

# Thèse

Présenté pour l'obtention du titre de

**DOCTEUR DE L'UNIVERSITE D'EVRY-VAL-D'ESSONNE**

**En BIOLOGIE MOLECULAIRE**

**Par BRESSON Aloïs**



## **Caractérisation de $R_1$ et $R_{US}$ : deux loci de résistance à la rouille foliaire sur le chromosome 19 du peuplier**

Soutenance publique le 11 Janvier 2011

### **Membres du jury :**

#### Rapporteurs :

Lefebvre Véronique  
Prat Daniel

Directeur de recherches, INRA d'Avignon  
Professeur de l'université de Lyon 1

#### Examineurs :

Bastien Catherine  
Duplessis Sébastien  
Sturbois Bénédicte  
Chalhoub Boulos  
Faivre Rampant Patricia

Directeur de recherches, INRA d'Orléans  
Chargé de recherches, INRA de Nancy  
Professeur de l'université d'Evry  
Directeur de recherches, INRA d'Evry (Directeur de thèse)  
Chargé de recherches, INRA d'Evry (Encadrant)



# TABLE DES MATIERES

<b>LISTE DES FIGURES.....</b>	<b>5</b>
<b>LISTE DES TABLEAUX .....</b>	<b>7</b>
<b>LISTE DES ANNEXES .....</b>	<b>7</b>
<b>REMERCIEMENTS.....</b>	<b>8</b>

## Partie I

### INTRODUCTION

<b>1 Introduction générale.....</b>	<b>12</b>
<b>2 Les Peupliers.....</b>	<b>12</b>
2.1 Description botanique .....	12
2.2 Les cultures de peupliers.....	14
2.3 Un arbre modèle pour la génomique des arbres forestiers .....	16
<b>3 Les rouilles foliaires des peupliers.....</b>	<b>19</b>
3.1 Les maladies du peuplier.....	19
3.2 <i>Melampsora larici-populina</i> .....	20
3.3 <i>Mlp</i> dans les peupleraies cultivées.....	22
3.4 Un pathosystème modèle .....	25
<b>4 Concepts de la résistance des plantes aux agents pathogènes... 25</b>	
4.1 Quelques définitions .....	26
4.2 Résistance constitutive.....	26
4.3 Résistance induite.....	26
<b>5 Les gènes de résistance .....</b>	<b>30</b>
5.1 Les différentes catégories de gènes R.....	30
5.2 Les gènes de type NBS-LRR .....	31
5.3 Organisation des gènes NBS-LRR.....	34
5.4 Évolution des gènes NBS-LRR.....	36
<b>6 Génétique de la résistance aux rouilles foliaires du peuplier... 37</b>	
6.1 Résistance qualitative et résistance quantitative .....	37
6.2 Les loci de résistance qualitative à <i>Mlp</i> .....	38
6.3 Un locus de résistance quantitative à <i>Mlp</i> : <i>R<sub>US</sub></i> .....	41
<b>7 Objectifs .....</b>	<b>42</b>
7.1 Les objectifs de recherche de l'INRA .....	42
7.2 Les objectifs de cette thèse .....	43

## Partie II

### MATÉRIELS ET MÉTHODES

<b>1 Matériel Biologique .....</b>	<b>46</b>
1.1 Famille 54.....	46
1.2 Clone Nisqually-1 .....	46
1.3 <i>Melampsora larici-populina</i> .....	46
<b>2 Ressources Génomiques .....</b>	<b>47</b>
2.1 Banque BAC de l'individu 101-74 <i>P. trichocarpa</i> .....	47
2.2 Séquences d'extrémité des clones BAC.....	48
<b>3 Méthodes .....</b>	<b>48</b>
3.1 Méthode générale .....	48
3.2 Evaluation de la résistance à <i>Mlp</i> .....	49
3.3 Développement de marqueurs .....	50
3.4 Réactions PCR et électrophorèse .....	54
3.5 Cartographie génétique.....	56
3.6 Cartographie physique.....	58
3.7 Analyse de séquences.....	62

## Partie III

### ORGANISATION STRUCTURALE DE L'EXTREMITÉ DU CHROMOSOME 19

<b>1 Introduction .....</b>	<b>65</b>
1.1 Vers le clonage positionnel de $R_{US}$ .....	65
1.2 Vers le clonage positionnel de $R_1$ .....	65
<b>2 Résultats .....</b>	<b>66</b>
2.1 Séquence du génome de Nisqually-1 .....	66
2.2 Développement de marqueurs .....	70
2.3 Test des couples d'amorces .....	71
2.4 Cartographie génétique de la famille 54 .....	74
2.5 Cartographie physique de l'individu 101-74 .....	80
2.6 Caractéristiques de la zone génomique.....	86
2.7 Vers le clonage positionnel de $R_{US}$ et $R_1$ .....	88

## Partie IV

### CLONAGE POSITIONNEL DES LOCI DE RESISTANCE

<b>1</b>	<b>Le clonage positionnel du locus <math>R_{US}</math></b>	<b>91</b>
1.1	Séquençage des clones BAC	91
1.2	Cartographie génétique fine de $R_{US}$	96
1.3	Caractérisation de l'allèle candidat pour le locus $R_{US}$	98
<b>2</b>	<b>Vers le clonage positionnel du locus <math>R_1</math></b>	<b>101</b>
2.1	Cartographie génétique fine du locus $R_1$	102
2.2	Cartographie comparée	103

## Partie V

### DISCUSSION GENERALE

<b>1</b>	<b>Clonage positionnel des loci de résistance</b>	<b>105</b>
1.1	Développement de marqueurs	105
1.2	Utilité des cartes génétiques	106
1.3	Utilité de la banque BAC et des cartes physiques	110
1.4	Variation du taux de recombinaison	111
1.5	Distorsions de ségrégation	112
<b>2</b>	<b>Le séquençage de génome</b>	<b>112</b>
2.1	Les erreurs d'assemblage de la séquence du génome	112
2.2	Les zones complexes dans le séquençage de génome	114
<b>3</b>	<b>Les loci de résistance</b>	<b>115</b>
3.1	Le locus $R_{US}$	115
3.2	Le locus $R_1$	118
3.3	$R_{US}$ et $R_1$ : deux loci différents	119
3.4	Le locus $MER$	120
<b>4</b>	<b>Cluster de gènes R</b>	<b>121</b>
4.1	Un super cluster de gène R	121
4.2	L'évolution des gènes NBS-LRR	121
4.3	Les éléments transposables	123

## Partie VI

### PERSPECTIVES GENERALES

<b>1</b>	<b>Validation fonctionnelle du gène candidat <math>R_{US}</math></b> .....	<b>125</b>
<b>2</b>	<b>Compréhension des mécanismes de cette résistance</b> .....	<b>126</b>
2.1	Etude de l'expression du gène candidat.....	126
2.2	Domaine NLS.....	126
2.3	Domaine LRR.....	127
2.4	Etude des éléments transposables .....	127
2.5	Génotype [ $R_{US}/R_{US}$ ].....	127
<b>3</b>	<b>L'interaction peuplier-<i>Mlp</i></b> .....	<b>128</b>
<b>4</b>	<b>Etude du locus <math>R_{US}</math> dans les populations naturelles de <i>P. trichocarpa</i></b> .....	<b>128</b>
<b>5</b>	<b>Vers le clonage positionnel de <math>R_1</math></b> .....	<b>129</b>
5.1	Développement de nouveaux marqueurs génétiques .....	129
5.2	Clonage physique de l'allèle $R_1$ .....	129
<b>6</b>	<b>Y a-t-il d'autres gènes proches des loci responsables de résistance à <i>Mlp</i> ?</b> .....	<b>130</b>
6.1	QTL.....	130
6.2	Etude des individus porteurs d'une recombinaison proche des loci de résistance	131
<b>7</b>	<b>L'inventaire des gènes de résistance à <i>Mlp</i></b> .....	<b>131</b>
7.1	Les marqueurs utiles .....	131
7.2	Le locus <i>MER</i> .....	131
7.3	Caractériser de nouveaux loci de résistance.....	132
<b>8</b>	<b>Séquençage de la région génomique couvrant les gènes <math>R_1</math> et <math>R_{US}</math></b>	<b>132</b>
<b>9</b>	<b>Gestion des peupleraies cultivées</b> .....	<b>133</b>
9.1	A court terme.....	133
9.2	A long terme .....	133
	<b>ABREVIATIONS</b> .....	<b>134</b>
	<b>BIBLIOGRAPHIE</b> .....	<b>135</b>
	<b>ANNEXE</b> .....	<b>144</b>
	<b>ARTICLE</b> .....	<b>152</b>

# LISTE DES FIGURES

- Figure 1 : Une feuille de peuplier infectée par le pathogène *Melampsora larici-populina*.
- Figure 2 : Comparaison entre un arbre sensible et un arbre résistant à la rouille foliaire.
- Figure 3 : Production de bois par surface des trois essences principales de feuillus en France.
- Figure 4 : Culture monoclonale de peupliers.
- Figure 5 : Obtention de clones hybrides à partir de croisements d'espèces sauvages.
- Figure 6 : Evolution des clones de peupliers plantés en France.
- Figure 7 : Exemples de débouchés pour le bois de peuplier.
- Figure 8 : La duplication du génome entier du peuplier.
- Figure 9 : Cycle de reproduction de *Melampsora larici-populina*.
- Figure 10 : Ecidies de *Mlp* sur un mélèze.
- Figure 11 : Illustration du modèle gène à gène.
- Figure 12 : Modèle ligand-récepteur et modèle de garde.
- Figure 13 : Le modèle en zig-zag, les différentes couches de résistance.
- Figure 14 : Les différentes catégories de gènes de résistance définies chez les plantes.
- Figure 15 : Modèle de fonctionnement des protéines NBS-LRR.
- Figure 16 : Phylogénie de 1600 gènes NBS-LRR issues de plusieurs espèces végétales.
- Figure 17 : Localisation des gènes NBS-LRR dans le génome d'*Arabidopsis*.
- Figure 18 : Localisation des gènes NBS-LRR dans le génome de *Populus trichocarpa* V1.1.
- Figure 19 : Le super cluster de gènes NBS-LRR de l'extrémité du chromosome 19 du peuplier.
- Figure 20 : Phénotypes d'individus descendant de la famille 54 lors de l'infection par *Mlp*.
- Figure 21 : Localisation des loci  $R_1$  et  $R_{US}$  sur le chromosome 19.
- Figure 22 : Stratégie pour le clonage positionnel du locus  $R_{US}$ .
- Figure 23 : Protocole expérimental du phénotypage de la famille 54.
- Figure 24 : Distribution des moyennes phénotypiques clonales des individus de la famille 54 pour la taille des sores.
- Figure 25 : Couples d'amorces développés dans les gènes TNL.
- Figure 26 : Marqueurs génétiques développés à partir de séquences de produits d'amplification PCR.
- Figure 27 : Protocole des réactions PCR.
- Figure 28 : Tableau des résultats de génotypage de la famille 54.
- Figure 29 : Alignement entre les 3 premiers Mb du chromosome 19 de la version V1 et la région équivalente de la version V2 du génome de Nisqually-1.
- Figure 30 : Phylogénie des gènes NBS-LRR présents sur le chromosome 19.
- Figure 31 : Phylogénie des gènes TNL de l'extrémité du chromosome 19 de Nisqually-1 V2.
- Figure 32 : Couples d'amorces développés sur le génome de Nisqually-1 V1.
- Figure 33 : Génotype du 101-74 et du 73028-62 et des descendants de la famille 54 pour les différents cas de marqueurs.
- Figure 34 : Comparaison entre les cartes génétiques du père 101-74 *P. trichocarpa*.

Figure 35 : Distribution de la famille 54 lors de deux expériences de phénotypage pour la taille des sores.

Figure 36 : Comparaison entre les cartes génétiques de la mère 73028-62 *P. deltoides*.

Figure 37 : Alignement entre les cartes génétiques du 101-74 et du 73028-62.

Figure 38 : Alignement des cartes génétiques des deux parents et des séquences des deux assemblages du génome de Nisqually-1.

Figure 39 : Construction des cartes physiques pour chacun des deux haplotypes du 101-74.

Figure 40 : Alignement entre les cartes physiques et la carte génétique du 101-74.

Figure 41 : Alignement entre la carte physique de l'haplotype  $hR_{US}$  et les deux assemblages du génome de Nisqually-1.

Figure 42 : Taux de recombinaison entre la carte génétique de 333 individus et la carte physique de l'haplotype  $hR_{US}$ .

Figure 43 : Taux de recombinaison entre la carte génétique de 1399 individus et la carte physique de l'haplotype  $hR_{US}$ .

Figure 44 : Alignement entre le génome et les cartes physiques : définition de cinq zones différentes.

Figure 45 : Localisation génétique des loci de résistance sur les deux assemblages du génome de Nisqually-1.

Figure 46 : Alignement de la carte génétique consensus des parents de la famille 54 avec les cartes physiques du 101-74.

Figure 47 : Annotation des clones BAC du 101-74.

Figure 48 : Alignement des séquences des clones BAC entre eux par Dot Plot.

Figure 49 : Alignement entre les séquences des clones BAC du 101-74 et le génome de Nisqually-1 V2.

Figure 50 : Alignement de la carte génétique fine de  $R_{US}$  avec les clones BAC de l'haplotype  $hR_{US}$  et les individus recombinants proches du locus.

Figure 51 : Alignement de la carte génétique fine de  $R_{US}$  avec les clones BAC des haplotypes  $hR_{US}$  et  $hr_{US}$  autour du locus.

Figure 52 : Structure des gènes candidats pour  $R_{US}$  et  $r_{US}$ .

Figure 53 : Comparaison des CDSs des gènes candidats pour  $R_{US}$  et  $r_{US}$ .

Figure 54 : Alignement des exons du gène candidat avec les autres gènes du cluster de TNL.

Figure 55 : Expression de l'allèle candidat par RTq-PCR.

Figure 56 : Alignement de la carte génétique fine de  $R_1$  et de la séquence du génome V2.

Figure 57 : Carte génétique fine du locus  $R_1$  alignée avec le génome de Nisqually-1 V2 et le clone BAC 97G09 du 101-74.

Figure 58 : Comparaison entre les logiciels de cartographie génétique MAPMAKER et Tmap.

Figure 59 : La carte physique de Nisqually-1 par fingerprint pour l'extrémité du chromosome 19.

Figure 60 : le domaine NLS.

Figure 61 : QTLs de résistance quantitative évaluée en laboratoire chez le 101-74 vis-à-vis de sept souches de *Mlp*.

Figure 62 : Les loci *MER*,  $R_1$  et  $R_{US}$  localisés sur le génome de Nisqually-1 V2.

Figure 63 : Comparaison entre deux haplotypes au locus RGA08 chez la banane.

Figure 64 : Comparaison entre deux haplotypes au locus *RI* chez la pomme de terre.

Figure 65 : Les clones BAC en cours de séquençage.

# LISTE DES TABLEAUX

Tableau 1 : Les sections et les principales espèces du genre *Populus*.

Tableau 2 : Liste non exhaustive de gènes R clonés chez les végétaux.

Tableau 3 : Nombre de gènes NBS-LRR chez *P. trichocarpa*, *A. thaliana*, et *Oryza sativa*.

Tableau 4 : Inventaire des gènes NBS-LRR sur l'extrémité du chromosome 19.

Tableau 5 : Les individus atypiques de la famille 54.

# LISTE DES ANNEXES

Annexe 1 : Composition des solutions

Annexe 2 : Motifs MEME utilisés pour trouver les gènes NBS-LRR sur les séquences.

Annexe 3 : Les couples d'amorces utilisés en cartographie génétique locale.

Annexe 4 : Les couples d'amorces utilisés en cartographie physique.

Annexe 5 : Les couples d'amorces développés pour la cartographie fine de  $R_{US}$ .

Annexe 5 : Les couples d'amorces développés pour la cartographie fine de  $R_1$ .

# REMERCIEMENTS

Je tiens à remercier mes rapporteurs Véronique Lefebvre et Daniel Prat ainsi que les autres membres du jury d'avoir accepté de se plonger quelque temps dans mon sujet de recherche.

Je remercie également Michel Caboche et Heribert Hirt de m'avoir accueilli dans l'Unité de Recherche en Génomique Végétale me permettant d'intégrer l'équipe peuplier guidée par Patricia Faivre Rampant.

Pendant ces quatre années, bien du monde est passé dans l'équipe "peuplier", de façon éphémère ou plus durable, tous y ont laissé leur empreinte : Louisa, Charlotte "Blondy", Clément "le jeune", Jeanne, Isabella, Séréna "petit poisson", Frédérique, Frank et enfin Patricia, qui a passé cinq ans à me supporter.

Mais l'équipe "peuplier" est proche d'autres équipes notamment celle de la vigne et de la "polyplœdie" : Aurélie, Marco, Cléa, Isabelle (ton soutien psychologique, qui sans aucun doute, ne m'aurait pas permis d'arriver jusqu'au bout. De ton lapin à ma biquette : je te dois, tout simplement, plus que beaucoup), Anne Françoise, Harry, Jérémy "JJ", Gwenaëlle "pitchoun", Cécile, Imen, Mathieu "Nounours, mon thésard référent", Nathalie et Boulos.

Cela a été une joie de passer cinq ans parmi vous, avec beaucoup de bonne humeur, de phrases amusantes écrites sur la porte du laboratoire qui regroupe ces trois équipes et aussi une entraide dans le but de faire un meilleur travail.

Je remercie particulièrement l'équipe de l'INRA d'Orléans UAGPF qui a œuvré à la réussite de ce travail sur la rouille foliaire du peuplier : Véronique, Arnaud, Vanina, Isabelle et Catherine. Les petits jeunes avec qui j'ai pu rigoler au cours de mes voyages, mais aussi beaucoup d'autres dont je ne connais même pas le nom et qui ont rendu ce travail de thèse possible.

Je n'oublie pas aussi l'équipe "(Eco)TILLING", qui m'a accueilli à mon arrivée à l'URGV : Erwan, Bénédicte, Isabelle et Abdel.

Sans oublier tous les thésards de l'URGV avec qui on se tient les coudes entre nous. Et tout l'URGV. Ce petit groupe où il y a tellement de gens intéressants et dont je garderai un bon souvenir.

Merci enfin à Francis Quetier, ma mère, Boulos, Bénédicte et Patricia pour leur correction, surtout orthographique.

Mes derniers remerciements sont pour ma famille, en commençant par mes parents qui m'ont toujours soutenu dans ces études qui n'en finissaient pas en m'apportant leur soutien indéfectible. Et mes amis qui ont su me soutenir quand tout allait mal et me remotiver pour en arriver là.

Et tellement d'autres...



**Partie I**  
**INTRODUCTION**



# Partie I

## INTRODUCTION

<b>1</b>	<b>Introduction générale.....</b>	<b>12</b>
<b>2</b>	<b>Les Peupliers.....</b>	<b>12</b>
2.1	Description botanique .....	12
2.1.1	Classification.....	12
2.1.2	Caractéristiques botaniques .....	12
2.1.3	Aire de répartition .....	13
2.1.4	Reproduction.....	13
2.2	Les cultures de peupliers.....	14
2.2.1	Les principales caractéristiques .....	14
2.2.2	Les cultivars .....	14
2.2.3	Les débouchés de la production de peuplier .....	15
2.3	Un arbre modèle pour la génomique des arbres forestiers.....	16
2.3.1	Espèce modèle.....	16
2.3.2	La séquence du génome de l'individu Nisqually-1 <i>P. trichocarpa</i> .....	17
<b>3</b>	<b>Les rouilles foliaires des peupliers.....</b>	<b>19</b>
3.1	Les maladies du peuplier .....	19
3.2	<i>Melampsora larici-populina</i> .....	20
3.2.1	Le cycle de <i>Mlp</i> .....	20
3.2.2	Symptômes de la maladie chez le peuplier .....	21
3.2.3	La diversité chez <i>Melampsora larici-populina</i> .....	21
3.3	<i>Mlp</i> dans les peupleraies cultivées .....	22
3.3.1	Les dégâts .....	22
3.3.2	Exemple d'un cultivar : "Beaupré" .....	23
3.3.3	Les luttés.....	23
3.3.4	Problème sanitaire prévisible des cultures de peuplier.....	24
3.3.5	Les pistes pour des résistances durables.....	24
3.4	Un pathosystème modèle.....	25
<b>4</b>	<b>Concepts de la résistance des plantes aux agents pathogènes. 25</b>	
4.1	Quelques définitions.....	26
4.1.1	Agents pathogènes nécrotrophes et biotrophes.....	26
4.1.2	Résistance non-hôte et hôte .....	26
4.2	Résistance constitutive .....	26
4.3	Résistance induite.....	26
4.3.1	Le modèle gène à gène .....	27
4.3.2	Résistance par reconnaissance d'éléments constitutifs de l'agent pathogène ....	28
4.3.3	Résistance par reconnaissance des effecteurs de l'agent pathogène .....	29
4.3.4	Evènements de signalisation et réactions de défense .....	30
<b>5</b>	<b>Les gènes de résistance .....</b>	<b>30</b>
5.1	Les différentes catégories de gènes R .....	30
5.2	Les gènes de type NBS-LRR .....	31
5.2.1	Les différents domaines.....	31
5.2.2	Les différentes catégories .....	33



<b>5.3</b>	<b>Organisation des gènes NBS-LRR.....</b>	<b>34</b>
5.3.1	Nombre de gènes NBS-LRR.....	34
5.3.2	Distribution des gènes NBS-LRR .....	35
<b>5.4</b>	<b>Évolution des gènes NBS-LRR .....</b>	<b>36</b>
5.4.1	Les mécanismes évolutifs .....	36
5.4.2	Diversité à l'intérieur des gènes.....	37
<b>6</b>	<b>Génétique de la résistance aux rouilles foliaires du peuplier .</b>	<b>37</b>
<b>6.1</b>	<b>Résistance qualitative et résistance quantitative.....</b>	<b>37</b>
<b>6.2</b>	<b>Les loci de résistance qualitative à <i>Mlp</i> .....</b>	<b>38</b>
6.2.1	Locus <i>MER</i> .....	38
6.2.2	Locus <i>R</i> <sub>1</sub> .....	39
6.2.3	Les résistances qualitatives identifiées pour d'autres espèces de <i>Melampsora</i> ..	40
<b>6.3</b>	<b>Un locus de résistance quantitative à <i>Mlp</i> : <i>R</i><sub>US</sub>.....</b>	<b>41</b>
6.3.1	Analyse de QTL .....	42
<b>7</b>	<b>Objectifs .....</b>	<b>42</b>
<b>7.1</b>	<b>Les objectifs de recherche de l'INRA .....</b>	<b>42</b>
<b>7.2</b>	<b>Les objectifs de cette thèse .....</b>	<b>43</b>
7.2.1	Organisation structurale de l'extrémité du chromosome 19 .....	43
7.2.2	Identification du gène <i>R</i> <sub>US</sub> par clonage positionnel.....	43
7.2.3	Identification du gène <i>R</i> <sub>1</sub> par cartographie comparée .....	44



Photo : P. Frey

**Figure 1** : Une feuille de peuplier infectée par le pathogène *Melampsora larici-populina*.



Photo : M. Villar

**Figure 2** : Comparaison entre un arbre sensible (à droite) et un arbre résistant à la rouille foliaire (à gauche).

**Tableau 1** : Les sections et les principales espèces du genre *Populus*.

Sections	Principales espèces	Origines	Remarques
<b>POPULUS</b>	<i>P. tremula</i>	Eurasie	Peupliers trembles
	<i>P. tremulodes</i>	Amérique du Nord	
	<i>P. alba</i>	Eurasie	Peupliers blancs
	<i>P. canescens</i>	Eurasie	Peupliers grisards
<b>TACAHAMACA</b>	<i>P. simonii</i>	Chine	
	<i>P. yunnanensis</i>	Chine	Peupliers baumiers
	<i>P. trichocarpa</i>	Amérique du Nord	
<b>LEUCOÏDE</b>	<i>P. lasiocarpa</i>	Chine	
	<i>P. glauca</i>	Asie	
	<i>P. heterophylla</i>	Amérique du Nord	
<b>TURANGA</b>	<i>P. euphratica</i>	Eurasie et Afrique	
	<i>P. pruinosa</i>	Asie	
<b>AIGEIOS</b>	<i>P. nigra</i>	Eurasie	Peupliers noirs
	<i>P. fremontii</i>	Amérique du Nord	
	<i>P. deltoides</i>	Amérique du Nord	

# 1 Introduction générale

La rouille foliaire du peuplier, principalement causée par le champignon *Melampsora larici-populina* (*Mlp*), est l'une des principales maladies affectant les peupleraies européennes (Paillassa, 2006 ; Pinon & Frey, 2005). Les symptômes chez le peuplier sont la formation de pustules orangées (urédies ou sores) sur la face inférieure des feuilles (figure 1). Les conséquences sont une diminution de la mise en réserve des substances carbonées et azotées mais aussi une perte prématurée des feuilles (figure 2). Des épidémies répétées conduisent à un défaut de croissance et à un affaiblissement général de l'arbre. Ces arbres sont ensuite la proie d'autres parasites qui, cumulés à ceux de la rouille foliaire, peuvent aboutir à la mort de l'arbre. Les premiers dégâts en peupleraie cultivée, relevés en Europe, datent des années 1980. Depuis cette période, on assiste à une rapide expansion de la maladie, due à la capacité du champignon à contourner rapidement les résistances introduites dans les cultures monoclonales de peuplier. Ce champignon oblige donc à explorer de nouvelles stratégies de sélection pour tenter d'obtenir des résistances plus durables.

