séquences sont les séquences des trois β 1,3 glucanases de chicorée disponibles dans NCBI. La séquence du bas représente la séquence consensus entre les différentes séquences alignées.

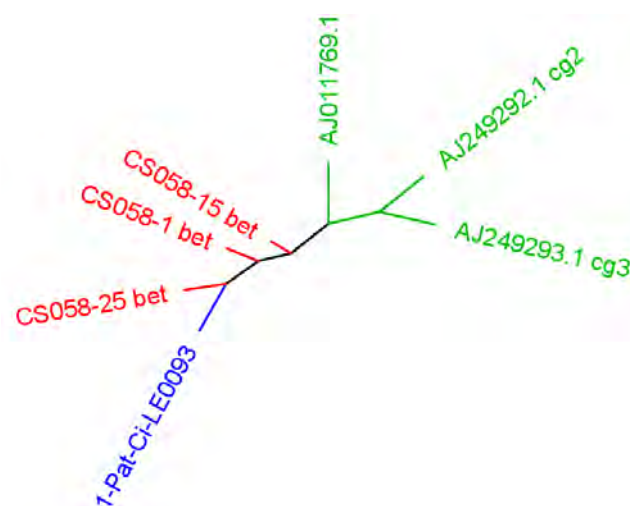


Figure 2.6 : Arbre phylogénique entre les différentes β 1,3 glucanases de chicorée sur base de l'alignement de celles-ci (séquences partielles). En rouge, séquences des trois groupes mis en évidence dans ce travail ; en bleu : séquence EST de la banque d'unigènes de chicorée mise en évidence par analyse *reciprocal best BLAST hit* ; en vert, les séquences des trois β 1,3 glucanases de chicorée disponibles dans NCBI.

CS062								
Clone 2	Clone 5	Clone 35	Clone 46	Clone 50	Clone 52	Clone 54	Clone 62	Changement aa
96								M/I
		168					168	
		206						Q/K
							214	K/R
	304							Q/L
		331	331					

Table 2.22 : Résultats de séquençage de 8 clones pour le gène CS062 (homologue de *Avr9/Cf-9 rapidly elicited protein 75*), modifications d'acides aminés. La position de chaque SNP est calculée à partir de l'amorce *reverse*. En jaune, n° du clone séquencé. La dernière colonne indique le changement éventuel d'un acide aminé induit par le SNP correspondant. En rouge, changement radical des propriétés de l'acide aminé modifié par le SNP.

CS076								
Clone 24	Clone 37	Clone 61	Clone 20	Clone 48	Clone 14	Clone 30	Clone 43	changement aa
Contig1								
55								
	82	82	82					
96				96				
104								
	107							
	167	167	167					
238								F/L
				397				
Contig2								
				60				
	155							
					238			délétion
					239			délétion
					240			délétion
					238			K/E
	382							

Table 2.23 : Résultats de séquençage de 8 clones pour le gène CS076 (homologue *Cf-5 disease resistance protein*), modifications d'acides aminés. En jaune, n° du clone séquencé ; En orange, Contig 1 = zone séquencée du coté de l'amorce *forward*, Contig2 = zone séquencée du coté de l'amorce *reverse*. Position de chaque SNP. La dernière colonne indique la modification d'acide aminé induite par le SNP. En rouge, changement radical des propriétés de l'acide aminé modifié par le SNP. Remarque : Les deux séquences *forward* et *reverse* ne se recouvrent pas.

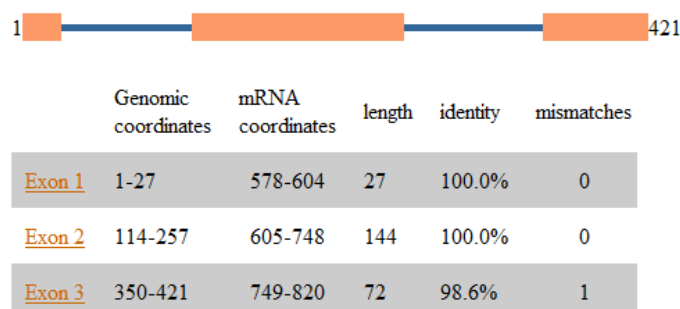


Figure 2.7 : Positions introns-exons dans la séquence CS076 (homologue *Cf-5 disease resistance protein*) séquencée du côté *forward*. Visualisation graphique de la position introns-exons (respectivement en orange et en bleu). Genomic coordinates, coordonnées de la séquence génomique correspondant à un exon (à partir du site de l'amorce *forward*); mRNA coordinates, position des exons sur la séquence codante EST ayant servi au design des amorces ; length, taille de l'exon ; identity, identité entre la séquence génomique et la séquence codante au niveau des exons ; mismatches, mismatches entre la séquence génomique et la séquence codante au niveau des exons.