## 2 Les Peupliers

### 2.1 Description botanique

#### 2.1.1 Classification

Le genre *Populus* appartient à la famille des Salicacées, qui comprend aussi le genre *Salix* (saule). Le genre *Populus* se divise en une trentaine d'espèces environ (ce chiffre peut varier selon les auteurs), qui sont réparties en 6 sections botaniques différentes : Tacamahaca, Turanga, Abaso, *Populus* (anciennement *Leuce*), Aigeiros et *Leucoides*. (Tableau 1 ; Cervera *et al*, 2005 ; Hamez & Dayanandan, 2004 ; Eckenwalder, 1996).

#### 2.1.2 Caractéristiques botaniques

Les peupliers sont des espèces pionnières, héliophiles, thermophiles et apprécient les sols bien alimentés en eau. On peut observer une grande diversité dans la forme et la couleur des feuilles, des bourgeons, des pétioles, et des inflorescences. Leurs feuilles sont simples, alternes, caduques et sont généralement grandes et à long pétiole.



Les peupliers sont des arbres à croissance rapide (2 mètres (m)/an) qui peuvent atteindre 25 à 30 m de hauteur en une vingtaine d'années et à courte durée de vie, ce qui en fait une essence de bon rapport pour l'exploitation.

### **2.1.3 Aire de répartition**

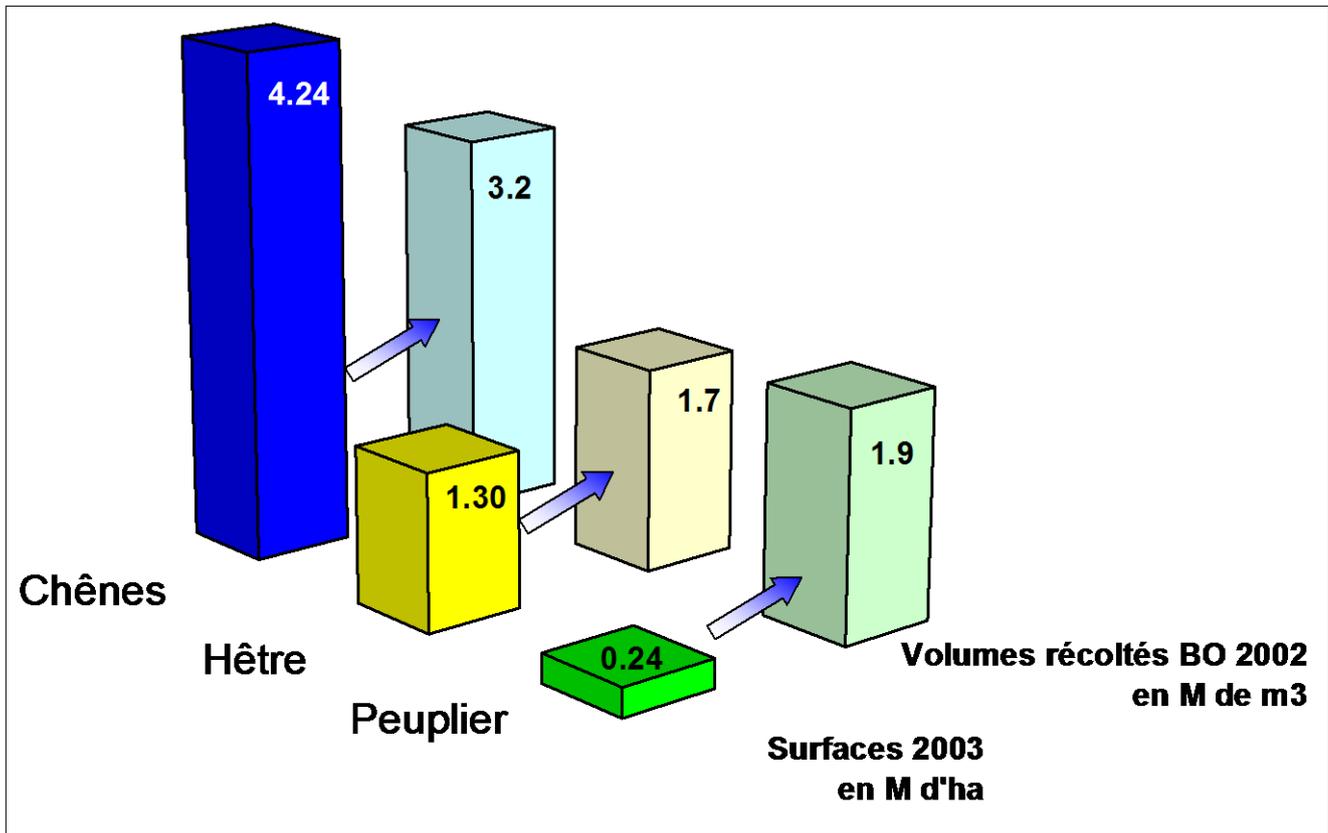
Les peupliers couvrent une vaste aire géographique des régions tempérées de l'hémisphère nord, comprenant l'Amérique du Nord, l'Europe et l'Asie. On les trouve généralement dans des terres sableuses et/ou limoneuses. Des inondations temporaires n'empêchent pas le développement de l'arbre. Les plaines alluviales répondent parfaitement à son exigence élevée en eau et en minéraux. Trois espèces sont indigènes en France : le peuplier tremble (*P. tremula*), le peuplier blanc (*P. alba*), le peuplier noir (*P. nigra*). Le peuplier noir et le peuplier blanc sont présents naturellement le long des ripisylves et en assurent l'équilibre écologique. Le peuplier tremble est trouvé en forêt.

### **2.1.4 Reproduction**

Les peupliers sont des arbres dioïques : les fleurs mâles et femelles sont portées par des arbres distincts. La maturité sexuelle du peuplier est atteinte après 6 à 10 ans. Les fleurs sont disposées en chatons et les fruits, appelés capsules, renferment de nombreuses graines cotonneuses. Ces dernières sont disséminées par le vent et l'eau à maturité et peuvent parcourir de grandes distances. L'espèce présente aussi une reproduction végétative importante par bouturage ou par drageonnage.

Les hybrides interspécifiques sont présents naturellement pour les espèces qui possèdent une même aire géographique (sympatrique), et sont fertiles, cependant certaines espèces ne sont pas compatibles entre elles (Eckenwalder, 1996). En effet, l'hybridation entre les espèces n'est pas toujours possible ; ainsi les peupliers de la section *Populus* ne peuvent s'hybrider avec les sections Aigeiros et Tacamahaca. De plus, certains croisements peuvent fonctionner seulement dans un sens et pas dans l'autre ; c'est le cas, par exemple, des croisements entre les espèces *P. deltoides* et *P. nigra*, où il est nécessaire que la mère soit *P. deltoides* et le père *P. nigra*. Dans le cas inverse, l'embryon avorte après la fécondation.

L'impossibilité d'autofécondation due au caractère dioïque de cette espèce et à la grande dispersion du pollen et des graines par le vent, amène à un fort niveau d'hétérozygotie.



D'après A. Berthelot

**Figure 3** : Production de bois par surface des trois essences principales de feuillus en France.

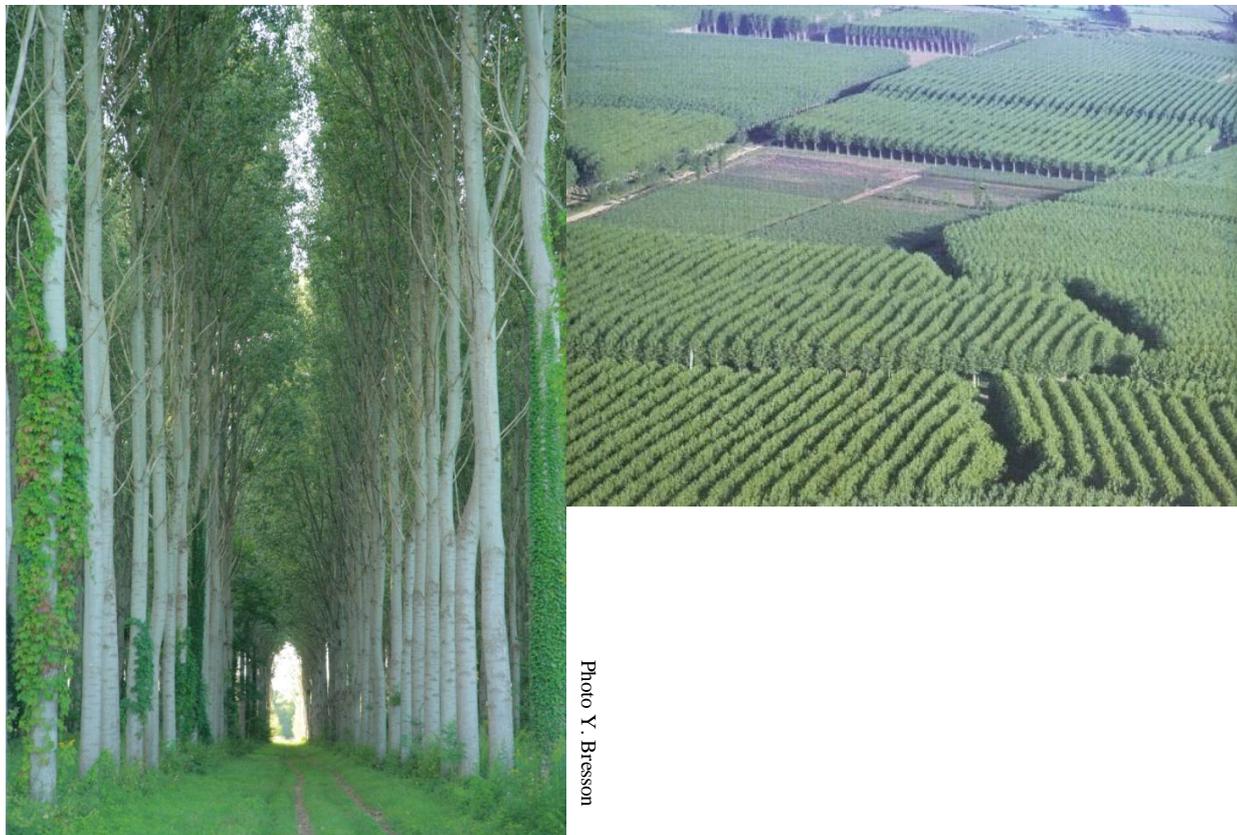


Photo Y. Bresson

©S. Bisolffi

**Figure 4** : Culture monoclonale de peupliers.

## 2.2 Les cultures de peupliers

### 2.2.1 Les principales caractéristiques

En Europe, les principales zones populicoles se trouvent en Italie, en Belgique, aux Pays-Bas, en France et en Espagne. Les cultures de peupliers en France appartiennent pour la plupart à des propriétaires privés. A l'image du reste de la forêt française, les plantations sont extrêmement morcelées, puisque la surface moyenne par propriétaire est de l'ordre d'un hectare (ha) seulement (Berthelot & Chantre, 2004).

Les cultures de peuplier nécessitent beaucoup d'eau, elles sont donc souvent plantées dans des vallées proches des cours d'eau. Présentes dans tout le pays, elles se situent traditionnellement dans les grandes vallées alluviales du Nord de la France (Champagne-Ardenne, Picardie), et elles se sont développées ces vingt dernières années dans l'ouest du pays (Pays-de la Loire, Centre, Aquitaine).

La densité moyenne d'une peupleraie est de 200 tiges/ha avec une productivité qui peut atteindre 20 m<sup>3</sup> par ha et par an. En France, la culture de peupliers occupait environ 240 000 ha en 2002, soit 1,6 % des surfaces forestières, avec 1,9 millions de m<sup>3</sup> de bois produit par an, soit 30 % de la production de grumes feuillus, en seconde position après le chêne qui couvre 4 240 000 ha et produit 3,2 millions de m<sup>3</sup> de bois par an (figure 3). La faible surface de peuplier cultivé est compensée par la rapidité de croissance et l'utilisation de mode de culture productiviste à la limite de l'agriculture : culture monoclonale, plantation régulière et de fortes quantités d'intrants. Ces cultures constituent l'un des écosystèmes forestiers les plus simplifiés : des plantations à mi-chemin entre l'agriculture et la forêt (figure 4).

Le peuplier ne supporte pas la compétition ; par conséquence, il a besoin de beaucoup de soins lors de ses premières années de croissance (entretien, élagage, taille), et nécessite des apports d'engrais, de pesticides et de fongicides afin d'optimiser son potentiel de productivité.

Le peuplier est aussi utilisé en aménagement paysager (bordure des routes et des cours d'eau). Le clone *P. nigra* var "italica" est un arbre d'alignement très répandu dans le paysage français depuis la fin du XVIII siècle.

### 2.2.2 Les cultivars

La populiculture est fondée sur la multiplication par bouturage des tiges issues d'un même génotype, ce qui permet la plantation et la commercialisation de variétés clonales. L'homogénéité de production d'une peupleraie monoclonale est un avantage d'un point de vue

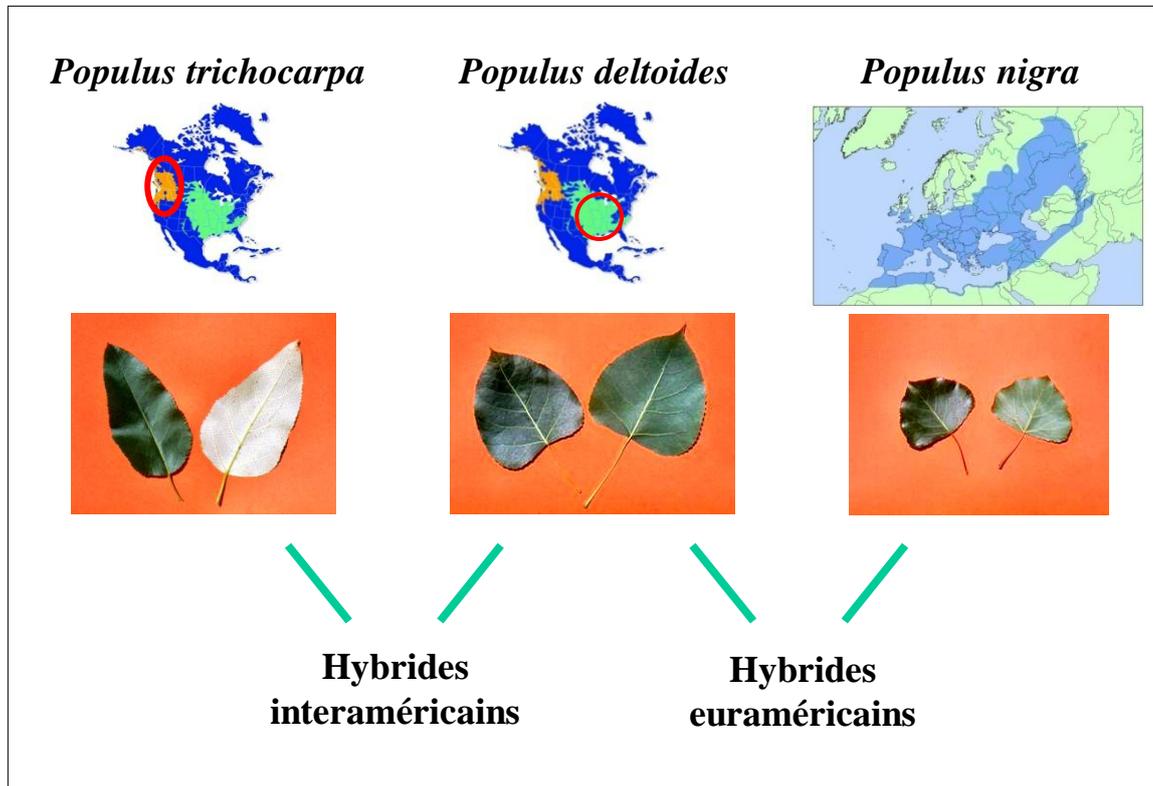
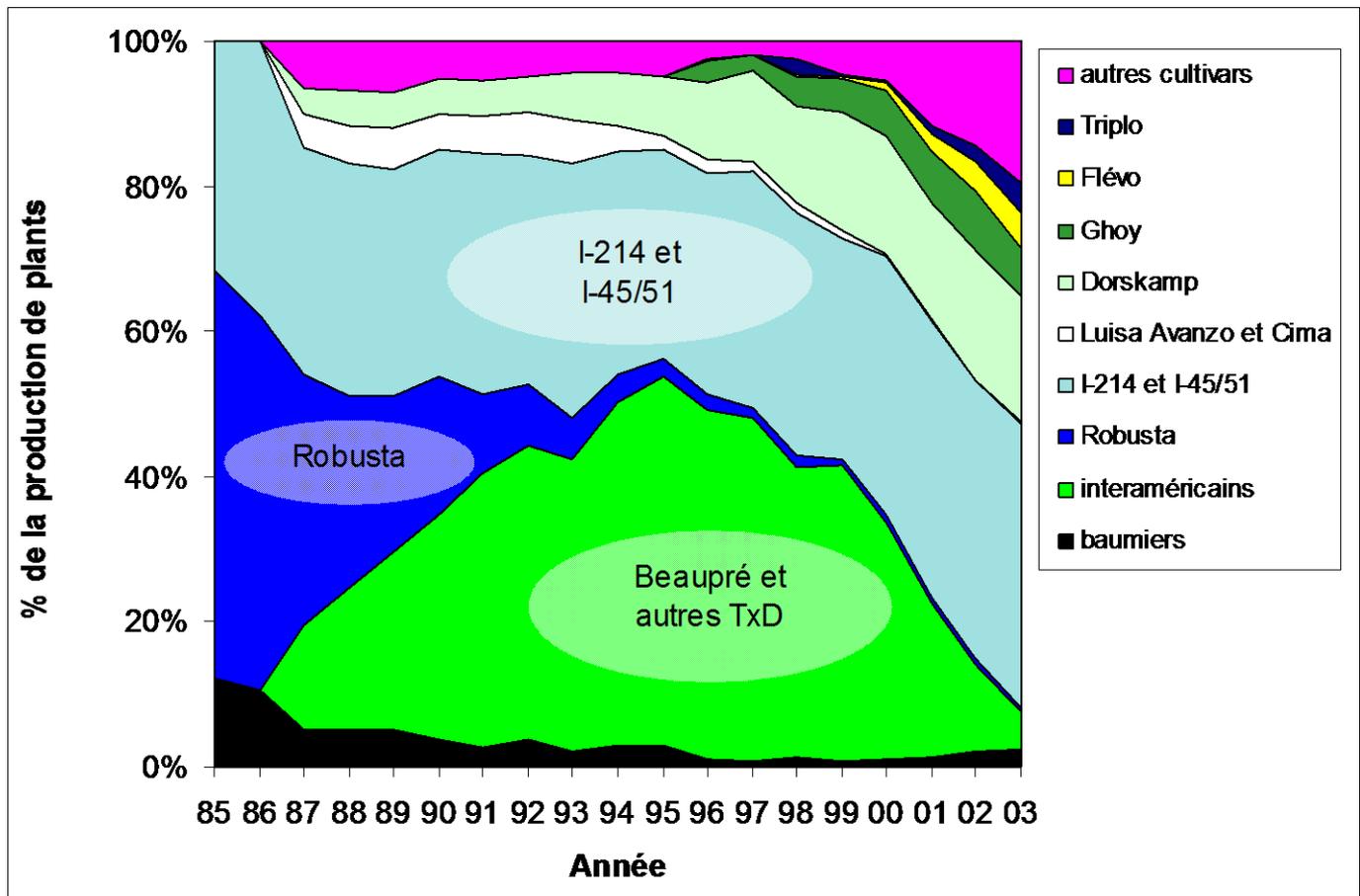


Figure 5 : Obtention de clones hybrides à partir de croisements d'espèces sauvages.



D'après Berthelot & Chantre, 2004

Figure 6 : Evolution des clones de peupliers plantés en France.

agronomique. Cependant, il représente aussi un important inconvénient : la vulnérabilité de la population face aux agresseurs biotiques (insectes, champignons, bactéries, virus, etc.) et abiotiques (vent, pollution, sécheresse, etc.).