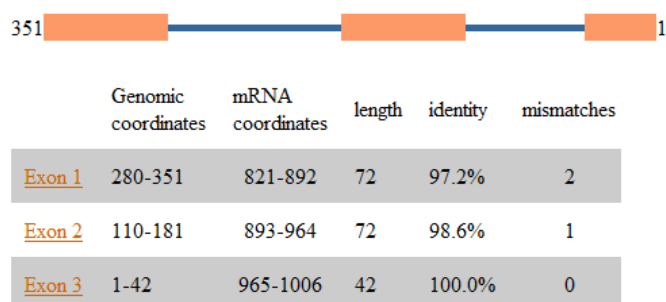


Figure 2.8 : Positions introns-exons dans la séquence CS076 (homologue *Cf-5 disease resistance protein*) séquencée du côté *reverse*. Visualisation graphique de la position introns-exons (respectivement en orange et en bleu). Genomic coordinates, coordonnées de la séquence génomique correspondant à un exon (à partir du site de l'amorce *reverse*); mRNA coordinates, position des exons sur la séquence codante EST ayant servi au design des amorces ; length, taille de l'exon ; identity, identité entre la séquence génomique et la séquence codante au niveau des exons ; mismatches, mismatches entre la séquence génomique et la séquence codante au niveau des exons.

CS078									
Clone 1	Clone 16	Clone 23	Clone 30	Clone 41	Clone 47	Clone 53	Clone 57	Clone 63	changement aa
	107				107				T/M
	125		125			125	125	125	
					205				
	276		276			276		276	H/R
	327								S/P

Table 2.23 : Résultats de séquençage de 9 clones pour le gène CS078 (Homologue *Glycine max putative Hs1pro-1-like receptor*), modifications d'acides aminés. La position de chaque SNP est calculée à partir de l'amorce *reverse*. En jaune, n° du clone séquencé. La dernière colonne indique la modification d'acide aminé induite par le SNP. En rouge, changement radical des propriétés de l'acide aminé modifié par le SNP.

CS160			
Clone 28	Clone 50	Clone 52	changement aa
	122		T/S
227			

Table 2.24 : Résultats de séquençage de 3 clones pour le gène CS0160 (homologue *ethylene responsive element binding factor 4*), modifications d'acides aminés. La position de chaque SNP est calculée à partir de l'amorce *reverse*. En jaune, n° du clone séquencé. La dernière colonne indique la modification d'acide aminé induite par le SNP. En rouge, changement radical des propriétés de l'acide aminé modifié par le SNP.

CS230										
Clone 1	Clone 6	Clone 8	Clone 19	Clone 27	Clone 30	Clone 42	Clone 49	Clone 56	Clone 62	changement aa
							81			K/R
284										I/F
	303			303					303	N/S
			345							
	373			373					373	
376										W/STOP
			378							

Table 2.25 : Résultats de séquençage de 10 clones pour le gène CS230 (NBS-LRR), modifications d'acides aminés. La position de chaque SNP est calculée à partir de l'amorce *forward*. En jaune, n° du clone séquencé. La dernière colonne indique la modification d'acides aminé induit par le SNP. En rouge, changement radical des propriétés de l'acide aminé modifié par le SNP ou codon STOP. En vert, SNP caractérisant l'allèle 1 ; En orange, SNP caractérisant l'allèle 2 ; En gris, séquence sans SNP caractérisant l'allèle 3 ; En bleu foncé, SNP caractérisant l'allèle 4 ; En bleu clair, SNP caractérisant l'allèle 5.

CS242 Groupe 1				
Clone 27	Clone 30	Clone 52	Clone 60	Changement aa
		156		délétion générant STOP
		167		
	236		236	
245				
		300		
316				délétion
317				délétion
318				délétion
319		319		délétion
320		320		délétion
321		321		délétion
		326		
		335		
		337		
		339		
		345		
			384	
	392			

Table 2.26 : Résultats de séquençage de 4 clones pour le gène CS242, modifications d'acides aminés. La position de chaque SNP est calculée à partir de l'amorce *forward*. En jaune, n° du clone séquencé. La dernière colonne indique la modification d'acide aminé induite par le SNP. En rouge, changement radical des propriétés de l'acide aminé modifié par le SNP ou codon STOP.