Les peupleraies cultivées sont constituées à 90 % de clones hybrides qui sont issus de deux types de croisement. D'une part, les hybrides interaméricains qui sont issues d'un croisement entre les espèces *P. deltoides* originaire de l'est du continent nord américain (pour la rectitude de son tronc) et *P. trichocarpa* originaire de l'ouest continent nord américain (pour la rapidité de sa croissance et son enracinement), et d'autres hybrides euraméricains qui sont issues d'un croisement entre les espèces *P. deltoides* et *P. nigra* (pour sa rusticité) (figure 5). Ces croisements ont permis de combiner des caractéristiques agro-économiques importantes apportées par chacune des espèces et de bénéficier de l'effet de vigueur hybride (hétérosis). Ces hybrides ont été sélectionnés pour leur forme, leur productivité, la qualité du bois, leur stabilité vis à vis des variations climatiques et de leurs résistances aux agents pathogènes principalement à *Mlp*. Les clones de peupliers cultivés en France sont généralement ceux inscrits au « catalogue des espèces et variétés cultivées ». Il existe au catalogue européen 130 clones et au catalogue national 44 clones, dont 10 couvrent à eux seuls plus de 80 % de la surface plantée en France.

Les cultivars interaméricains « Beaupré », « Boelare », « Raspalje », « Unal », « Hunnegen » ont dominé le marché durant les années 1990 en raison de leur croissance, leur forme, leur bois et leur valeur marchande. Ils sont aujourd'hui remplacés par les hybrides euraméricains du fait de leur sensibilité moins élevée aux attaques de rouilles foliaires (figure 6).

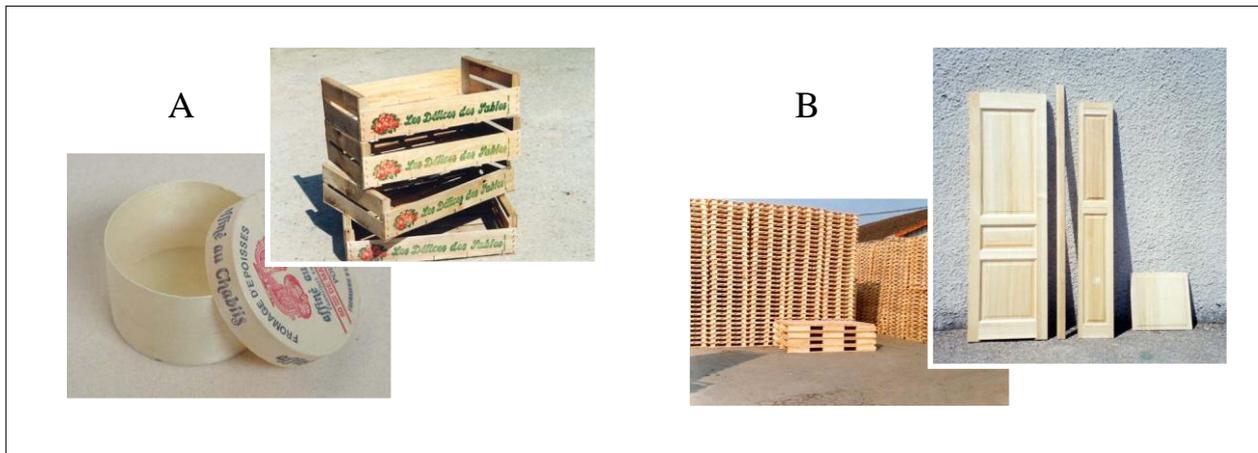
### **2.2.3 Les débouchés de la production de peuplier**

Le bois de peuplier a des usages variés :

- Le bois déroulé permet la fabrication de : contreplaqué, allumettes, confection d'emballage léger notamment pour l'alimentation (boîte à fromage, cagette pour fruits ou légumes,...). Le bois de peuplier a comme particularité importante de pouvoir être en contact avec l'aliment sans en modifier le goût (figure 7 A).

- Le bois scié est utilisé pour la confection de palettes, de caisses, d'ameublement et pour le bâtiment (figure 7 B).

- Seul le bois résiduel après extraction des grumes sert pour la pâte à papier. Le peuplier génère 750 000 m<sup>2</sup>/an de bois de "trituration", dont 500 000 pour le papier, ce qui représente 4 % du volume de bois utilisé par cette industrie.



**Figure 7 :** Exemples de débouchés pour le bois de peuplier.

A : bois déroulé, B : bois scié.

Cependant, dans certains pays, la consommation de bois de peuplier a diminué. Dans le secteur de l'emballage, par exemple, le bois est souvent remplacé par le carton et le plastique, et les allumettes par les briquets.

## 2.3 Un arbre modèle pour la génomique des arbres forestiers

### 2.3.1 Espèce modèle

*Arabidopsis thaliana* a été choisi comme espèce modèle pour la génomique des plantes. Les ressources et techniques développées pour cette espèce sont considérables pour comprendre la fonction de ces gènes. Cependant, on ne peut pas étudier avec ce modèle, les différents caractères spécifiques des arbres, tels que la formation du bois et les cycles saisonniers de croissance d'une espèce pérenne. Le peuplier permet aussi d'étudier une espèce végétale fortement hétérozygote comparé à celui d'*A. thaliana*.

En plus de sa valeur économique, le peuplier s'impose aujourd'hui comme une espèce modèle pour la génomique des arbres forestiers, de par ses caractéristiques biologiques et les ressources génomiques disponibles (Jansson & Douglas, 2007) :

- Son cycle de reproduction est relativement court pour un arbre forestier (6 à 10 ans), ce qui facilite l'obtention de croisements contrôlés.

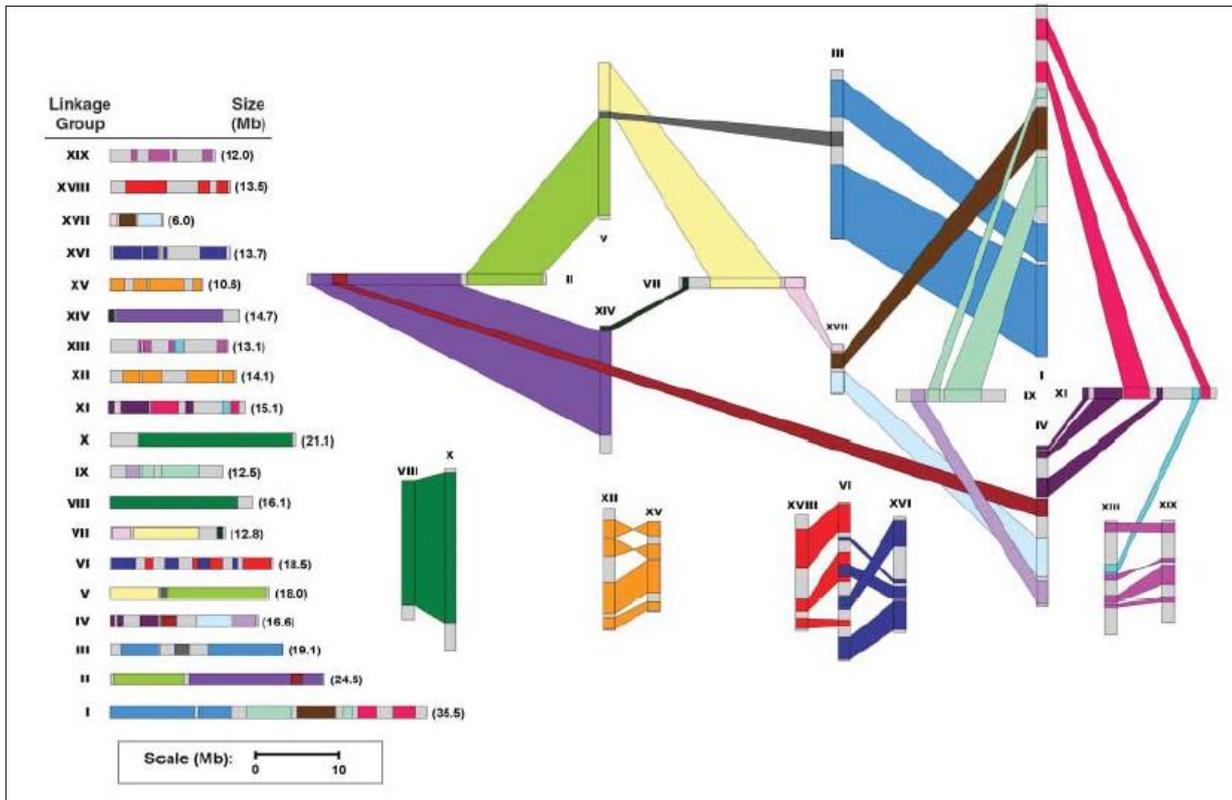
- La multiplication végétative est possible, ce qui facilite et optimise les analyses génétiques.

- La culture *in vitro* a été développée.

- Le genre *Populus* est diploïde avec  $2n = 38$  chromosomes, et son génome relativement petit : sa taille physique est estimée à 500 Mb (1 n) soit 4 fois le génome d'*A. thaliana* et 40 fois plus petit que celui du pin maritime.

- Les croisements contrôlés à l'intérieur d'une section et entre certaines des différentes sections, sont possibles, ce qui permet de créer des populations en ségrégation. Ceci a facilité l'élaboration de plusieurs cartes génétiques de différentes espèces (Paolucci *et al*, 2010 ; Drost *et al*, 2009 ; Pakull *et al*, 2009 ; Gaudet *et al*, 2008 ; Woolbright *et al*, 2007 ; Yin *et al*, 2004a ; Yin *et al*, 2002 ; Cervera *et al*, 2001). La cartographie comparée a permis de voir une bonne conservation et une bonne colinéarité entre les 19 chromosomes des différentes espèces. De même, plusieurs cartes de QTL ont pu être construites.

- Il existe, pour les principales espèces, des collections *in-situ* et *ex-situ*.



D'après Tuskan et al, 2006

**Figure 8 :** La duplication du génome entier du peuplier.  
 Les alignements des différentes couleurs représentent les segments chromosomiques homologues après la duplication du génome. Les mêmes couleurs sont utilisées sur la représentation des chromosomes à gauche.

- La transformation génétique est possible grâce à *Agrobacterium tumefaciens* (génotype dépendant).

- Plusieurs banques BAC sont disponibles pour les espèces *P. trichocarpa*, *P. tremuloides*, et un hybride *P. deltoides* x *P. nigra*.

- Dans la base d'EST (expressed sequence tag) du NCBI, 454 999 séquences sont accessibles pour différentes espèces et clones de peupliers (octobre 2011). Ces EST proviennent du clonage d'ADNc (ADN complémentaire) provenant de différents tissus, de différentes espèces ou d'hybrides, et de différents traitements. 89 943 ESTs sont issues de l'espèce *P. trichocarpa*, et beaucoup d'autres, issues d'hybrides entre *P. trichocarpa* et d'autres espèces du genre *Populus*.

- En 2002, le peuplier a été choisi pour être le premier génome séquencé chez les arbres forestiers (Tuskan *et al*, 2006).

### **2.3.2 La séquence du génome de l'individu Nisqually-1 *P. trichocarpa***

L'importance de cette ressource dans la recherche actuelle et dans ce travail de thèse a nécessité un chapitre plus amplement développé.

Le séquençage du génome a révélé un événement de duplication du génome entier (figure 8), qui se caractérise par 8 000 gènes dupliqués. La comparaison de ces gènes avec les EST des espèces du genre *Salix* a montré que cette duplication était antérieure à la séparation entre les deux genres, il y a 60 à 65 millions d'années. Des événements de réarrangement après cette duplication ont été observés, qui seraient antérieurs à la séparation entre les différentes espèces du genre *Populus* ; en effet, les cartes génétiques de différentes espèces de peuplier ont montré une bonne colinéarité.

#### **2.3.2.1 Premier assemblage**

##### **2.3.2.1.1 Assemblage v1**

L'individu Nisqually-1, femelle de l'espèce *P. trichocarpa*, individu diploïde et hétérozygote, a été choisi pour être séquencé au Joint Genome Institute (JGI, Walnut Creek, Californie USA). Ce séquençage a été effectué en whole genome shotgun (WGS) à hauteur de 7,5 X (Tuskan *et al*, 2006). L'assemblage de cette séquence (version v1) effectué avec le logiciel Jazz aboutit à une séquence de 486 Mb. Ce premier assemblage était disponible en 2004 sur le site web <http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>. Dix neuf scaffolds (dans cette version, appelés LG dans le "genome browser") représentent les 19 chromosomes pour un total de 308 Mb sur les 500 Mb environ du génome. La numérotation de référence pour ces



chromosomes est la numérotation donnée par Cervera *et al*, (2001). Il y a 21 993 scaffolds qui ne sont pas assignés aux 19 chromosomes représentant 178 Mb.

### **2.3.2.1.2 Annotation**

Deux versions différentes d'annotation automatique de ce premier assemblage ont été réalisées (v1.0 et v1.1). Ces annotations sont le résultat de l'utilisation de quatre logiciels de prédiction de gènes : Eugene, Fgenesh, GrailEXP6 et Genewise. Dans ce travail de thèse, c'est la deuxième version de l'annotation qui a été utilisée (v1.1). Cette annotation a conduit à 45 555 modèles de gènes codant des protéines. Dans le "genome browser" développé au JGI (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>), les gènes et les EST de peuplier sont alignés sur la séquence du génome. Les gènes non codants, tels que les miARNs, n'y sont pas précisés.

Dans un souci de simplicité, dans ce manuscrit, le premier assemblage v1 et la seconde annotation v1.1 qui est celle utilisée, sont nommés "V1".

### **2.3.2.2 Second assemblage**

#### **2.3.2.2.1 Assemblage v2**

Un second assemblage du séquençage du génome (v2) a été effectuée avec le logiciel ARACHNE. Ce second assemblage a été disponible sur le web début 2010 (<http://www.phytozome.net/poplar.php>). Jerry Tuskan (JGI) a aimablement donné accès au chromosome 19 dès le début 2009 à l'équipe dont je fais partie. Il y a 2 518 scaffolds, représentant 417 Mb (dans cette seconde version ils sont tous appelés scaffold), avec les 19 premiers scaffolds correspondants aux 19 chromosomes pour 378 Mb sur les 500 Mb environ du génome, les 2 499 scaffolds suivants ne sont pas assignés aux chromosomes, et représentent 39 Mb.

#### **2.3.2.2.2 Annotation v2.0**

L'annotation du second assemblage a été faite au JGI avec les logiciels FGESH+ et GenomeScan, avec l'aide de l'alignement des EST de peuplier. Dans l'annotation du second assemblage, il y a 45 778 modèles de gènes codants répartis dans 41 377 loci de gènes. Quarante-vingt quatorze % des modèles de gènes portent un codon d'initiation et de terminaison.

Dans ce manuscrit, le premier assemblage v2 et son annotation v2.0 sont nommés "V2".



## 3 Les rouilles foliaires des peupliers

### 3.1 Les maladies du peuplier

Diverses maladies touchent les peupleraies cultivées (Pinon & Frey, 2005), telles que le chancre bactérien causé par *Xanthomonas populi*, qui a été un problème majeur durant les années 1950 pour les peupleraies cultivées dans la partie nord de l'Europe. Durant les années 1960, la maladie majeure des peupleraies était la brunissure des feuilles causée par *Marssonina brunnea* après son introduction depuis l'Amérique du nord.

Dans le passé, la rouille foliaire n'était pas un problème car les infections étaient sporadiques, et seulement quand les conditions lui étaient très favorables. Cette maladie foliaire est devenue réellement un problème depuis les années 1980. Elle peut induire des pertes considérables de rendement chez les peupliers pouvant aller jusqu'à 50% (Paillassa, 2006). Les rouilles foliaires du peuplier sont causées par les champignons du genre *Melampsora*. La majorité des espèces de *Melampsora* sont hétéroïques : deux plantes hôtes sont indispensables au développement complet de la rouille (Laurans & Pilate, 1999). Treize espèces de *Melampsora* sont capables d'infecter le genre *Populus* dont 7 sont présentes en France, trois seulement infectent les peupleraies cultivées françaises :

- *Melampsora medusae* Thümen dont l'aire de répartition est limitée, en France, au sud-ouest du pays (Aquitaine et Midi-Pyrénées) et en Europe, à l'Espagne et au Portugal. Cette espèce est originaire de l'Amérique du nord. Il a plusieurs hôtes alternants, dont le mélèze. Jusqu'à présent il n'y a pas eu de dégât significatif dans les peupleraies cultivées, même si la plupart des cultivars y sont sensibles.

- *Melampsora allii-populina* Kleb., est présent dans l'ouest de la France. Ses hôtes alternes sont les plantes de la famille des *Aliaceae*. Depuis l'année 1997, cette espèce est en diminution devant l'importance des infections par *Melampsora larici-populina*.

- *Melampsora larici-populina* Kleb., est présent dans toute l'Europe, et particulièrement dans les régions du nord de la France. Aujourd'hui, *Mlp* est considéré comme l'un des dix plus importants problèmes sanitaires mettant en danger les peupleraies européennes. Cet agent pathogène, originaire d'Europe, s'est propagé en Australie et en Nouvelle-Zélande il y a 30 ans. Bien que rare, il a été détecté aux États-Unis il y a 10 ans, et plus récemment, en Islande et au Chili.

Les plus gros dégâts en Europe étant plus particulièrement causés par *Melampsora larici-populina* (*Mlp*), nous ne retiendrons par la suite que *Mlp*.

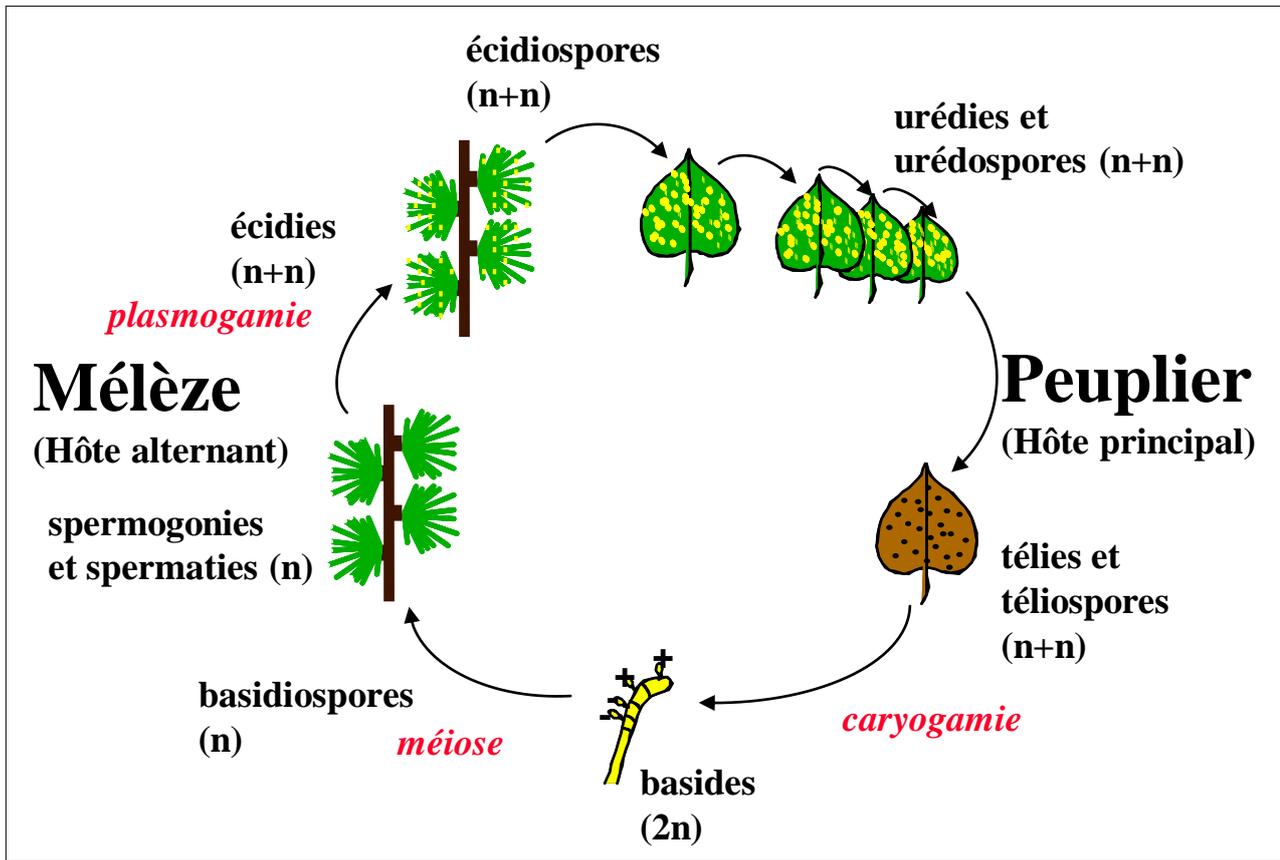


Figure 9 : Cycle de reproduction de *Melampsora larici-populina*.



Photo : J. Pinon, INRA

Figure 10 : Ecidies de *Mlp* sur un mélèze.

## 3.2 *Melampsora larici-populina*

*Melampsora larici-populina* Kleb. est un champignon Basidiomycète de la classe des Téliomycètes, de l'ordre des Urédinales et de la famille des *Melampsoraceae*.

### 3.2.1 Le cycle de *Mlp*

*Mlp* a un cycle hétéroïque (figure 9) combinant une phase de reproduction sexuée sur un hôte alterne : le mélèze (figure 10), et une phase de reproduction asexuée sur son hôte principal : le peuplier.

De début mai à mi-juin, et durant tout l'été, le champignon se développe sur les feuilles de peuplier (figure 9 ; Frey & Pinon, 2004). Le champignon est au stade de l'urédie, pendant lequel il est sous forme dicaryotique ( $n + n$  chromosome). Les urédospores germent à la surface des feuilles, le champignon pénètre dans la feuille par les stomates et envahit le parenchyme foliaire. Après une semaine environ, les urédies apparaissent sur la face inférieure des feuilles. Les urédies produisent de grandes quantités d'urédospores (2 500 à 5 000 urédospores par urédie et par jour), qui sont disséminées par le vent et infectent d'autres feuilles de peuplier. Si les conditions sont favorables, une dizaine de cycles de multiplication asexuée peuvent se succéder pendant l'été. Pendant l'automne (chute des températures extérieures), le champignon évolue sous forme de télies contenant des téliospores ( $n + n$ ), une forme résistante qui lui permet d'hiverner sur les feuilles tombées au sol. Il prend alors l'aspect de « croûtes noirâtres » sur la face supérieure des feuilles. Au printemps, à partir de fin février, la caryogamie (fusion des deux noyaux) se produit dans les téliospores, qui émettent alors des basides diploïdes ( $2n$ ). Ces basides subissent immédiatement une méiose et produisent des basidiospores haploïdes ( $n$ ), qui sont disséminées à faible distance par le vent. Ces basidiospores colonisent un hôte alterne au moment de son débourrement et forme des fructifications : les spermogonies. Ces spermogonies produisent des spermaties haploïdes ( $n$ ). Il se produit alors la fusion des spermaties : c'est la plasmogamie (fusion de deux cytoplasmes), qui aboutit à la formation d'un mycélium dicaryotique ( $n + n$ ). Ce mycélium forme une fructification sur l'aiguille de mélèze, l'écidie, qui produit des écidiospores dicaryotique ( $n + n$ ). Enfin ces écidiospores se disséminent par le vent jusqu'à plusieurs kilomètres pour s'installer sur son hôte principal, le peuplier. Débute alors la phase épidémique du cycle, où le champignon colonise successivement plusieurs peupliers.

*Mlp* semble incapable de survivre sous forme d'urédospores d'une année sur l'autre. Le passage par l'hôte alternant est donc obligatoire. Or, pendant cette phase de reproduction sexuée,



des possibilités de recombinaison génétique existent et favorisent l'apparition de nouveaux pathotypes, ce qui confère une très grande adaptabilité au champignon.

### **3.2.2 Symptômes de la maladie chez le peuplier**

Les urédies ou sores présentes sur la face inférieure de quelques feuilles, peuvent couvrir la totalité de la surface des feuilles en l'espace d'une à quatre semaines selon les conditions météorologiques et les cultivars, puis l'infection s'étend à l'arbre entier.

*Mlp* puise grâce à des suçoirs (haustoria), qui passent par les stomates de la feuille (Laurans & Pilate, 1999), les nutriments essentiels à sa survie (composés carbonés et azotés) au détriment de son hôte, perturbant ainsi le métabolisme de ce dernier. Ces perturbations entraînent une réduction de l'activité photosynthétique avec une désorganisation des chloroplastes et une activité respiratoire anormalement élevée. En effet, ce phénomène accroît la production d'énergie nécessaire à la synthèse de composés impliqués dans les réactions de défense de la plante et la translocation des composés carbonés et azotés aux sites d'infection. Les réserves des peupliers ne seront pas suffisantes pour un bon débourrement l'année suivante.

### **3.2.3 La diversité chez *Melampsora larici-populina***

La variabilité de pathogénécité de *Mlp* est détectée à deux niveaux : la virulence et l'agressivité.

Dans le pathosystème peuplier-*Mlp*, chaque souche de *Mlp* peut être définie selon ses virulences. Ici une virulence est définie par la possibilité de contourner un gène de résistance qualitatif d'un cultivar particulier qui était à l'origine résistant. A l'heure actuelle, huit virulences différentes ont pu être découvertes chez *Mlp*, définies par huit cultivars différentiels discriminants (Pinon & Frey, 2005). Jusqu'à présent, il n'a pas été possible de connaître le déterminisme génétique de ces virulences ; il ne peut donc pas être défini un gène d'avirulence pour chacune des virulences. Chaque souche de *Mlp* présente une combinaison de plusieurs virulences qu'il est possible de connaître en l'inoculant sur des disques foliaires de la gamme différentielle de cultivars discriminants. On appelle pathotype l'ensemble des souches qui possèdent la même combinaison de virulence. Avec huit virulences, il y a un potentiel de 256 pathotypes. A ce jour, une soixantaine de pathotypes a été identifiée dans la nature.

Il a été observé aussi une variabilité sur des critères quantitatifs de leur pouvoir infectieux (vitesse d'apparition des symptômes, le nombre d'urédosores et d'urédospores produits) indépendamment de la combinaison en virulence des souches de *Mlp*. Ceci montre une



différence d'agressivité très variable entre les souches de même pathotype envers le peuplier (Pinon & Frey, 2005 ; Dowkiw & Bastien, 2004 ; Dowkiw, 2003).

Une souche ou isolat de champignon est une mono-urédie que l'on a multipliée de façon végétative afin d'en obtenir une grande quantité. Pour *Mlp*, chaque isolat est défini par un nom et son année de récolte. On procède ainsi pour avoir des conditions d'expériences contrôlées afin de permettre la reproductibilité des expériences que l'on réalise. En condition naturelle, les isolats de *Mlp* sont mélangés.

### **3.3 *Mlp* dans les peupleraies cultivées**

#### **3.3.1 Les dégâts**

Les dégâts causés par *Mlp* dépendent essentiellement de la précocité et de la sévérité de l'attaque. Lorsque les défeuillaisons sont précoces (juillet) et fortes, elles provoquent une réduction de la croissance en diamètre et en hauteur (jusqu'à 30%) pendant l'année en cours et a des conséquences l'année suivante. En revanche, une infection mi-septembre a très peu d'effet sur la croissance mais entraîne des dérèglements physiologiques susceptibles de perturber l'aoûtement et la constitution des réserves en fin de saison de végétation, ce qui peut entraîner un débourrement tardif et/ou une sensibilité accrue aux gelées tardives du printemps suivant. La présence de l'hôte alternant, le mélèze, à proximité de la peupleraie facilite la réalisation du cycle de développement du parasite, et risque d'accroître la précocité et la gravité de l'attaque. Il est donc un facteur de risque important.

La sensibilité d'une peupleraie aux rouilles dépend grandement des cultivars qui la composent. Jusqu'à présent, pour la résistance à *Melampsora*, les stratégies de sélection du peuplier étaient exclusivement orientées vers la valorisation de résistances qualitatives (également appelées totales ou complètes) apportées par l'espèce *P. deltoides*. Ces résistances introduites contre *Melampsora larici-populina*, ont été rapidement contournées par de nouvelles souches de *Mlp* portant de nouvelles virulences. Les peupliers restent résistants aux inoculum naturels pendant les tests de la sélection, mais cette résistance est contournée quelques années plus tard par de nouvelles virulences de l'agent pathogène (Pinon & Frey, 1997). L'apparition de ces nouveaux pathotypes est la conséquence du haut potentiel adaptatif du champignon, mais aussi de la culture monoclonale des peupliers cultivés qui renforce la pression de sélection sur les agents pathogènes. Ces contournements de résistance montrent les limites de ces stratégies de sélection et conduisent à réorienter la création variétale vers la production de cultivars présentant une résistance durable.



Les conditions climatiques printanières et estivales jouent également un rôle considérable dans la dissémination, la contamination et le développement des rouilles. Des conditions sèches et chaudes ( $> 25^{\circ}\text{C}$ ) peuvent bloquer l'évolution de la maladie, alors que des températures moyennes et une forte hydrométrie sont, au contraire, susceptibles de provoquer son explosion. Le parasite est également favorisé par les fonds de vallée où le brouillard persiste et crée un microclimat humide (Pinon & Valadon, 1997).

D'autres facteurs tels que les excès d'azote et les déficits en potassium pourraient aussi rendre les peupliers plus sensibles à ces champignons pathogènes. De plus, les défeuillaisons répétées peuvent entraîner des affaiblissements capables de favoriser l'installation et le développement d'autres parasites.

### **3.3.2 Exemple d'un cultivar : "Beaupré"**

De grandes surfaces de peupleraies ont été plantées en clone Beaupré entre 1990 et 2000. En France, plus de 50% des peupliers plantés en 1997 étaient des clones Beaupré, proportion qui passe à 80% si l'on se restreint à la moitié nord de la France. Depuis 1997, ces peupleraies, et en particulier celles du quart nord-est de la France, sont attaquées chaque année par la rouille (Paillassa, 2006). Ceci est dû à l'apparition en 1994 d'une nouvelle virulence, la virulence 7 (Cf. § I-3.2.3), qui contourne la résistance qualitative introduite chez ce cultivar. La perte de croissance en circonférence est de l'ordre de 30% pour un clone Beaupré malade comparé à un Beaupré sain (Paillassa, 2006). Cette perte de rendement, par l'augmentation du temps de rotation et de la moindre qualité du bois, rend cette culture non rentable. En 2002, la proportion de clone Beaupré a chuté à moins de 3%.

A chaque fois, les crises sanitaires trouvent leur origine dans la monoculture : les clones d'un même cultivar réagissent de la même façon à l'infection. La maladie se propage très vite, un seul arbre atteint peut contaminer la peupleraie entière. Devant cette situation, Lefèvre *et al*, (1994) ont proposé l'utilisation d'une diversité de clone plus importante notamment avec une diversité des gènes de résistance introduite.

### **3.3.3 Les luttes**

#### **3.3.3.1 Lutte préventive**

La lutte préventive s'appuie sur quatre actions :

- Il faut bien choisir des cultivars adaptés aux conditions du lieu de plantation et de tolérance à la rouille foliaire.



- Il faut limiter l'humidité au sein des parcelles et assurer la ventilation en élaguant régulièrement les arbres.
- Il faut enterrer les feuilles tombées au sol pendant l'hiver.
- Il faut éviter la présence de mélèzes à proximité des parcelles.

### **3.3.3.2 La lutte chimique**

La lutte chimique se fait par application de fongicides par canon atomiseur ou par hélicoptère. Les fongicides ont un effet préventif et non curatif : il convient donc de traiter dès l'apparition des sores. Mais cette lutte chimique engendre un coût supplémentaire d'exploitation.

### **3.3.4 Problème sanitaire prévisible des cultures de peuplier.**

Le risque d'avoir un problème sanitaire important était prévisible en raison de plusieurs paramètres :

- Le flux de gènes important dans l'espace car les urédospores sont transportées sur de longues distances par le vent pouvant atteindre plusieurs centaines de kilomètres.
- Le cycle biologique de *Mlp* combine la reproduction sexuée qui provoque l'apparition de nouveaux pathotypes, et la reproduction asexuée qui amène à une multiplication rapide des souches les mieux adaptées.
- Les cultures monoclonales avec une très faible diversité de résistance entre les différentes plantations provoquent une forte pression de sélection sur *Mlp*.
- Les résistances apportées ont été uniquement des résistances qualitatives et aucune résistance quantitative n'a été introduite ; il y a donc eu l'invasion de nouvelles souches contournant ces résistances qualitatives, sans possibilité de ralentissement par une seconde ligne de défense.

### **3.3.5 Les pistes pour des résistances durables**

Une résistance est durable si elle demeure efficace pendant une utilisation prolongée dans un environnement favorable à l'agent pathogène. Dans le cas d'un arbre forestier, ce temps de durabilité doit être en adéquation avec le long temps de rotation de la culture d'une espèce pérenne. Il est donc demandé un temps de résistance bien plus importante que pour des espèces annuelles.

La fréquence des virulences nécessaires à l'infection de certains cultivars a augmenté rapidement. Malheureusement, contrairement aux productions agricoles annuelles, pour lesquelles on peut remplacer rapidement un gène de résistance par un autre, la pérennité du



peuplier rend cette stratégie caduque. Ce constat renforce l'intérêt d'une stratégie d'amélioration du peuplier basé sur des résistances plus durables.

### **3.3.5.1 Résistance quantitative**

Au-delà des interactions qualitatives, il existe également des interactions quantitatives entre la rouille et le peuplier. En effet, dans le cas d'une interaction compatible entre un isolat de *Mlp* et un cultivar de peuplier, les dégâts causés par la maladie peuvent varier, en fonction de l'agressivité de l'isolat et en fonction de la résistance partielle du cultivar. La résistance partielle, contrairement à la résistance complète, n'empêche pas la maladie, mais elle en réduit l'incidence, et elle est considérée en général comme efficace contre tous les isolats de l'agent pathogène. De plus, son déterminisme génétique est souvent polygénique, donc a priori, moins facilement contournable par l'agent pathogène. Le premier changement dans cette direction est une augmentation de la proportion des cultivars euraméricains, qui ont une meilleure résistance générale (Cf. figure 6, page I-15). Mais ces résistances quantitatives seront-elle réellement plus durables ?

## **3.4 Un pathosystème modèle**

Le pathosystème peuplier-*Mlp* a été choisi par plusieurs laboratoires comme modèle d'études des interactions arbre-microorganismes. De nombreuses connaissances ont été accumulées sur le pathosystème peuplier-*Mlp* ces dernières années, mais il reste cependant beaucoup de zones d'ombre. Le séquençage du génome du peuplier et le futur séquençage du génome de *Melampsora larici-populina* ouvre des perspectives extraordinaires sur les compréhensions des mécanismes moléculaires qui régissent l'interaction entre ces deux organismes. Ce pathosystème constitue donc un excellent modèle scientifique, associé à un réel enjeu économique pour l'étude des interactions arbre-microorganismes (Frey & Pinon, 2004).

# **4 Concepts de la résistance des plantes aux agents pathogènes**

Afin de mieux comprendre les interactions entre le peuplier et *Mlp*, ce chapitre regroupe les bases de la résistance des plantes aux agents pathogènes.

Les plantes ne peuvent pas se déplacer pour échapper aux agressions extérieures, par conséquent elles doivent subir les stress notamment biotiques dont sont responsables de



nombreux types d'agents pathogènes : virus, mycoplasmes, bactéries, champignons, nématodes, insectes, protozoaires etc.

Les plantes ont élaboré au cours de leur évolution différentes stratégies qui leur permettent de se défendre contre les agents pathogènes, si bien que la résistance chez les plantes est la règle, et la sensibilité l'exception.

## **4.1 Quelques définitions**

### **4.1.1 Agents pathogènes nécrotrophes et biotrophes**

Les agents pathogènes des plantes peuvent être divisés en deux catégories (Dangl & Jones, 2001) : ceux qui tuent l'hôte et s'en nourrissent (nécrotrophes) et ceux qui ont besoin d'un hôte vivant pour terminer leur cycle de vie (biotrophes), c'est le cas pour le pathosystème peuplier-*Mlp*.

### **4.1.2 Résistance non-hôte et hôte**

La résistance est qualifiée de non-hôte quand tous les génotypes d'une espèce donnée sont résistants à toutes les souches d'un agent pathogène donné, ce niveau de résistance est associé à des mécanismes de défense passive. Les barrières physiques et chimiques constitutives telles que la cuticule, la paroi cellulaire et les composés antimicrobiens produits constitutivement, protègent la plante.

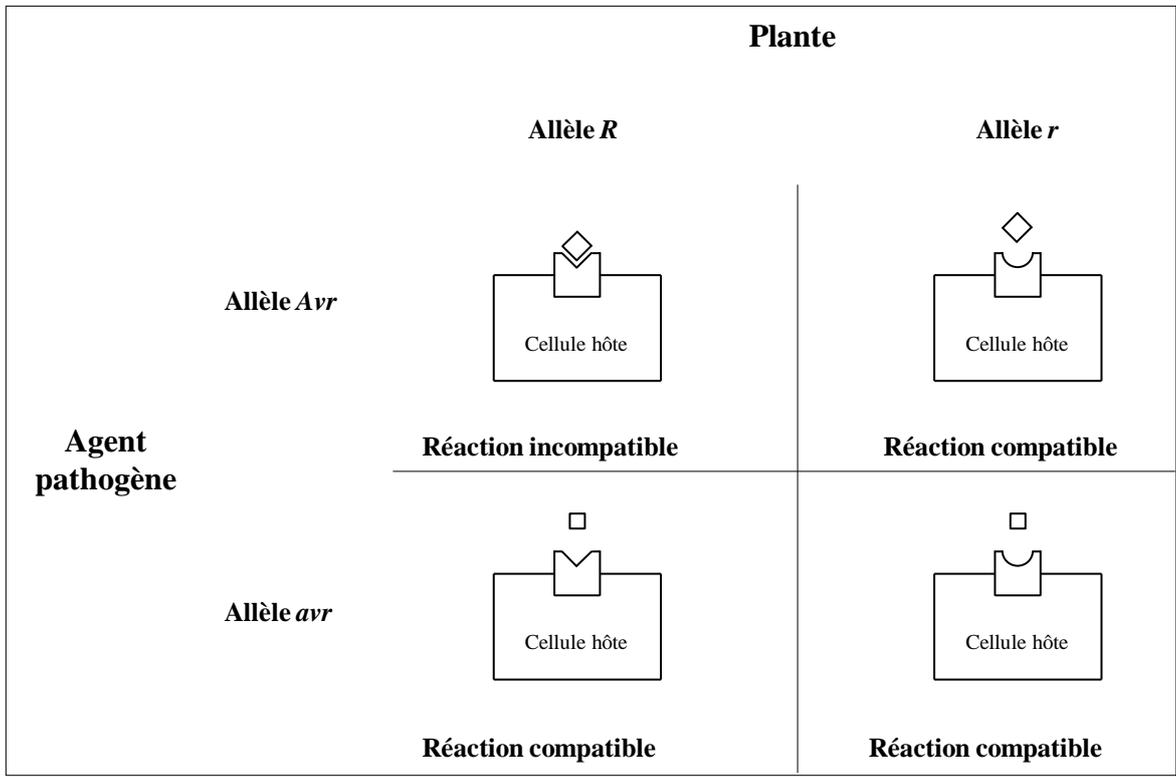
La résistance est qualifiée d'hôte lorsque l'agent pathogène est capable d'infecter une espèce donnée mais que certains génotypes de l'espèce sont résistants.

## **4.2 Résistance constitutive**

Pour pouvoir se nourrir, les organismes phytopathogènes doivent accéder aux cellules des plantes. Les premières barrières que doivent contourner ces organismes sont les barrières physiques et chimiques constitutives de la plante. Cette résistance est passive car elle n'implique pas une reconnaissance de l'agent pathogène par la plante. Les barrières sont les cuticules, les parois cellulaires, les enzymes antimicrobiennes et les métabolismes secondaires préformés.

## **4.3 Résistance induite**

Lorsque ces barrières sont franchies, les plantes déclenchent des mécanismes de résistance induite.



**Figure 11** : Illustration du modèle gène à gène.

La perception de l'agent pathogène par la plante hôte active un réseau de signaux moléculaires dont la finalité est l'induction des mécanismes de défense. Trois phases se distinguent : la reconnaissance de l'agent pathogène, les événements de signalisation et enfin les réactions de défense.

### 4.3.1 Le modèle gène à gène

Le concept du modèle gène à gène énoncé par Flor, (1971), en étudiant le pathosystème lin/rouille, implique un gène de résistance (R) de la plante codant une protéine, qui reconnaît un effecteur codé par un gène d'avirulence (Avr) de l'agent pathogène (figure 11).

En réalisant des croisements entre les variétés de lin, Flor a montré que la résistance était mono-génique et dominante par rapport à la sensibilité. De façon similaire, en croisant des isolats de *Melampsora lini*, il a montré que l'avirulence était mono-génique et dominante par rapport à la virulence.

- Quand l'allèle résistant du gène R est présent chez la plante, et que l'allèle Avr est présent chez l'agent pathogène, il y a reconnaissance du gène Avr par l'allèle R : il y a induction des mécanismes de défense et l'infection est inhibée (réaction incompatible). L'absence de maladie est le résultat d'une interaction incompatible entre l'hôte et l'agent pathogène, l'agent pathogène est avirulent et l'hôte est résistant.

- Quand l'allèle résistant du gène R est présent chez la plante, et que l'allèle Avr est absent chez l'agent pathogène, il n'y a pas de reconnaissance de l'agent pathogène : il n'y a pas induction des mécanismes de défense et donc il y a une réaction compatible. La maladie est le résultat d'une interaction compatible entre l'hôte et l'agent pathogène, l'agent pathogène est virulent et l'hôte est sensible.

- Quand l'allèle résistant du gène R n'est pas présent chez la plante, et que l'allèle Avr est présent chez l'agent pathogène, il n'y a pas de reconnaissance de l'agent pathogène : il n'y a pas induction des mécanismes de défense et il y a une réaction compatible.

- Quand l'allèle résistant du gène R n'est pas présent chez la plante, et que l'allèle Avr est absent chez l'agent pathogène, il n'y a pas de reconnaissance de l'agent pathogène : il n'y a pas induction des mécanismes de défense et il y a une réaction compatible.

Flor en a conclu qu'à un gène de résistance dominant chez la plante correspond un gène d'avirulence dominant chez le champignon. Ce modèle implique donc une spécificité du génotype de la plante et de l'agent pathogène.



Cette relation, appelée gène-pour-gène, a ensuite été suggérée dans de nombreux pathosystèmes, avec des agents pathogènes et des ravageurs divers (champignons, bactéries, virus, nématodes, insectes).

Deux types de résistance induite peuvent être distingués suivant la reconnaissance de l'agent pathogène, (i) la reconnaissance d'éléments constitutifs à l'agent pathogène (éliciteurs) et (ii) la reconnaissance des effecteurs de l'agent pathogène.

### **4.3.2 Résistance par reconnaissance d'éléments constitutifs de l'agent pathogène**

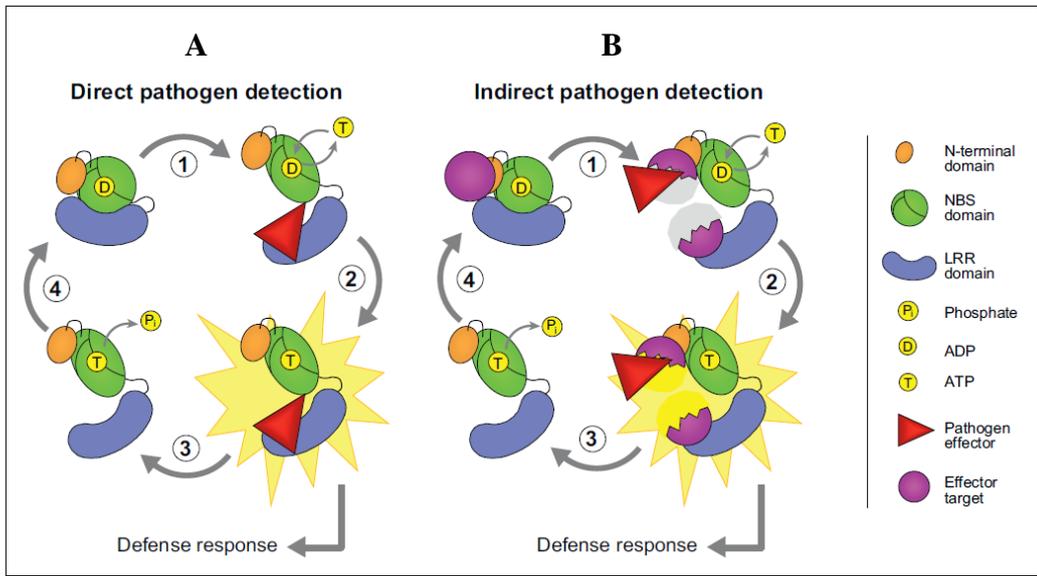
#### **4.3.2.1 La reconnaissance de l'agent pathogène**

La reconnaissance d'un agent pathogène est faite par des récepteurs des plantes, capables de reconnaître des éléments conservés chez les agents pathogènes (éliciteurs), appelés pathogen-associated molecular patterns (PAMPs), ceux-ci ont été changés en microbe-associated molecular patterns (MAMPs), car ces éléments conservés ne sont pas spécifiques des agents pathogènes et sont présents aussi chez de nombreux microorganismes qui n'ont aucun pouvoir pathogène (Boller & Felix, 2009). Un MAMP est une molécule indispensable au micro-organisme, qui peut être détectée par des récepteurs des plantes appelés pattern recognition receptor (PRR). Ce sont des protéines transmembranaires qui signalent à la cellule végétale la présence des MAMPs dans l'espace extracellulaire. Les PRRs provoquent une cascade de signalisation qui permet à la plante de se protéger contre l'agent pathogène, c'est ce qu'on appelle la PAMP-triggered immunity (PTI).