CS284								
Clone 7	Clone 37	Clone 28	Clone 1	Clone 54	Clone 45	Clone 22	Clone 16	Changement aa
29	29	29						
38	38	38						
			39	39	39			
45	45	45						I/F
			47					
	56							
57	57	57						H/D
						79	79	69 V/L
90	90	90						Y/H
			108	108	108			Y/H
			121					Y/H
					128			L/S
153	153	153						A/M
154	154	154						
							175	C/S
			239	239	239			
			262	262	262			D/G
			263	263	263			
			275	275	275			
			284	284	284			
			296	296	296			
		349						G/E
			359	359	359			
	456							N/D
			461	461	461			

Table 2.27. : Résultats de séquençage de 8 clones pour le gène CS284 (*WAK- related gene*), modifications d'acides aminés. La position de chaque SNP est calculée à partir de l'amorce *forward*. En jaune, n° du clone séquençé. La dernière colonne indique la modification d'acide aminé induite par le SNP. CS284 est potentiellement un pseudogène et l'impact des mutations est donc potentiellement nul. En rouge, changement radical des propriétés de l'acide aminé modifié par le SNP ou codon STOP. En vert, SNP caractérisant le paralogue 1 ; En orange, SNP caractérisant le paralogue 2 ; En bleu, les SNP caractérisant le paralogue 3.

CS284-1-T7 NRTENCNQs-VINLID-LHYQVYKCIHTYTCRNLYLKLHFLAPVVAGSAFMILMLCLVIFIWRRYKSNPLSYFPSKDKSPNVEGSLFFGVSSFFSYTELEDATRNFDPSQELGNGGFGV
CS284-16-T7 NRTENCNQs-VIYLID-LHYQVYKSIHTYTCRNLYLKLHFLAPVVAGSAFMILMLCLVIFIWRRYKSNPLSYFPSKDKSPNVEDGSLFFGVSSFFSYTELEDATRNFDPSQELGNGGFGV
CS284-22-T7 NRTENCNQs-VIYLID-LHYQVYKSIHTYTCRNLYLKLHFLAPVVAGSAFMILMLCLVIFIWRRYKSNPLSYFPSKDKSPNVEDGSLFFGVSSFFSYTELEDATRNFDPSQELGNGGFGV
CS284-28-T7 NRTENCNQs-VMYLFD-LDYQVYKCIHTYTCRNLYLKLHFLAPVVAGSAFMILMLCLVIFIWRRYKSNPLSYFPSKDKSPNVEDGSLFFGVSSFFSYTELEDATRNFDPSQELGNGGFGV
CS284-37-T7 NRTENCNQs-VMYLFD-LDYQVYKCIHTYTCRNLYLKLHFLAPVVAGSAFMILMLCLVIFIWRRYKSNPLSYFPSKDKSPNVEDGSLFFGVSSFFSYTELEDATRNFDPSQELGNGGFGV
CS284-7-T7 NRTENCNQs-VMYLFD-LDYQVYKCIHTYTCRNLYLKLHFLAPVVAGSAFMILMLCLVIFIWRRYKSNPLSYFPSKDKSPNVEDGSLFFGVSSFFSYTELEDATRNFDPSQELGNGGFGV
CS284-45-T7 NRTENCNQs-VINLID-LHYQVYKCIHTYTCRNLYLKLHFLAPVVAGSAFMILMLCLVIFIWRRYKSNPLSYFPSKDKSPNVEGSLFFGVSSFFSYTELEDATRNFDPSQELGNGGFGV
CS284-54-T7 NRTENCNQs-VINLID-LHYQVYKCIHTYTCRNLYLKLHFLAPVVAGSAFMILMLCLVIFIWRRYKSNPLSYFPSKDKSPNVEGSLFFGVSSFFSYTELEDATRNFDPSQELGNGGFGV
Consensus 1|NRTENCNQs VIYLID LHYQVYKCIHTYTCRNLYLKLHFLAPVVAGSAFMILMLCLVIFIWRRYKSNPLSYFPSKDKSPNVEDGSLFFGVSSFFSYTELEDATRNFDPSQELGNGGFGV

Figure 2.9 : Alignement de la séquence protéique de 8 clones de CS284 (*WAK- related gene*), mise en évidence de 2 codons stop. Description de la figure : à gauche, liste des clones alignés (CS284-n° du clone- T7). En bas, séquence consensus obtenue sur base de l'alignement des séquences protéiques des 8 clones. Les acides aminés des séquences protéiques sont sous forme code à une lettre. Les codons stop sont représentés par «- ».

CS286							
Clone 1	Clone 4	Clone 9	Clone 11	Clone 22	Clone 46	Clone 54	Changement aa
47							
80							
179							
215							
224							
263							
377							M/I
		441					C/R

Table 2.28. : Résultats de séquençage de 7 clones pour le gène CS286 (*WAK- related gene*), modifications d'acides aminés. En jaune, n° du clone séquençé ; Position de chaque SNP à partir de l'amorce *forward*. La dernière colonne indique la modification d'acide aminé induite par le SNP. En rouge, changement radical des propriétés de l'acide aminé modifié par le SNP ou codon STOP. En bleu, SNP caractérisant l'allèle 1 ; En rose, le SNP caractérisant l'allèle 2 ; Les clones 4-9-11-22-46-54 appartiennent à l'allèle 3.