Les plantes reconnaissent aussi des molécules de dégradations cellulaires issues des lésions provoquées par les agents pathogènes. Ces éliciteurs sont appelés damage-associated molecular pattern (DAMP). De même que pour les MAMPs, ils sont reconnus par des PRRs (Boller & Felix, 2009).

#### **4.3.2.2 Contournement de la PTI**

Pour contourner la résistance induite par la PTI, les agents pathogènes sont capables de sécréter une série de molécules, appelées effecteurs, qui inhibent les voies de signalisation induites par la reconnaissance de l'agent pathogène. Ces effecteurs sont des protéines de structure et de fonction très diverses et peuvent agir dans l'espace inter et intracellulaire. On parle d'effector-triggered susceptibility (ETS), lorsque l'agent pathogène inhibe la PTI par l'intermédiaire de ses effecteurs : la plante est alors infectée.

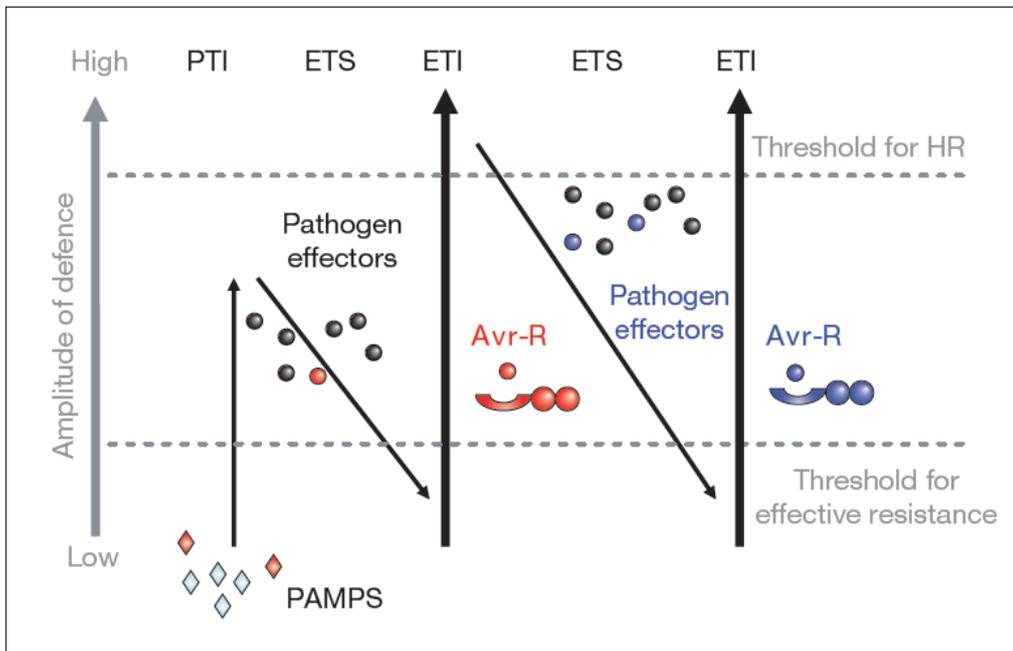


D'après Bent & Mackey, 2005

**Figure 12 :** Modèle ligand-récepteur et modèle de garde.

A : modèle ligand-récepteur : le gène R reconnaît directement l'effeteur de l'agent pathogène.

B : modèle de garde : le gène R reconnaît la modification de la protéine qui est ciblée par l'effeteur de l'agent pathogène.



D'après Jones & Dangl, 2006

**Figure 13 :** Le modèle en zig-zag, les différentes couches de résistance.

### 4.3.3 Résistance par reconnaissance des effecteurs de l'agent pathogène

En réponse à l'ETS, les plantes ont développé un autre système de défense stimulé par la présence des effecteurs. C'est l'effector-triggered immunity (ETI) (Chisholm *et al*, 2006).

#### 4.3.3.1 Interactions entre protéine R et effecteur

Une reconnaissance directe entre le gène R et le gène d'Avr est appelé modèle ligand-récepteur (figure 12 A). Ce type d'interaction est très peu décrit dans la littérature, l'exemple le plus récent est la reconnaissance directe entre les formes alléliques L5 et L6 d'un gène R du lin avec l'effecteur AvrL567 de *Melampsora lini* (Dodds *et al*, 2004). La détection de l'effecteur par la protéine R peut se faire aussi de façon indirecte ; l'effecteur cible une protéine de la plante, ce qui a pour conséquence sa modification, celle-ci est reconnue par la protéine R. Ce modèle est appelé modèle de garde (figure 12 B ; Bent & Mackey, 2007 ; Jones & Dangl, 2006).

#### 4.3.3.2 Le modèle en zigzag

Le modèle en zigzag proposé par Jones & Dangl, 2006 (figure 13) est le modèle admis à ce jour pour décrire les différents niveaux de résistance. Il comporte différentes phases résumées ci-dessous :

- Les plantes reconnaissent les MAMPs des agents pathogènes grâce aux PRRs et enclenchent la PTI : la plante est résistante (Cf. § I-4.3.2.1). Cette réaction est de faible amplitude, mais suffisante pour déjouer l'attaque de l'agent pathogène.
- Grâce à leur effecteur, certains agents pathogènes arrivent à contourner la PTI et à infecter la plante : c'est l'ETS, la plante est sensible (Cf. § I-4.3.2.2).
- Certaines plantes sont capables de reconnaître spécifiquement la présence directe ou indirecte de l'agent pathogène grâce à leurs protéines de résistance, ce qui déclenche la PTI : la plante est résistante (Cf. § I-4.3.3).
- Les agents pathogènes évoluent avec de nouveaux effecteurs pour échapper à l'ETI et inhiber la PTI. Les plantes elles aussi évoluent afin de reconnaître les nouveaux effecteurs avec de nouveaux gènes R. Il s'agit donc d'une coévolution entre la plante et l'agent pathogène, une course entre les armes d'attaque et de défense. Récemment, Boller & Felix, (2009) ont proposé un modèle généraliste dans lequel les agents pathogènes sont reconnus par des protéines différentes : la reconnaissance déclenche alors un signal de danger universel.

**Tableau 2** : Liste non exhaustive de gènes R clonés chez les végétaux.

Gène	Classe	Espèce	Agent pathogène	Référence
Bs2	CC-NBS-LRR	Poivron	<i>Xanthomonas campestris</i>	Tai et al, 1999
Cf 2/5	RLP	Tomate	<i>Cladosporium fulvum</i>	Dixon <i>et al</i> , 1996; Dixon <i>et al</i> , 1998
Cf 4/9	RLP	Tomate	<i>Cladosporium fulvum</i>	Jones <i>et al</i> , 1994
Dm3	CC-NBS-LRR	Laitue	<i>Bremia lactucae</i>	Shen et al, 2002
Gpa2	CC-NBS-LRR	Pomme de terre	<i>Globodera pallida</i>	Van der Vossen et al, 2000
HcrVf2	RLP	Pomme	<i>Ventria inaequalis</i>	Belfanti et al, 2004
Hero	CC-NBS-LRR	Tomate	<i>Globodera</i>	
HRT	CC-NBS-LRR	Arabidopsis	TCV	
Hm1	Toxine réductase	Maïs	<i>Cochliobolus</i>	
I2	CC-NBS-LRR	Tomate	<i>Fusarium oxysporum</i>	Ori et al, 1997
L6	TIR-NBS-LRR	Lin	<i>Melampsora lini</i>	Lawrence et al, 1995
M	TIR-NBS-LRR	Lin	<i>Melampsora lini</i>	Anderson et al, 1997
Mi	CC-NBS-LRR	Tomate	<i>Macrosiphum euphorbiae</i>	Milligan et al, 1998
Mi	CC-NBS-LRR	Tomate	<i>Meloidogyne incognita</i>	Rossi et al, 1998
Mla-1	CC-NBS-LRR	Orge	<i>Blumeria graminis</i>	Zhou et al, 2001
Mla-6	CC-NBS-LRR	Orge	<i>Blumeria graminis</i>	Halterman et al, 2001
Mlo	proteine membranaire	Orge	<i>Erysiphe graminis</i>	Buschges et al, 1997
N	TIR-NBS-LRR	tabac	<i>Virus mosaïque du tabac</i>	Whitham et al, 1994
P	TIR-NBS-LRR	Lin	<i>Melampsora lini</i>	
Pi-b	CC-NBS-LRR	Riz	<i>Magnaporthe grisea</i>	Wang et al, 1999
Pi-ta	CC-NBS-LRR	Riz	<i>Magnaporthe grisea</i>	Bryan <i>et al</i> , 2000
Prf	CC-NBS-LRR	Tomate	<i>Pseudomonas</i>	
Pto	STK	Tomate	<i>Pseudomonas syringae</i>	Martin et al, 1993
R1	CC-NBS-LRR	Pomme de terre	<i>Phytophthora infestans</i>	Collins et al, 1999
Rp1	CC-NBS-LRR	Maïs	<i>Puccinia</i>	
Rpg1	2STK en tandem	Orge	<i>Puccinia</i>	
Rpg1-b	CC-NBS-LRR	Soja	<i>Pseudomonas syringae</i>	Ashfield et al, 2004
RPM1	CC-NBS-LRR	Arabidopsis	<i>Pseudomonas syringae</i>	Grant et al, 1995
RPP1	TIR-NBS-LRR	Arabidopsis	<i>Hyaloperonospora parasitica</i>	Botella et al, 1998
RPP13	CC-NBS-LRR	Arabidopsis	<i>Hyaloperonospora parasitica</i>	Bittner Eddy et al, 2000
RPP27	RLP	Arabidopsis	<i>Hyaloperonospora parasitica</i>	Tor et al, 2004
RPP4	TIR-NBS-LRR	Arabidopsis	<i>Hyaloperonospora parasitica</i>	Van der Biezen et al, 2002
RPP5	TIR-NBS-LRR	Arabidopsis	<i>Hyaloperonospora parasitica</i>	Parker et al, 1997
RPP8	CC-NBS-LRR	Arabidopsis	<i>Hyaloperonospora parasitica</i>	McDowell et al, 1998
RPS1	TIR-NBS-LRR	Arabidopsis	<i>Ralstonia solanacearum</i>	Deslandes et al, 2002
RPS2	CC-NBS-LRR	Arabidopsis	<i>Pseudomonas syringae</i>	Bent et al, 1994
RPS4	TIR-NBS-LRR	Arabidopsis	<i>Pseudomonas syringae</i>	Gassmann et al, 1999
RPS5	CC-NBS-LRR	Arabidopsis	<i>Pseudomonas syringae</i>	Warren et al, 1998
Rx1	CC-NBS-LRR	Pomme de terre	virus de la pomme de terre	Bendahmane et al, 1999
Rx2	CC-NBS-LRR	Pomme de terre	virus de la pomme de terre	
Sw-5	CC-NBS-LRR	Tomate	<i>Tospovirus</i>	
Ve1 et Ve2	RLP	Tomate	<i>Verticillium</i>	Vinatzer et al, 2001
Xa1	CC-NBS-LRR	Riz	<i>Xanthomonas oryzae</i>	Yoshimura et al, 1998
Xa21	RLK	Riz	<i>Xanthomonas oryzae</i>	Song et al, 1995
XA26	RLK	Riz	<i>Xanthomonas oryzae</i>	Sun et al, 2004
Xa27	-	Riz	<i>Hyaloperonospora parasitica</i>	Gu et al, 2005

#### 4.3.4 Evènements de signalisation et réactions de défense

Les voies de signalisation impliquées lors de la PTI et l'ETI sont généralement les mêmes. Il s'agit principalement de phosphatases et kinases, intervient aussi une régulation d'hormones notamment l'acide jasmonique, l'acide salicylique et l'éthylène, et des mouvements transmembranaires d'ions. La contribution de chacune des voies dépend de la nature de l'agent pathogène ainsi que celle du récepteur impliqué dans sa reconnaissance.

Ces voies de signalisation aboutissent à la mise en place d'un ensemble complexe de réactions de défense en vue de bloquer le développement de l'agent pathogène. Il y a un renforcement des parois cellulaires et une accumulation de PRR. Il peut y avoir une réponse d'hypersensibilité (HR). Cette réponse consiste en une mort cellulaire localisée aux points d'infection et qui a pour objectif de limiter la croissance de l'agent pathogène en le privant d'un moyen de multiplication.

## 5 Les gènes de résistance

Depuis le milieu des années 1990, plusieurs dizaines de gènes de résistance ont été isolés et séquencés (tableau 2), ce qui a permis d'étudier leurs séquences protéiques et d'en étudier leur distribution dans les génomes nucléaires des plantes. La plupart de ces gènes ont été isolés par clonage positionnel, par transposon-tagging ou à l'aide de mutants (Hulbert *et al*, 2001).

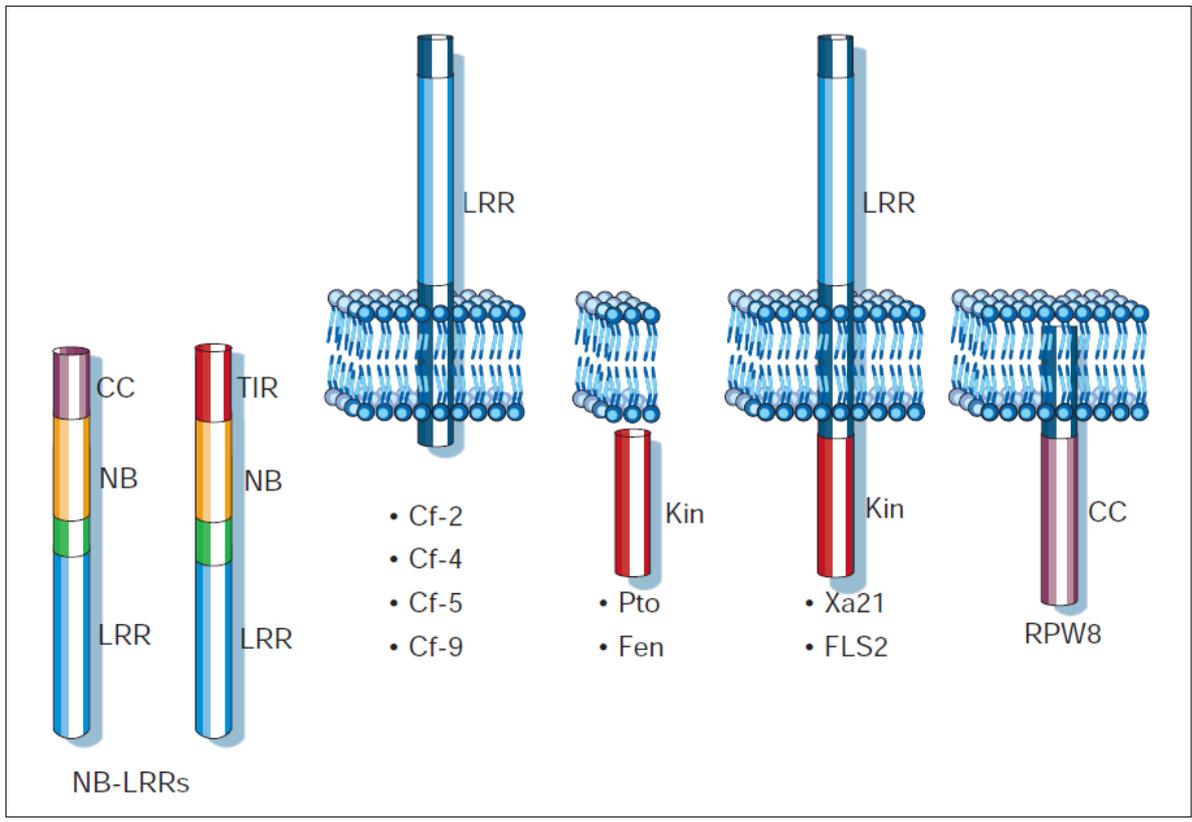
La plupart des résistances introduites dans les plantes cultivées appartiennent aux gènes R qui sont responsables de résistances hôtes, c'est ce qui nous intéressera ici.

### 5.1 Les différentes catégories de gènes R

Les gènes de résistance peuvent être classés en cinq catégories différentes, basées sur la structure de leurs domaines protéiques pour lesquels ils codent (figure 14 ; d'après Padmanabhan *et al*, (2009).

- La première catégorie code pour des protéines qui possèdent un domaine serine - thréonine kinase (STK). Ces protéines possèdent un domaine catalytique ainsi qu'un domaine de liaison à l'ATP. Le gène Pto, le premier gène de résistance cloné chez les plantes, appartient à cette catégorie (Martin *et al*, 1994). Ces protéines peuvent aussi intervenir dans d'autres fonctions biologiques. Il y a peu de gènes R qui appartiennent à cette catégorie.

- Les seconde et troisième catégories contiennent la famille des gènes possédant un domaine nucleotide binding site (NBS) et un domaine leucine rich repeat (LRR). Ces gènes représentent l'une des plus grandes familles de gènes chez les plantes (McHale *et al*, 2006). En



D'après Dangl & Jones, 2001

**Figure 14 :** Les différentes catégories de gènes de résistance définies chez les plantes.

plus de ces domaines, ces gènes portent un troisième domaine en position N-terminale : pour la seconde catégorie, c'est généralement un domaine Coiled-Coil (CC) et pour la troisième catégorie, celui-ci est un domaine Toll-interleukin receptor (TIR). Les deux catégories de gènes NBS-LRR sont plus amplement présentées dans le paragraphe I-5.2. Pour exemple, le gène L6 du lin qui confère la résistance au champignon *Melampsora lini*, code une protéine de type TIR-NBS-LRR (TNL) (Lawrence *et al*, 1995). Le gène 12c-1 de la tomate, qui confère la résistance au champignon *Fusarium oxysporum*, code une protéine de type CC-NBS-LRR (Simons *et al*, 1998).

Les 3 premières catégories de gènes de résistance codent pour des protéines localisées à l'intérieur de la cellule (figure 14).

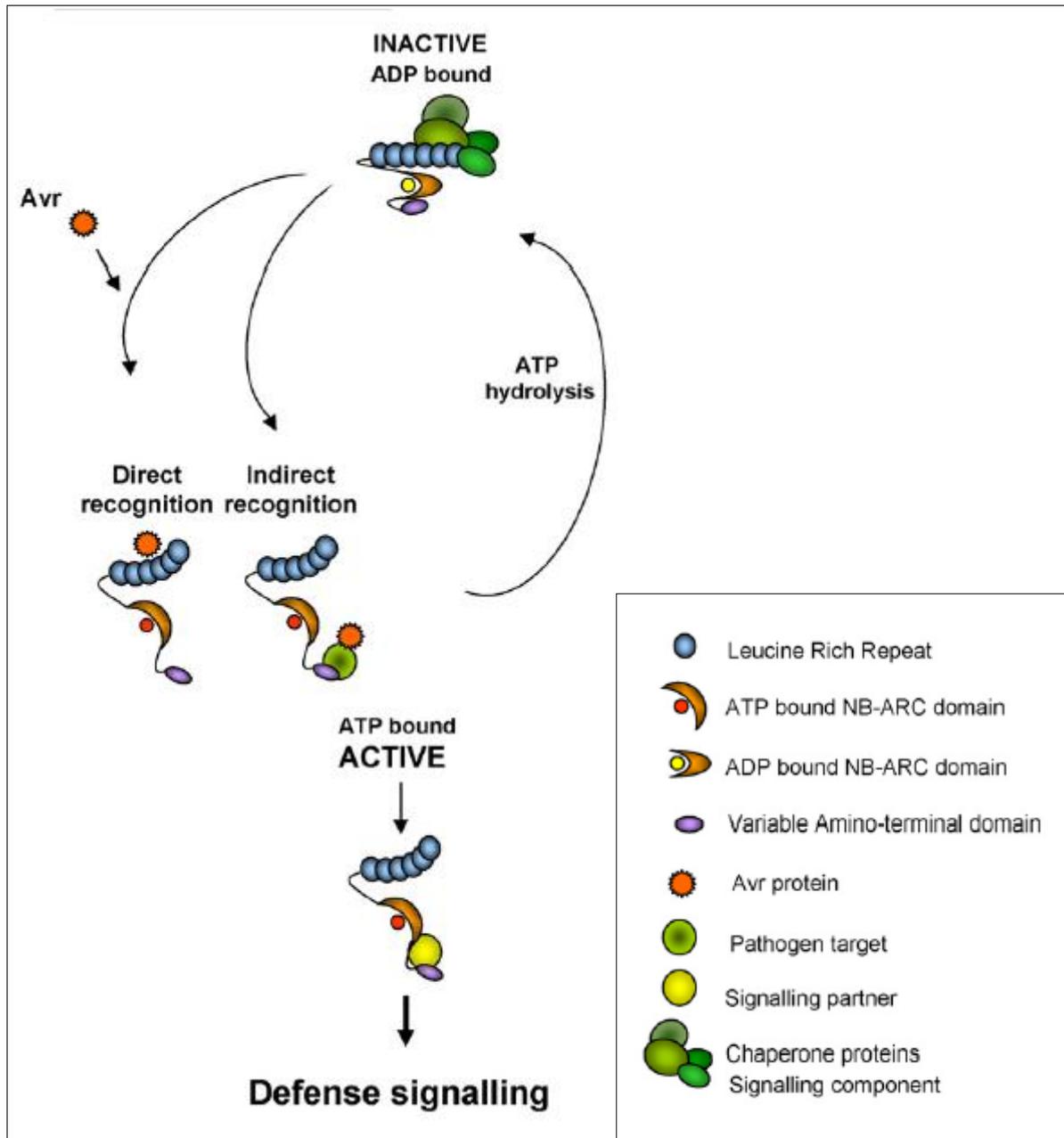
- Les quatrième et cinquième catégories codent pour des protéines possédant un domaine LRR extracellulaire et un domaine transmembranaire en C-terminal. La quatrième catégorie possède un domaine intracellulaire inconnu nommé *receptor-like protein* (RLP) tandis que la cinquième possède un domaine intracellulaire STK nommé *receptor-like kinase* (RLK) (figure 14). Les RLP sont représentés, par exemple, par la famille de gènes *Cf* conférant la résistance race-spécifique à *Cladosporium fulvum* chez la tomate (Hammond-Kosack & Jones, 1997 ; Thomas *et al*, 1997). Le gène *Xa21* est un représentant de la cinquième catégorie contrôlant la résistance du riz à *Xanthomonas oryzae* pv. *oryzae* (Song *et al*, 1995) et *Ve* de résistance à *Verticillium albo-atrum* de la tomate (Kawchuk *et al*, 2001 ; Dixon *et al*, 1996).

## 5.2 Les gènes de type NBS-LRR

Une grande majorité des gènes de résistance isolés jusqu'à présent, appartiennent à la grande famille des gènes NBS-LRR (Cf. tableau 2, page I-30). Il a été montré depuis peu que ces gènes pouvaient intervenir dans d'autres fonctions telles que le développement morphologique pour un gène TIR-NBS-LRR chez *A. thaliana* (Faigón-Soverna *et al*, 2006) ou la tolérance à la sécheresse pour un gène CC-NBS-LRR chez *A. thaliana* (Chini *et al*, 2004).

### 5.2.1 Les différents domaines

Les différents domaines constituant les gènes NBS-LRR se retrouvent dans d'autres règnes, mais ces gènes composés de l'enchaînement de ces différents domaines sont spécifiques du règne végétal.



**Figure 15** : Modèle de fonctionnement des protéines NBS-LRR.

En dehors de la période d'infection, le domaine NBS est liée a l'ADP. Lors de l'interaction directe ou indirecte avec un effecteur (Avr), ceci induit un changement de conformation qui libère l'ADP et qui est remplacé par l'ATP. Ceci forme l'état actif de la protéine NBS-LRR. Après induction du signal de défense, l'ATP est hydrolysé en ADP et la protéine retourne à son état inactif.

### 5.2.1.1 Domaine LRR

Le domaine LRR comporte une série de motifs protéiques contenant des leucines ou isoleucines et d'autres résidus hydrophiles. Le nombre de ces répétitions est extrêmement variable suivant les gènes.

Ce domaine se retrouve dans de nombreuses protéines à fonctions biologiques diverses chez les plantes, les animaux et les bactéries. Ce domaine est impliqué dans les interactions protéine-protéine. Chez les plantes, il a été montré que ce domaine était à l'origine de la spécificité des gènes de résistance (Padmanabhan *et al*, 2009). Ce domaine pourrait donc correspondre au domaine de reconnaissance des MAMPs ou des effecteurs de l'agent pathogène.

### 5.2.1.2 Domaine NBS

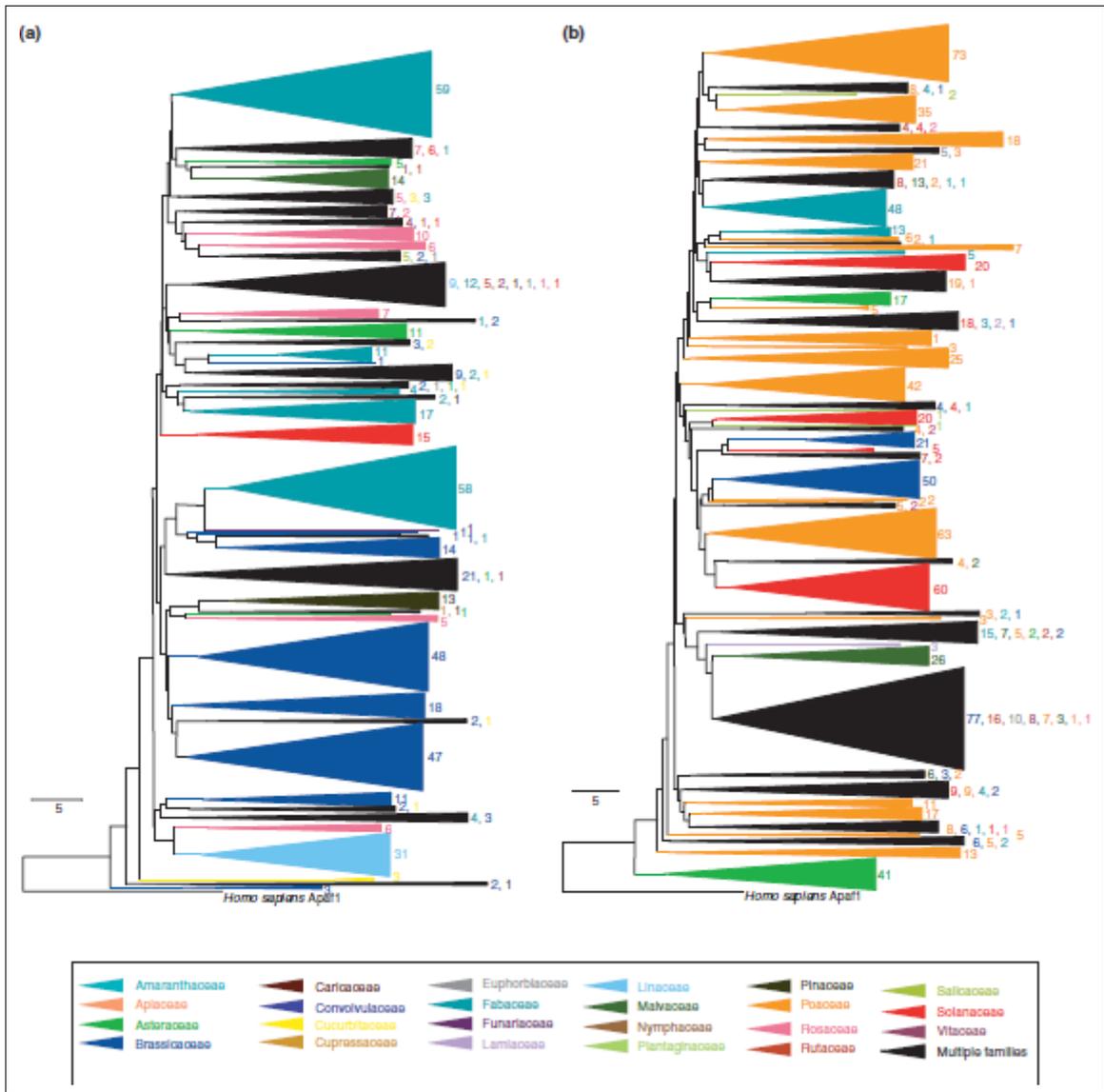
Le domaine NBS possède des similarités avec des inducteurs de la mort cellulaire chez de nombreux animaux (Van der Biezen & Jones, 1998). Sa fonction biochimique est la fixation et l'hydrolyse de l'ATP et/ou le GTP, et sa fonction biologique pourrait être l'activation de la résistance (figure 15 ; Rafiqi *et al*, 2009 ; Tameling *et al*, 2006 ; Tameling *et al*, 2002). Sans la présence de l'effecteur, la protéine NBS-LRR présente une forme inactive avec le domaine NBS qui est lié à l'ADP. Lors de la reconnaissance de l'effecteur par le domaine LRR, un changement de conformation entraîne la libération de l'ADP, le domaine fixe alors l'ATP, ce qui rend alors la protéine active. Après induction du signal de défense, l'ATP est hydrolysé en ADP et la protéine retourne à son état inactif.

Ce domaine est très conservé à travers le monde végétal et est composé de huit motifs différents : ils sont nommés P-loop (kinase 1), RNBS-A, Kinase 2, RNBS-B (Kinase 3), RNBS-C, GLPL, RNBS-D et MHDV. Parmi ces huit motifs, trois sont des kinases qui seraient responsables de la liaison à l'ATP ou ADP et de l'hydrolyse de l'ATP (Tameling *et al*, 2002)

### 5.2.1.3 Domaine TIR et CC

La famille des gènes NBS-LRR se divise en deux sous-classes d'après la nature du domaine N-terminal : TIR-NBS-LRR et CC-NBS-LRR (Cf. § I-5.1). Ces domaines jouent probablement un rôle dans l'interaction entre les gènes de résistance et les gènes situés en aval dans la voie de transduction du signal.

- Le domaine CC est connu pour permettre les interactions protéine-protéine.
- Le domaine TIR : similaire au domaine Toll du système immunitaire de la drosophile et similaire à l'interleukine-I du système immunitaire humain chez les animaux.



D'après McHale et al, 2006

**Figure 16** : Phylogénie de 1600 gènes NBS-LRR issues de plusieurs espèces végétales. Les numéros indiquent le nombre de séquences à l'intérieur de chaque clade. (a) : catégorie des TNLs, (b) : catégorie des CNLs.

#### 5.2.1.4 Domaine NLS

Un quatrième domaine a été identifié dans certains gènes NBS-LRR ; c'est le domaine NLS (nuclear localization signal) qui a été retrouvé dans des gènes TNL chez *A. thaliana*. Il s'agit d'un domaine d'adressage au compartiment nucléaire. Ce domaine a été identifié chez 5 gènes TNL de *P. trichocarpa* en position N-terminale (Kohler *et al*, 2008), et deux gènes TNL chez *P. deltoides* (Lescot *et al*, 2004). Une transformation chez le tabac d'un gène TNL de *P. tremula*, avec et sans ce domaine, fusionné au gène rapporteur GFP (green fluorescent protein) a permis de démontrer la fonctionnalité de ce domaine (Cf. figure 60, page V-116 ; Kohler *et al*, 2008).

### 5.2.2 Les différentes catégories

On retrouve les gènes TIR-NBS-LRR (TNL) et CC-NBS-LRR (CNL) chez des espèces angiospermes et gymnospermes (figure 16 ; McHale *et al*, 2006). Cela implique que l'origine de ces gènes est très ancienne, ils sont apparus avant la séparation entre angiospermes et gymnospermes. Aucun gène de la catégorie des TNL n'a été identifié chez les monocotylédones.

Les deux catégories de gènes TIR-NBS-LRR (TNL) et CC-NBS-LRR (CNL) sont largement représentées chez le peuplier. Il existe une troisième catégorie de gènes NBS-LRR qui diffère par un autre type de domaine en 5' ; ce domaine est le domaine BED-finger (BED), il code pour un motif zinc-finger, dont la fonction serait la liaison à l'ADN. La découverte du domaine BED a été publiée en 2000 et a été nommé BED-finger pour deux protéines de drosophile connues sous les noms de BEAF et DREF contenant ce domaine (Aravind, 2000). Il a été identifié 34 BED-NBS-LRR (BNL) chez le peuplier (Kohler *et al*, 2008). Cette catégorie de gènes pourrait correspondre à des régulateurs transcriptionnels induits en réponse à une attaque d'un agent pathogène. Ce type de gène de résistance n'a pas encore été identifié chez d'autres dicotylédones mais est présent chez le riz, une monocotylédone. Il semble donc que le riz et le peuplier ont acquis de manière indépendante cette architecture de gène. Un gène BNL nommé *Xa1*, a été identifié comme résistant à la bactérie responsable de blight du riz .

Les gènes de résistance de type NBS-LRR sont regroupés en clade selon leur similarité de séquence. Certains clades de ces gènes, principalement pour les CNL, sont conservés chez différentes espèces, ce qui implique une divergence ancestrale (figure 16 ; McHale *et al*, 2006). A l'inverse, il y a des clades spécifiques à certaines espèces qui montrent une expansion récente. Les gènes BNL sont proches phylogénétiquement de la catégorie des gènes CNL (Kohler *et al*, 2008).

**Tableau 3 :** Nombre de gènes NBS-LRR chez *P. trichocarpa*, *A. thaliana*, et *Oryza sativa*.

<sup>a</sup> Kolher *et al*, 2008

<sup>b</sup> Meyer *et al*, 2003

<sup>c</sup> Zhou *et al*, 2004

Predicted protein domains	Populus <sup>a</sup>	Arabidopsis <sup>b</sup>	Rice <sup>c</sup>
Non-TIR-NBS-LRR	224	55	480
CC-NBS-LRR (CNL)	119	51	159
BED-NBS-LRR (BNL)	34	–	3
NBS <sub>CC/BED</sub> -LRR (NL)	71	4	40
Other Non-TIR-NBS-LRR	–	–	278
TIR-NBS-LRR	91	94	–
TIR-NBS-LRR (TNL)	64	83	–
TIR-NBS-LRR-TIR (TNLT)	13	–	–
TIR-NBS-LRR-NBS (TNLN)	1	–	–
NBS-LRR-TIR (NLT)	1	–	–
NBS <sub>TIR</sub> -LRR (NL)	12	2	–
Other TIR-NBS-LRR	–	9	–
MIXED NBS-LRR	2	–	–
TIR-CC-NBS-LRR (TCNL)	2	–	–
Total NBS-LRR	317	149	480
Non-TIR-NBS	53	6	52
CC-NBS (CN)	19	4	7
BED-NBS (BN)	6	–	–
NBS-BED (NB)	1	–	–
NBS <sub>CC</sub>	27	1	45
Other Non-TIR-NBS	–	1	–
TIR-NBS	32	23	3
TIR-NBS	13	21	3
NBS-TIR	19	–	–
Other TIR-NBS	–	2	–
Total NBS	85	29	55
Total	402	178	535

Les gènes de la catégorie des TNLs ont plus d'introns que les gènes de la catégorie des CNLs, qui peuvent même ne pas posséder d'introns (Kohler *et al*, 2008).

Il a été montré aussi la présence de gènes apparentés aux gènes NBS-LRR ; ces gènes ne portent pas le domaine LRR. Il s'agit donc de gènes TIR-NBS (TN), CC-NBS (CN) et BED-NBS (BN). Il y a aussi des gènes qui ne portent que le domaine NBS (N).

## 5.3 Organisation des gènes NBS-LRR

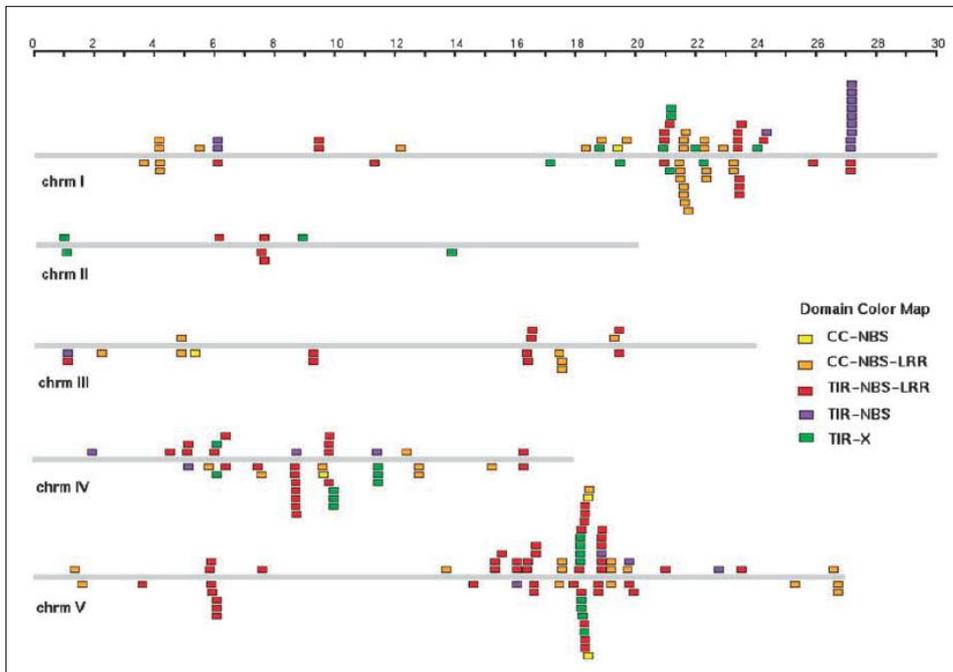
Le séquençage de génomes complets apporte beaucoup d'informations sur le nombre, la diversité et l'organisation des gènes NBS-LRR au sein des génomes. *A. thaliana* est la première plante chez qui un inventaire exhaustif et complet des gènes NBS-LRR a été établi ; elle forme donc une référence pour les autres plantes séquencées où cet inventaire a pu être fait.

Bien qu'il y ait de nombreuses informations sur les gènes NBS-LRR chez beaucoup d'espèces de monocotylédones et de dicotylédones, peu d'informations sont disponibles chez les espèces pérennes. Le séquençage du peuplier et de la vigne donne les premières informations qui demandent à être complétées par d'autres.

### 5.3.1 Nombre de gènes NBS-LRR

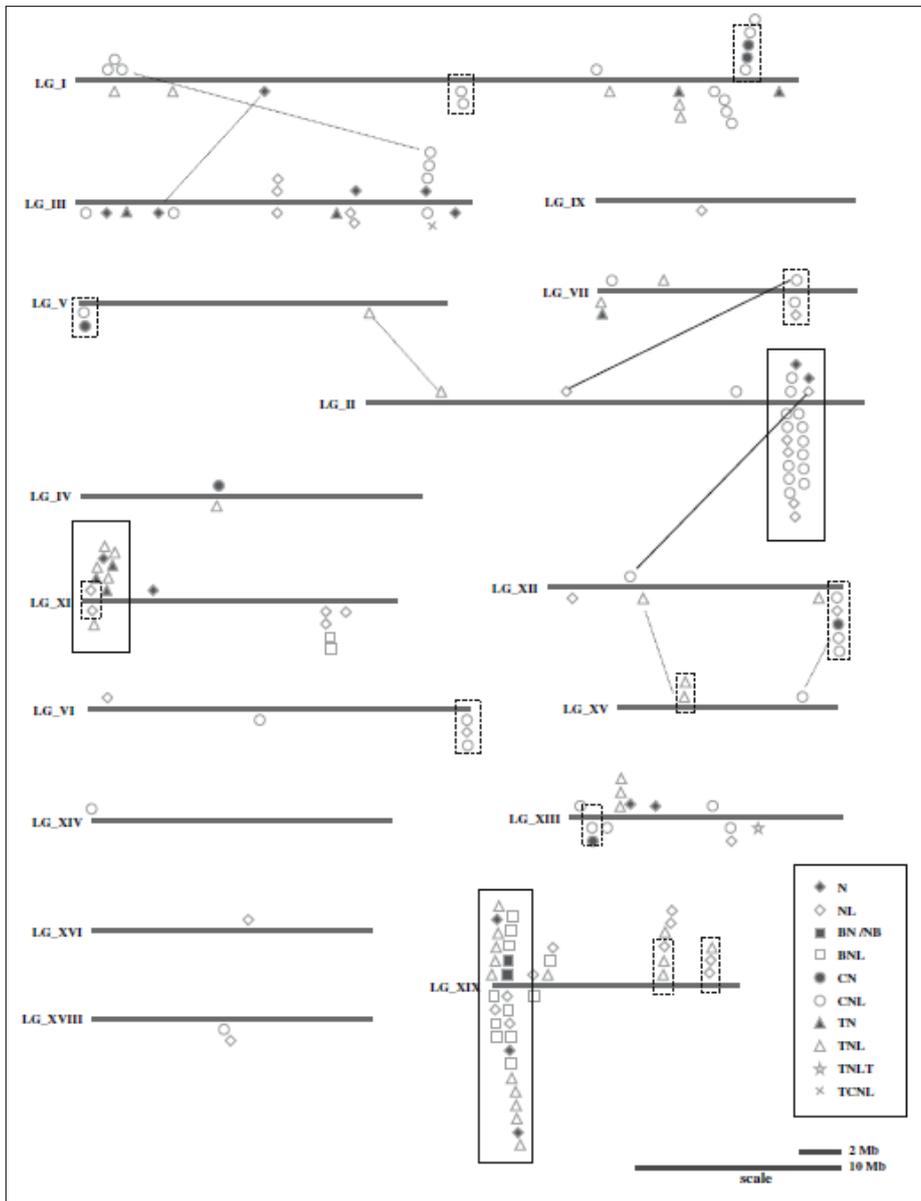
Le tableau 3 résume le nombre de gènes NBS-LRR trouvés chez quelques génomes séquencés (Kohler *et al*, 2008 ; Zhou *et al*, 2004 ; Meyers *et al*, 2003). Chez *A. thaliana*, 149 gènes NBS-LRR ont été identifiés, dont la majorité appartient à la classe des TNLs (Meyers *et al*, 2003 ; Meyers *et al*, 1999). A partir de la séquence du génome de *P. trichocarpa* v1.1, 402 gènes de la famille des NBS-LRRs ont pu être identifiés par Kohler *et al*, (2008) et 330 dans une autre étude (Yang *et al*, 2008), le double d'*A. thaliana*. Le peuplier possède un nombre inférieur de gènes NBS-LRR à celui du riz, qui en compte 464 (McHale *et al*, 2006 ; Wisser *et al*, 2005), et de la vigne, qui en compte 459 (Yang *et al*, 2008).

Chez le peuplier, de nombreux gènes porteurs du domaine NBS sont dépourvus de domaines LRR : 13 TIR-NBS, 19 CC-NBS, 7 BED-NBS et 49 gènes porteurs uniquement du domaine NBS (Kohler *et al*, 2008). La phylogénie de ces gènes avec le domaine NBS montre qu'ils sont disséminés dans les différentes catégories, indiquant une relation étroite avec les autres gènes TNL, CNL et BNL (Kohler *et al*, 2008). Il est possible qu'au moins certains de ces gènes aient été mal assemblés et que d'autres domaines uniques n'aient pas été identifiés à ce jour.



D'après Meyer *et al.*, 2003

**Figure 17** : Localisation des gènes NBS-LRR dans le génome d'*Arabidopsis*.



D'après Kolher *et al.*, 2008

**Figure 18** : Localisation des gènes NBS-LRR dans le génome de *P. trichocarpa* V1.1.

### 5.3.2 Distribution des gènes NBS-LRR

Les gènes NBS-LRR sont rarement trouvés isolés dans le génome des plantes, ils sont généralement regroupés dans des régions génomiques précises en clusters plus ou moins complexes.

L'exemple le plus connu de singleton est celui du gène L chez le lin, qui contrôle la résistance à la rouille du lin *Melampsora lini*, 13 formes alléliques ont été décrites.

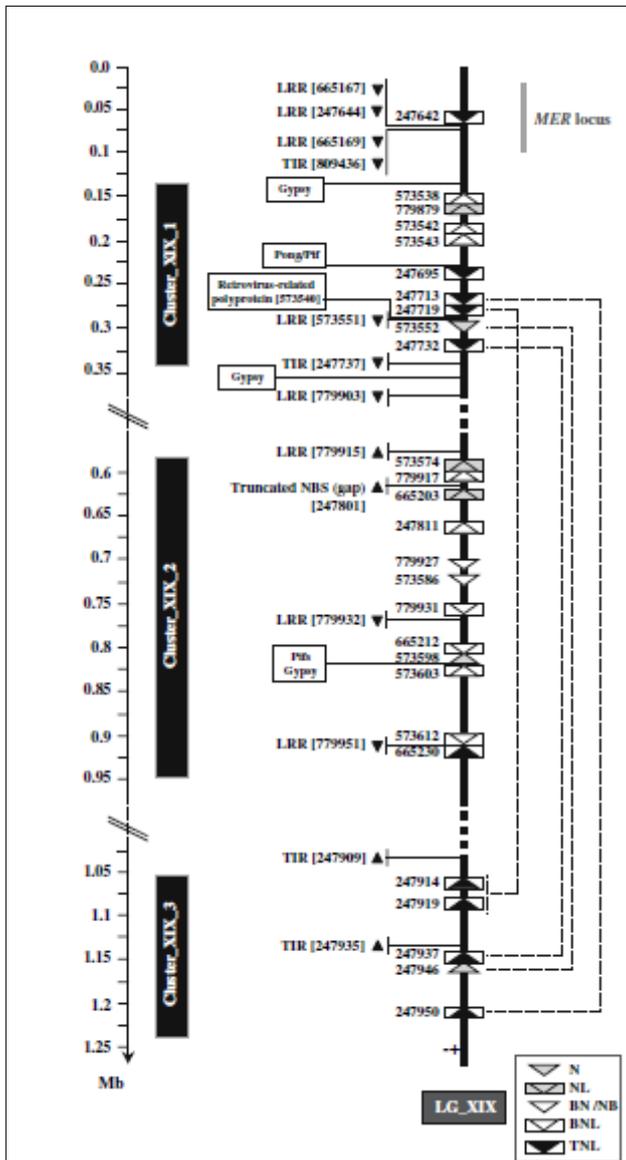
Chez *A. thaliana*, 200 gènes NBS-LRR sont distribués en 40 singletons et 43 clusters (figure 17 ; Meyers *et al*, 2003). Chez le peuplier, les 170 gènes NBS-LRR assignés aux chromosomes sont répartis en 37 clusters et seulement 28% des gènes sont en singletons (figure 18). Une répartition comparable a été observée chez le riz, la vigne, mais c'est aussi le cas chez d'autres espèces, où les clusters de gènes NBS-LRR sont alors identifiés par cartographies génétique et physique (Geffroy *et al*, 2009 ; Ameline-Torregrosa *et al*, 2008).

La distribution des clusters n'est pas homogène sur les chromosomes : chez *A. thaliana*, les chromosomes 1, 4 et 5 portent des clusters de gènes NBS-LRR, tandis que les chromosomes 2 et 3 sont relativement dépourvus de ce type de gènes (figure 17 ; Meyers *et al*, 2003). Chez le peuplier, les gènes de résistance sont concentrés dans trois clusters, sur les chromosomes 2, 11 et 19 (Kohler *et al*, 2008). Le plus gros cluster est celui de l'extrémité du chromosome 19 qui comporte 31 gènes qui se répartissent sur près de 1,16 Mb (figure 19 ; Kohler *et al*, 2008).

Les clusters sont plus ou moins complexes, comprenant un nombre variable de gènes et d'allèles. Deux des plus gros clusters étudiés sont le locus Dm3 chez la laitue, et le locus B4 du haricot. La proximité de gènes entre eux dans un même cluster fait que ce sont des régions complexes et redondantes, qui sont difficiles à étudier.

En général, la plupart des gènes localisés dans un même cluster sont proches phylogénétiquement et appartiennent donc à un même clade. Mais il peut y avoir aussi des clusters regroupant plusieurs catégories et/ou plusieurs clades de gène NBS-LRR de plusieurs catégories comme chez *A. thaliana*. Ce sont donc des clusters de clusters appelés "super-clusters" comme l'extrémité du chromosome 19 du peuplier où sont présents les catégories TNL et BNL (Kohler *et al*, 2008).

Les clusters de gènes de résistance auraient une fonction de réservoirs de variations alléliques pour résister à un nombre important d'agents pathogènes.



D'après Kolher *et al*, 2008

**Figure 19 :** Le super cluster de gènes NBS-LRR de l'extrémité du chromosome 19 du peuplier.

## 5.4 Évolution des gènes NBS-LRR

Dans le cadre de l'évolution entre les plantes et les agents pathogènes (Cf. § I-4.3.3.2), les plantes doivent générer constamment de nouveaux allèles afin de créer de nouvelles spécificités envers les nouveaux gènes d'avirulence des agents pathogènes. Les mécanismes "birth and death process" développés par Michelmore & Meyers, (1998) seraient à l'origine de la grande variabilité observée. Plusieurs étapes de recombinaison intra et inter-locus, mutations et conversions, seraient à l'origine de la grande variabilité des gènes R et en particulier du domaine LRR (Wicker *et al*, 2007 ; Meyers *et al*, 2003). Deux types d'évolution des gènes R ont été proposés : le premier est basé sur des échanges fréquents entre des paralogues via des crossing-over intra-locus ; et le deuxième est basé sur une évolution par mutation ponctuelle (Kuang *et al*, 2004).

### 5.4.1 Les mécanismes évolutifs

#### 5.4.1.1 Expansion du nombre de gènes

Le nombre important de gènes NBS-LRR s'explique premièrement par des mécanismes qui ne leur sont pas spécifiques :

- Les événements de polyploïdisation, c'est-à-dire de duplication complète d'un génome, ont été observés et ont sûrement joué un rôle dans l'expansion de ces gènes.

- La duplication segmentale conduit à la dispersion dans le génome des gènes appartenant à un même clade.

L'existence de cluster est expliquée par des mécanismes de duplication en tandem :

- La recombinaison inégale entre paralogues (recombinaison non allélique homologue) génère de nouvelles copies en tandem.

- La recombinaison non homologue ou illégitime génère aussi de nouvelles copies en tandem.

- Les éléments transposables (ETs) de classe II ont la capacité de transporter des gènes ou des fragments de gènes lors de leur transposition.

Il est difficile de dire lequel des mécanismes est prépondérant sur l'autre, ces modes d'évolution peuvent tous intervenir, à des degrés divers.

#### 5.4.1.2 Création d'allèles

- Les gènes R évoluent aussi de manière classique avec des mutations ponctuelles et des recombinaisons alléliques qui génèrent de nouveaux allèles (Kuang *et al*, 2004).



- Des recombinaisons hétérologues au sein d'un même gène : avec les domaines LRR répétés, il peut arriver qu'une recombinaison se fasse entre deux allèles d'un même gène au niveau de deux répétitions LRR mais sans homologie. Cela joue sûrement un rôle dans la diversité observée au sein de ce domaine, conférant à la plante un vaste réservoir allélique et donc un réservoir de spécificité.

#### **5.4.1.3 Conservation des nouvelles copies de gènes.**

- Néofonctionalisation : une modification dans l'une des deux copies donne une nouvelle fonction. Pour les gènes R, la modification peut donner une nouvelle spécificité pour un nouveau gène Avr.

- Pseudogénisation : il s'agit d'élimination d'une des deux copies du gène, avec des codons stop. Mais ces pseudo-gènes sont des réservoirs pour la création de nouveaux allèles avec des recombinaisons inégales ou non homologues.

#### **5.4.2 Diversité à l'intérieur des gènes**

Les domaines TIR et NBS sont très conservés, contrairement au domaine LRR qui est très variable en séquence et en taille. Il apparaît donc que la partie N-terminale de ces gènes est plus conservée que leur partie C-terminale. Cette variabilité du domaine LRR semble logique, elle permet aux organismes de posséder une spécificité de résistance à une large gamme d'agents pathogènes.

## **6 Génétique de la résistance aux rouilles foliaires du peuplier**

### **6.1 Résistance qualitative et résistance quantitative**

Deux types de résistance sont distingués :

- Résistance complète (qualitative) : l'agent pathogène n'infecte pas la plante, la plante est résistante. Cette résistance est généralement contrôlée par un seul gène.

- Résistance partielle (quantitative) : l'agent pathogène infecte la plante mais la plante y est plus ou moins sensible. Cette résistance partielle est généralement contrôlée par de nombreux gènes (résistance polygénique). Bien souvent, il est fait un abus de langage : on parle de résistant pour un individu peu sensible et de non résistant pour un individu très sensible.



Les premières analyses génétiques de la résistance à *Mlp* menées en France ont montré l'existence de résistance qualitative et quantitative en fonction des souches de *Mlp* dans des hybrides *P. deltoides* x *P. trichocarpa* dont les résistances qualitatives proviennent de l'espèce *P. deltoides* (Pichot & Tessier du Cros, 1993a et b). Les espèces *P. trichocarpa* et *P. nigra* présentent une résistance quantitative à *Mlp*, mais aucune résistance qualitative n'y a été détectée.

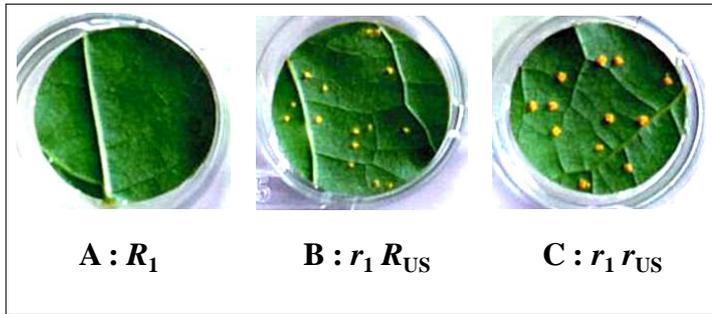
Dans une étude d'AFLP-RAPD dans un plan factoriel 2 X 2, Villar *et al*, (1996) ont développé un marqueur RADP (M03/04\_480) qui co-ségrège avec les résistances qualitatives, hérité d'un parent de l'espèce *P. deltoides* dans des hybrides interspécifiques. Ce même marqueur a été utilisé par Lefèvre *et al*, (1998) dans un plan multifactoriel 9 X 9. Ils ont pu montrer que ce marqueur était lié aux résistances qualitative et quantitative, qui sont héritées d'un parent de l'espèce *P. deltoides*. Les auteurs ont proposé une hypothèse : un cluster de gène gouvernerait la résistance qualitative et quantitative à différentes souches de *Mlp* chez l'espèce *P. deltoides*.

## 6.2 Les loci de résistance qualitative à *Mlp*

### 6.2.1 Locus *MER*

#### 6.2.1.1 Identification

L'étude de la résistance chez une famille issue du croisement entre la femelle *P. deltoides* S9-2 et le mâle *P. nigra* Ghoy (famille 87001) a permis de mettre en évidence un locus de résistance à *Mlp* (Cervera *et al*, 1996), il a été appelé *MER* pour « Melampsora Résistance ». L'inoculation en laboratoire des 262 individus de cette famille montre une ségrégation 1:1 d'un locus unique de résistance qualitative avec 3 souches différentes. Plus tard, Pinon & Frey, (2005) ont montré que la virulence 7 contourne le locus *MER*. Pour identifier des marqueurs proches du locus, les auteurs ont combiné deux techniques : l'AFLP (Amplification Fragment Length Polymorphism) et le BSA (Bulked Segregant Analysis). Le principe de cette dernière est de mélanger des ADN d'individus (bulk) de phénotype similaire pour un caractère donné et de comparer les marqueurs AFLP entre deux bulks dont les phénotypes sont opposés (Michelmore *et al*, 1991). Il permet de trouver des marqueurs AFLP qui sont associés à l'un des deux phénotypes et donc génétiquement liés au caractère étudié. Cent quarante quatre couples d'amorces ont été testés sur des bulks d'individus sensibles et résistants et 123 descendants de la



**Figure 20** : Phénotypes d'individus descendants de la famille 54 lors de l'infection par *Mlp*.

A : phénotype absence de sores ( $R_1$ ).

B : phénotype petit sore ( $r_1 R_{US}$ ).

C : phénotype gros sore ( $r_1 r_{US}$ ).

famille 87001, ce qui a permis de trouver trois marqueurs AFLP liés génétiquement à la résistance.

### **6.2.1.2 Localisation**

En produisant une carte génétique complète de l'individu S9-2, Cervera *et al.*, (2001) ont localisé le locus *MER* à l'extrémité du groupe de liaison XIX. Par la suite, une étude menée sur 512 individus a permis de localiser finement le locus *MER* (Zhang *et al.*, 2001). Les individus sont issus de trois familles différentes dont la mère est l'individu S9-2 ou un descendant de cet individu porteur du locus *MER*. Un total de 556 couples d'amorces a été testé en BSA-AFLP. Ce locus est entouré par deux marqueurs génétiques séparés par 1 cM. Le séquençage des produits d'amplification AFLP a montré la présence de gènes NBS-LRR à l'extrémité du chromosome 19, à proximité du locus *MER* (Zhang *et al.*, 2001).

### **6.2.1.3 Séquence d'un clone BAC proche du locus *MER***

En 2004, Lescot *et al.*, (2004) dans une tentative de clonage positionnel du locus *MER*, a séquencé un clone BAC de 95 Kb distant de 0,4 cM du locus. Cette séquence est issue de *P. deltoides*. Ce clone BAC a été entièrement annoté ; il contient 3 gènes de type TIR-NBS-LRR et de nombreux éléments transposables. Le marqueur R7, développé sur la séquence de ce clone BAC, a été localisé à 33,6 cM d'une extrémité du groupe de liaison XIX dans un autre pedigree (Yin *et al.*, 2004b).

## **6.2.2 Locus *R*<sub>1</sub>**

Une famille issue du plan factoriel, la famille 54 (Villar *et al.*, 1996 ; Cf. § I-6.1), a été particulièrement étudiée et est devenue une famille modèle pour l'étude de la résistance chez le peuplier à *Mlp*. Cette famille, résultat du croisement entre *P. deltoides* (individu 73028-62) et *P. trichocarpa* (individu 101-74), a permis de mettre en évidence l'existence du locus *R*<sub>1</sub> (figure 20 ; Dowkiw *et al.*, 2003). Ce locus, hérité de la mère *P. deltoides*, contrôle une résistance qualitative. Les 348 individus de la famille 54 ont été phénotypés avec 7 souches de *Mlp* avec plusieurs pathotypes, ce qui a permis de caractériser son contournement par la virulence 1 ; c'est pourquoi ce locus est appelé *R*<sub>1</sub>. La ségrégation 1:1 de cette résistance qualitative, avec quatre pathotypes ne possédant pas la virulence 1, renforce l'hypothèse que le locus est gouverné par un seul gène ou plusieurs très proches physiquement. Avec 348 individus de la famille 54, la première carte génétique de l'individu 73028-62 a montré que le locus *R*<sub>1</sub> était sur le groupe de



liaison (LG) XIX (Jorge *et al*, 2005). Cette carte génétique a été réalisée avec des marqueurs microsatellites et des marqueurs AFLP.

### 6.2.2.1 Cluster de gènes à proximité de $R_1$

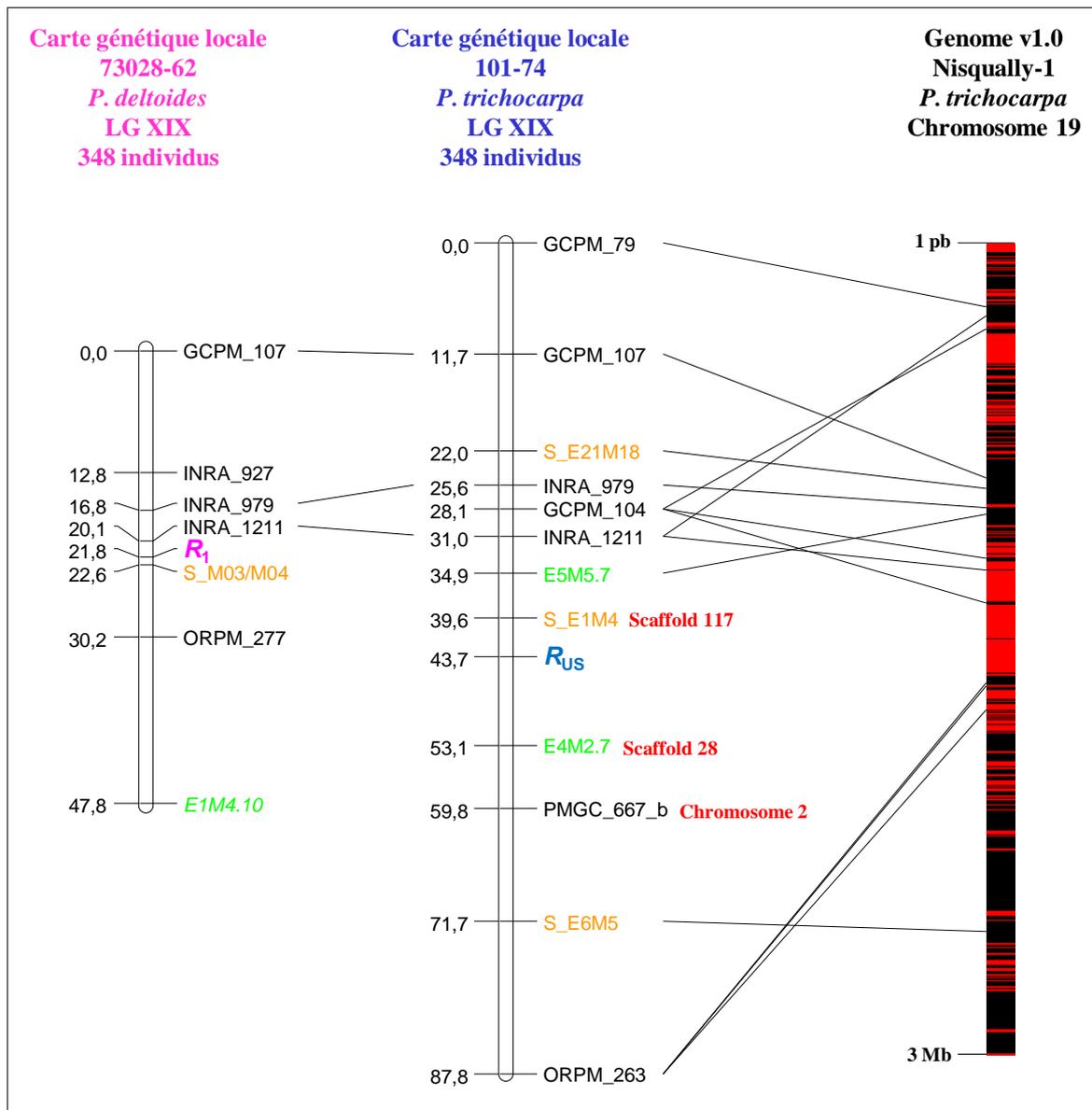
Le 73028-62 porte une résistance qualitative à la souche de *Mlp* 93CV1 mais la descendance de la famille 54 n'est pas résistante à cette souche. Lefèvre *et al*, (1998) a émis l'hypothèse que la résistance à cette souche est gouvernée par deux gènes proches en répulsion. Dans l'étude de 350 individus de la famille 54, Dowkiw, (2003) a trouvé 5 individus qui sont résistants à la souche 93CV1. Il a donc émis l'hypothèse d'une recombinaison entre les deux gènes en répulsion, ce qui permet de conserver la résistance hérité du 73028-62. Comme ces 5 individus sont de phénotype  $R_1$  pour les souches qui ne portent pas la virulence 1, ces deux gènes en répulsion seraient liés au locus  $R_1$ , dont l'un des deux pourrait être  $R_1$ . Il y aurait donc un cluster de gènes de résistance à proximité de  $R_1$ . La souche 93CV1 est différente aux autres souches qui portent aussi la virulence 1, par la présence de la virulence 7, par conséquent, ces deux gènes en répulsion contrôlèrent une résistance contournée par la virulence 7. Ils ont donc été nommés  $R_{7a}$  et  $R_{7b}$ .

### 6.2.3 Les résistances qualitatives identifiées pour d'autres espèces de *Melampsora*

Le locus *Lrd1* d'une résistance qualitative à une souche de *M. medusae* a été identifié dans une famille  $F_1$  intra-espèce *P. deltoides* (Tabor *et al*, 2000). Les marqueurs RAPD utilisés dans cette étude n'ont pas permis de localiser ce gène sur le génome et donc il n'est pas possible d'étudier les relations entre *Lrd1* et les autres gènes de résistance identifiés.

Le locus *MXC3* (*Melampsora x columbiana* pathotype 3) d'une résistance qualitative est hérité de *P. trichocarpa* a été identifié dans une population  $F_1$  inter-espèce *P. deltoides* x *P. trichocarpa*. Ce locus gouverne une résistance à *M. x columbiana*. Il a été localisé sur le groupe de liaison IV (Yin *et al*, 2004b).

Le locus *Mmd1* contrôle une résistance qualitative dans une population  $F_2$  *P. trichocarpa* x *P. deltoides* infectés par *M. medusae* f.sp. *deltoideae* (Newcombe *et al*, 1996 ; Newcombe, 1998). Ce locus a également été impliqué dans la résistance quantitative à l'agent pathogène dans des conditions contrôlées et sur le terrain. Il a été cartographié sur un LG contenant aucun des marqueurs que nous avons utilisés.



**Figure 21** : Localisation des loci  $R_1$  et  $R_{US}$  sur le chromosome 19.

Dans la barre représentant le génome, le rouge indique les trous de séquence et le noir indique la présence de séquence. Les marqueurs microsatellites sont indiqués en noir, les marqueurs AFLP sont indiqués en vert et les marqueurs SCAR ont indiqués en orange.

### 6.3 Un locus de résistance quantitative à *Mlp* : $R_{US}$

La technique d'inoculation de disques foliaires excisés permet la décomposition de la résistance quantitative du peuplier à *Mlp* pour ses principaux éléments épidémiologiques : la période de latence (LP), le nombre urédies (UN), la taille des urédies (US) pour un isolat donné de *Mlp* dans des conditions environnementales contrôlées.

Dans l'étude de la même famille, l'analyse génétique des individus en ségrégation montre une distribution bimodale de la valeur phénotypique de la taille des sores (figure 24, page II-67), paramètre épidémiologique de la résistance quantitative à *Mlp*, ce qui indique un locus à déterminisme simple avec probablement un seul gène, ou plusieurs gènes liés génétiquement (Dowkiw *et al*, 2003).

Ce locus de résistance, ségrége aussi dans les autres familles du plan factoriel dont le père est le 101-74 ce qui montre qu'il est hérité de cet individu *P. trichocarpa*. L'allèle  $R_{US}$ , dominant, est responsable de la petite taille des sores, comparé à l'allèle  $r_{US}$  (figures 20, page I-39). Ce locus est appelé  $R_{US}$  pour "résistance uredinia size".

Ce caractère quantitatif a été rendu qualitatif afin de le cartographier génétiquement (figure 24, page II-67). La première cartographie génétique de l'individu 101-74, établi avec des marqueurs AFLP et microsatellite, a permis de localiser  $R_{US}$  sur un petit LG contenant trois marqueurs génétiques, non assigné aux chromosomes (Jorge *et al*, 2005).

Afin d'identifier des marqueurs proches du locus  $R_{US}$ , une seconde démarche AFLP a été réalisée avec des bulks d'individus (BSA) (Masle, 2005). Afin d'augmenter la puissance de cette technique et de limiter les faux positifs, quatre couples de bulks (mélanges de 10 individus "gros sores" et mélange de 10 individus "petits sores") provenant de trois familles différentes issues du père 101-74 ont été réalisés : 2 couples de bulks de la famille 54, 1 couple de bulks de la famille 64, 1 couple de bulks de la famille 84. Soixante deux couples d'amorces AFLP ont été testés, 55 de ces couples ont donné des résultats lisibles. Huit amplification des 4 422 obtenues se sont révélées polymorphes entre les bulks de phénotypes opposés, présentes chez le 101-74 et absentes chez le 73028-62. Ces marqueurs AFLP ont été confirmés sur 90 individus de la famille 54. Ces huit bandes AFLP ont été clonées puis séquencées. L'alignement des séquences AFLP sur la séquence du génome a permis de cartographier  $R_{US}$  sur l'extrémité du LG XIX (figure 21 ; Masle, 2005). Il y a aussi deux séquences qui s'alignent sur le scaffold 117. Ce scaffold porte aussi des gènes NBS-LRR. Ces séquences ont été utilisées pour développer des marqueurs SCAR (sequence characterized amplified region) pour cartographier les 348 individus de la



famille 54. Ces couples d'amorces AFLP ont été redessinés avec l'une des deux amorces sur la séquence du génome V1, ce qui a permis d'en cartographie génétiquement 3 d'entre eux.

### 6.3.1 Analyse de QTL

Le locus  $R_{US}$  est particulièrement intéressant car il co-localise avec des QTL (quantitative trait loci) pour les autres paramètres épidémiologiques : la période de latence et le nombre de sores. De plus, la mesure de la résistance en champs, dans 3 dispositifs différents, montre une co-localisation avec le locus  $R_{US}$  (Jorge *et al*, 2005).

Une analyse de QTL a été effectuée sur les descendants de la famille 54 qui a montré que des QTLs des trois paramètres de la résistance quantitative étaient liés aux deux loci  $R_1$  et  $R_{US}$ .

## 7 Objectifs

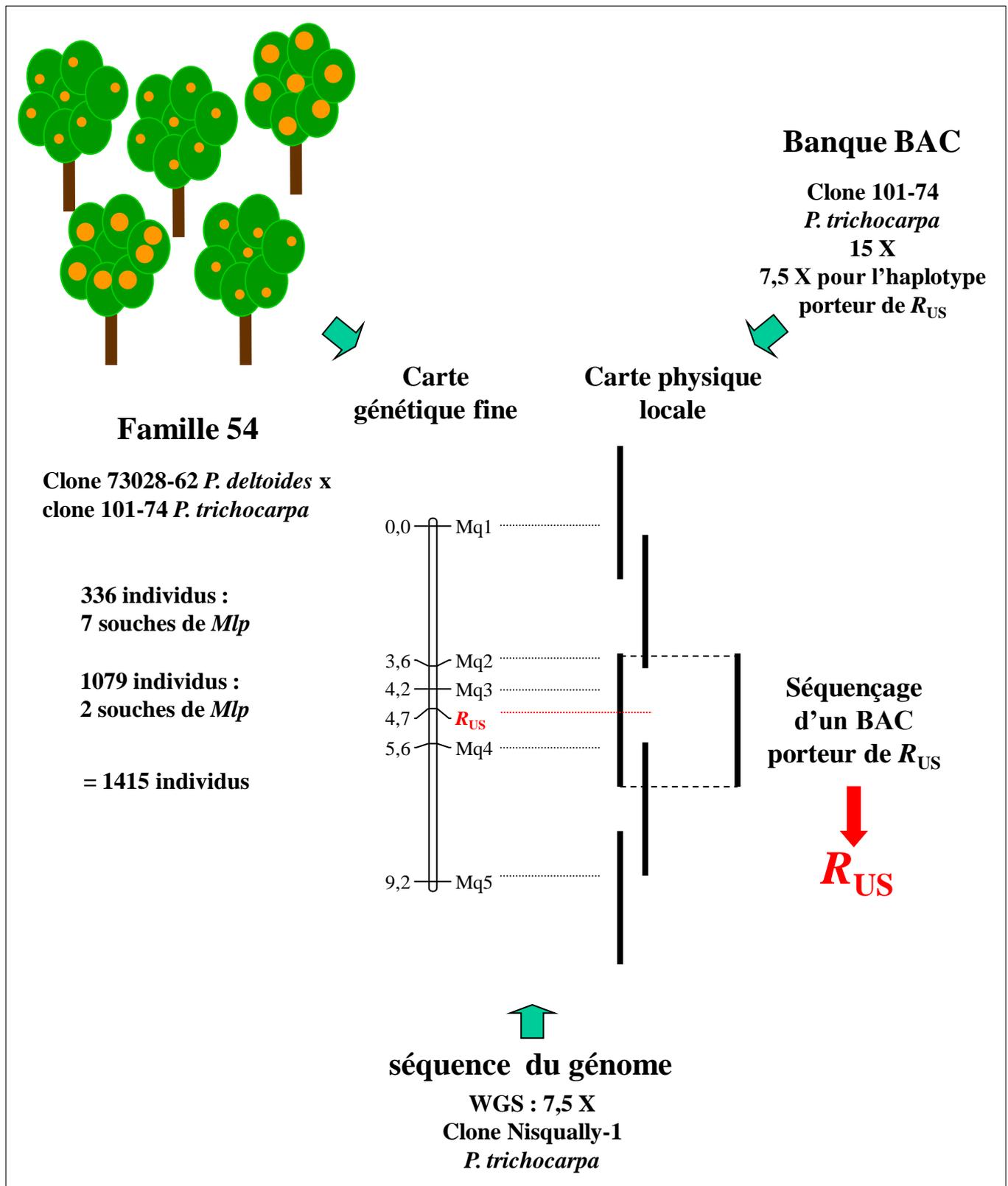
### 7.1 Les objectifs de recherche de l'INRA

La recherche d'une résistance durable à la rouille foliaire pour les peupliers cultivés est devenue la problématique majeure de la filière populicole (Jorge *et al*, 2005). En 2001, le GIS Peuplier (Groupement d'Intérêt Scientifique) a été créé par trois organismes (AFOCEL - Cemagref - INRA), afin de mettre en commun leurs moyens et leurs compétences en matière de sélection du peuplier. L'objectif du GIS Peuplier est de définir la stratégie la plus appropriée pour l'amélioration et la sélection du peuplier, et donc de proposer des cultivars adaptés aux conditions françaises notamment par rapport aux agents pathogènes, mais aussi d'évaluer les nouveaux cultivars étrangers.

Un projet coopératif entre quatre équipes de l'INRA (Orléans, Nancy, Evry et Bordeaux) a été mis en place afin de répondre au problème de résistance durable à la rouille du peuplier. Les objectifs de ce projet sont avant tout d'identifier les solutions permettant de limiter les pressions de sélection exercées sur des populations de *Mlp* grâce à une meilleure connaissance des bases génétiques et moléculaires de l'interaction peuplier-*Mlp*. Cette étude déterminera les niveaux de variabilité génétique présents dans les populations de la plante hôte et de l'agent pathogène, ainsi que les facteurs déterminant la dynamique d'évolution des populations de *Mlp*.

Les énormes progrès en biologie moléculaire ont permis de développer des outils pouvant explorer le génome des deux espèces et de mieux en comprendre leurs complexités.

C'est dans cette thématique de recherche développée à l'INRA que s'inscrit cette thèse.



**Figure 22** : Stratégie pour le clonage positionnel du locus  $R_{US}$ .

## 7.2 Les objectifs de cette thèse

### 7.2.1 Organisation structurale de l'extrémité du chromosome 19

La séquence du génome est de mauvaise qualité pour l'extrémité du chromosome 19. De plus, cette région porte de nombreux gènes de résistance tels que des gènes NBS-LRR et des gènes STK. Vu la complexité de la zone où se trouvent les loci  $R_{US}$  et  $R_1$ , le premier travail de cette thèse est de caractériser correctement l'extrémité du chromosome 19.

Le séquençage du génome du Peuplier, réalisé techniquement par le Joint Genome Institute de la DOE (Californie, USA), a été mené en technologie WGS. Il s'agissait de l'un des tous premiers séquençages en WGS et les lectures ne portaient que sur des extrémités d'inserts de quelques Kb, sans liens clones de mini-BAC, de Fosmides ou de clone BAC. Cette absence est en majeure partie responsable de la mauvaise qualité de l'assemblage, le statut hétérozygote constituant la partie mineure. Une amélioration de la qualité de l'assemblage nécessite la construction de liens clones complémentaires, de plus grande longueur ; les clones BAC constituent l'approche de choix pour cette ressource.

Une banque BAC de l'individu *P. trichocarpa* 101-74 porteur de l'allèle d'intérêt  $R_{US}$ , a été construite à l'URGV. Elle est utilisée tout d'abord pour la construction d'une carte physique locale.

### 7.2.2 Identification du gène $R_{US}$ par clonage positionnel

Les résistances à *Mlp* introduites dans les peupleraies cultivées ont été jusqu'alors uniquement des résistances qualitatives. Le rapide contournement de ces résistances oblige donc à explorer d'autres pistes. Une autre approche peut être l'étude d'une résistance quantitative contrôlée par le locus  $R_{US}$ , qui produirait une pression de sélection moins importante sur *Mlp*.

L'objectif principal de ma thèse porte sur l'identification du locus  $R_{US}$  par clonage positionnel (figure 22).

Le clonage positionnel consiste à déterminer précisément la position d'un gène dans le génome grâce à des marqueurs moléculaires et à isoler un fragment d'ADN qui contient ce gène. Le clonage positionnel est envisageable puisque le ratio de la distance physique sur la distance génétique est de 200 Kb/cM, soit environ deux fois celui observé chez *A. thaliana* (Yin *et al.*, 2004a).



### **7.2.3 Identification du gène $R_1$ par cartographie comparée**

L'objectif secondaire est l'identification du locus  $R_1$ . En l'absence d'une banque BAC de l'individu 73028-62 déjà construite, l'identification du gène  $R_1$  ne peut être menée que par cartographie comparée, en s'appuyant sur la carte physique et génétique du 101-74 et la séquence du génome de Nisqually-1.

Comme les loci  $R_1$  et  $R_{US}$  sont sur la même extrémité du chromosome 19, cette démarche pourra se faire en parallèle avec le clonage du locus  $R_{US}$ . De plus, elle permettra de connaître la relation entre les deux loci : s'agit-il de deux allèles d'un même gène ou de deux gènes différents ?



**Partie II**  
**MATERIELS ET METHODES**



# Partie II

## MATERIELS ET METHODES

<b>1</b>	<b>Matériel Biologique .....</b>	<b>46</b>
1.1	Famille 54 .....	46
1.2	Clone Nisqually-1 .....	46
1.3	<i>Melampsora larici-populina</i> .....	46
<b>2</b>	<b>Ressources Génomiques .....</b>	<b>47</b>
2.1	Banque BAC de l'individu 101-74 <i>P. trichocarpa</i> .....	47
2.1.1	Construction de la banque BAC.....	47
2.1.2	Caractéristiques de la banque BAC.....	47
2.1.3	Organisation des pools .....	47
2.2	Séquences d'extrémité des clones BAC .....	48
<b>3</b>	<b>Méthodes .....</b>	<b>48</b>
3.1	Méthode générale .....	48
3.2	Evaluation de la résistance à <i>Mlp</i> .....	49
3.2.1	Inoculation .....	49
3.2.2	Analyse de la taille des sores : le marqueur génétique $R_{US}$ .....	50
3.3	Développement de marqueurs .....	50
3.3.1	Vérification de la spécificité. ....	50
3.3.2	Types de locus.....	51
3.3.3	Développement de marqueurs génétiques sur des séquences de produits d'amplification PCR .....	52
3.4	Réactions PCR et électrophorèse.....	54
3.4.1	Protocole PCR.....	54
3.4.2	Electrophorèse.....	55
3.4.3	Test d'amplification des couples d'amorces.....	55
3.5	Cartographie génétique .....	56
3.5.1	Stratégie pseudo-testcross.....	56
3.5.2	Génotypage de la famille 54 .....	56
3.5.3	Construction des cartes génétiques locales .....	56
3.5.4	Carte génétique consensus .....	57
3.5.5	Carte génétique fine.....	58
3.6	Cartographie physique.....	58
3.6.1	Criblage de la banque BAC de l'individu 101-74.....	58
3.6.2	Contigage des cartes physiques.....	59
3.6.3	Assemblage des contigs.....	60
3.6.4	Extraction de BAC .....	60
3.6.5	Evaluation de la taille des cartes physiques.....	61
3.6.6	Fingerprints de clone BAC .....	62
3.7	Analyse de séquences .....	62
3.7.1	Annotation automatique. ....	62
3.7.2	Identification des gènes NBS-LRR .....	62
3.7.3	RepBase .....	63
3.7.4	Alignement de séquence de grande taille .....	63



# 1 Matériel Biologique

## 1.1 Famille 54

La famille 54 est la famille principalement utilisée dans ce travail. Elle est issue de croisements contrôlés entre la mère *P. deltoides* (clone 73028-62, Etat de l'Illinois, USA) et le père *P. trichocarpa* (clone 101-74, Etat de Washington, USA). Le parent femelle (73028-62) est hétérozygote au locus  $R_1$  et donc de génotype  $[R_1/r_1]$ . Le parent mâle (101-74) est hétérozygote au locus  $R_{US}$ , donc de génotype  $[R_{US}/r_{US}]$ . Des 348 premiers individus utilisés dans les études précédentes, 336 ont pu être utilisés dans cette thèse (nommés série 0). Ils sont conservés en pépinière à l'INRA d'Orléans avec 2 pieds par clone.

Une première extension de la famille 54 a été réalisée en 2003 pour obtenir 1079 individus supplémentaires. Cette campagne supplémentaire de croisements contrôlés a été effectuée en trois séries successives (nommées série 1, 2 et 3). Ils sont conservés en pépinière à l'INRA d'Orléans avec 3 pieds par clone.

Une seconde extension de la famille 54 a été réalisée pour obtenir 729 individus supplémentaires, durant cette thèse en 2007 (nommée série 4).

La taille de la famille 54 étudiée s'élève donc à 2 144 individus. Les individus sont nommés par 66130 qui définit la famille 54 puis par le numéro propre à chaque individu.

## 1.2 Clone Nisqually-1

Des copies végétatives de l'individu femelle Nisqually-1, utilisées pour la séquence publiée (Tuskan *et al.*, 2006), sont présentes à l'INRA d'Orléans. Cet individu, sensible à *Mlp*, forme de gros sores (Catherine Bastien, donnée personnelle); il est donc supposé ne pas porter l'allèle  $R_{US}$ .

## 1.3 *Melampsora larici-populina*

Dans cette étude, les souches de *Mlp* 93ID6 et 98AG69 ont été sélectionnées pour phénotyper la famille 54. La souche 93ID6 (souche référence pour la définition de  $R_1$ ) possède les virulences 3 et 4, ce qui permet de suivre la ségrégation de l'allèle  $R_1$  car elle ne possède pas la virulence 1. Les individus de la famille 54 sont par conséquent résistants ou sensibles à cette souche. Cette souche permet aussi de suivre la ségrégation de l'allèle  $R_{US}$  dans le cas où l'individu est sensible : la taille des sores peut donc être évaluée quand l'individu est de



phénotype  $r_1$ . La souche 98AG69 (souche référence pour la définition de  $R_{US}$ ) possède les virulences 1, 3, 4, 5 et 7. Elle permet de suivre la ségrégation de l'allèle  $R_{US}$  pour la totalité de la famille 54 car la virulence 1 contourne le locus  $R_1$  : tous les individus sont sensibles à cette souche et l'on peut donc mesurer les différents paramètres quantitatifs dont la taille des sores.

## 2 Ressources Génomiques

### 2.1 Banque BAC de l'individu 101-74 *P. trichocarpa*

#### 2.1.1 Construction de la banque BAC

La banque BAC a été constituée à partir du génotype 101-74 *P. trichocarpa* père, de la famille 54 ; ce clone a été choisi car il est porteur de l'allèle  $R_{US}$ . La banque a été construite à l'URGV par Boulos Chalhoub et Patricia Faivre Rampant à partir des noyaux extraits de feuilles étiolées et avec différentes concentrations de l'endonucléase de restriction *HindIII* (Chalhoub *et al*, 2004). Le vecteur utilisé est Pindigo (Epicenter), porteur d'un gène de résistance à l'antibiotique chloramphénicol, qui a été introduit dans la bactérie *Escherichia coli*, souche DH10B. La banque est constituée de 52 224 clones répartis dans 136 plaques de 384 puits chacune.

#### 2.1.2 Caractéristiques de la banque BAC

L'évaluation de la taille de 93 clones BAC a permis d'estimer la taille moyenne des inserts à 150 Kb. Des marqueurs spécifiques du génome chloroplastique et mitochondrial ont été utilisés et ont montré une amplification dans respectivement 3% et 0,3% des clones. La représentativité de cette banque est de 15 fois la taille du génome haploïde du peuplier (équivalents génome de la banque = 15 X). Il y a donc théoriquement 7,5 X pour l'allèle  $R_{US}$  (Giraut, 2005).

#### 2.1.3 Organisation des pools

Pour optimiser le criblage, la banque a été organisée en pools plaques et en pools lignes. Un pool plaque correspond au mélange de 384 clones d'une plaque, un pool ligne est le mélange des 24 clones d'une ligne (Giraut, 2005).



## 2.2 Séquences d'extrémité des clones BAC

Les séquences d'extrémité des clones BAC (BES) provenant des 20 premières plaques de la banque BAC ont été déterminées au Human Center of Stanford University (ce qui représente 7680 clones BAC = 15 360 séquences, la couverture en clone représente 2,3 X). Elles ont été obtenues avec la méthode de Sanger. La qualité des séquences a été vérifiée avec une fenêtre glissante de 50 pb : les bases ont été éliminées quand le score de qualité était en dessous de 20. Il en résulte 13 249 séquences (815 clones BAC n'ont pas de BES, 481 clones n'ont qu'un seul BES et 6384 clones ont leurs 2 BESs), la longueur moyenne des BESs est de 800 pb. Les séquences du vecteur n'ont pas été éliminées dans cet ensemble. Pour les clones BAC correspondants à la zone d'intérêt, ces séquences ont été recherchées et éliminées manuellement.

## 3 Méthodes

### 3.1 Méthode générale

Pour cloner un gène par clonage positionnel, deux approches complémentaires sont généralement utilisées : le chromosome "landing" et le chromosome "walking".

- Définition du chromosome landing : des marqueurs génétiques sont utilisés pour cribler une banque BAC. Les clones BAC chevauchants sont assemblés en contigs. Ces contigs sont allongés et ancrés à la carte génétique par l'intermédiaire des marqueurs génétiques (Ren *et al*, 2005).

- Définition du chromosome walking (marche sur le chromosome) : des marqueurs ou des sondes spécifiques d'une zone sont utilisés pour le criblage d'une banque BAC. Les BES sont produites et utilisées pour développer de nouveaux marqueurs ou sondes, devant cribler de nouveau la banque BAC. Plusieurs cycles sont nécessaires pour que des clones BAC recouvrent la totalité de la région étudiée (Ren *et al*, 2005).

Le jeu des premiers marqueurs génétiques qui permet de cribler la banque BAC, est souvent insuffisant pour mener à terme le clonage positionnel (chromosome landing). Il est nécessaire de poursuivre avec "la marche sur le chromosome" (chromosome walking).

Dans cette thèse, la stratégie est différente du fait de la disponibilité de la séquence du génome de Nisqually-1. Après l'identification des premiers marqueurs génétiques à proximité des loci de résistance, il n'était pas nécessaire de faire une marche sur le chromosome ; en effet, les premiers marqueurs génétiques indiquaient une zone précise sur un chromosome et il était donc possible de développer de nouveaux marqueurs directement sur la séquence du génome.

**A : Dispositif expérimental**



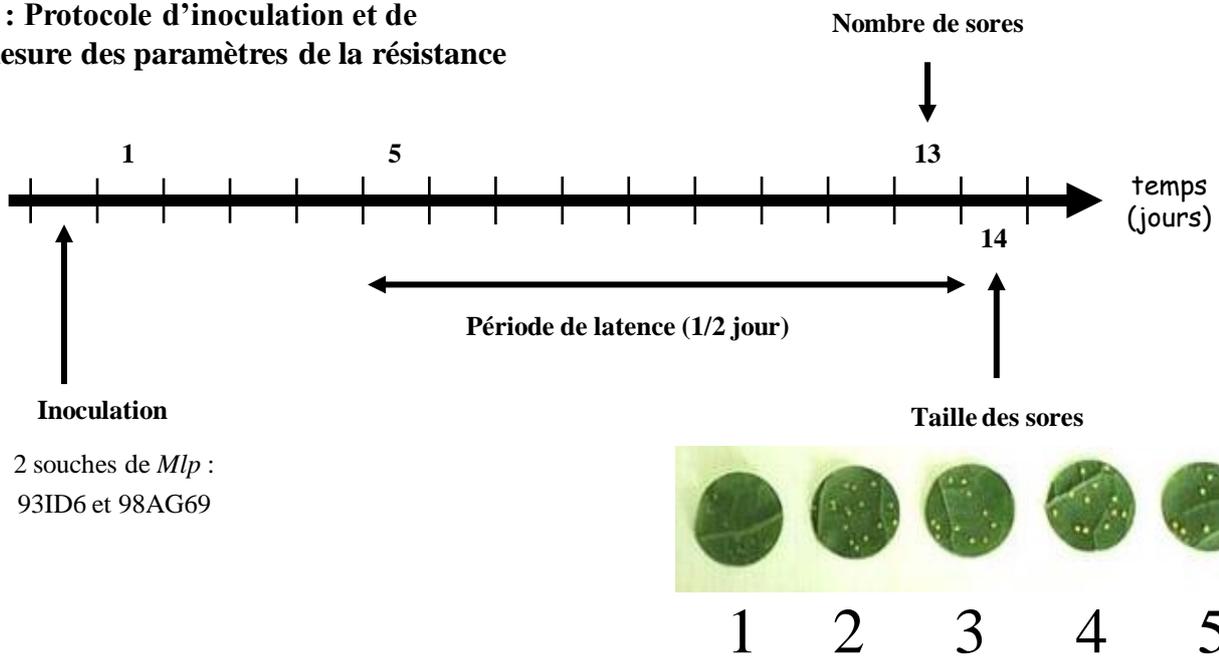
(Dowkiw, 2003)

**B : Disques foliaires après inoculation**



(Dowkiw, 2003)

**C : Protocole d'inoculation et de mesure des paramètres de la résistance**



**Figure 23** : Protocole expérimental du phénotypage de la famille 54.

A : disques foliaires de 3 cm en flottaison sur de l'eau distillée, 14 jours après inoculation par *Mlp*. B : Dispositif expérimental en condition contrôlée. C : Protocole d'inoculation et de mesure de la résistance qualitative et des différents paramètres de la résistance quantitative.

Pour la localisation de  $R_1$ , la démarche est la cartographie comparée ; l'établissement des cartes physiques et génétiques de l'individu 101-74 *P. trichocarpa*, en plus de la séquence du génome de l'individu Nisqually-1 *P. trichocarpa*, ont servi pour développer la carte génétique de l'individu 73028-62 porteur de l'allèle  $R_1$ , d'une autre espèce *P. deltoides*. Cette étape permettait de connaître les différences entre ces deux espèces pour l'extrémité du chromosome 19 et ainsi de savoir si cette cartographie comparée pouvait être utile à la localisation du locus  $R_1$ .

## 3.2 Evaluation de la résistance à *Mlp*

L'évaluation phénotypique de la résistance des individus de la famille 54 a été réalisée en conditions contrôlées sur des disques foliaires. L'évaluation de la résistance à *Mlp* et l'analyse des résultats a été réalisée à l'INRA d'Orléans par Arnaud Dowkiw et Catherine Bastien (Dowkiw *et al*, 2003).

Les souches de *Mlp* sont conservées à -80 °C. Les souches de pathogènes référencées en virulence sont multipliées sur des feuilles du cultivar "Robusta" qui ne possède pas de résistance qualitative.

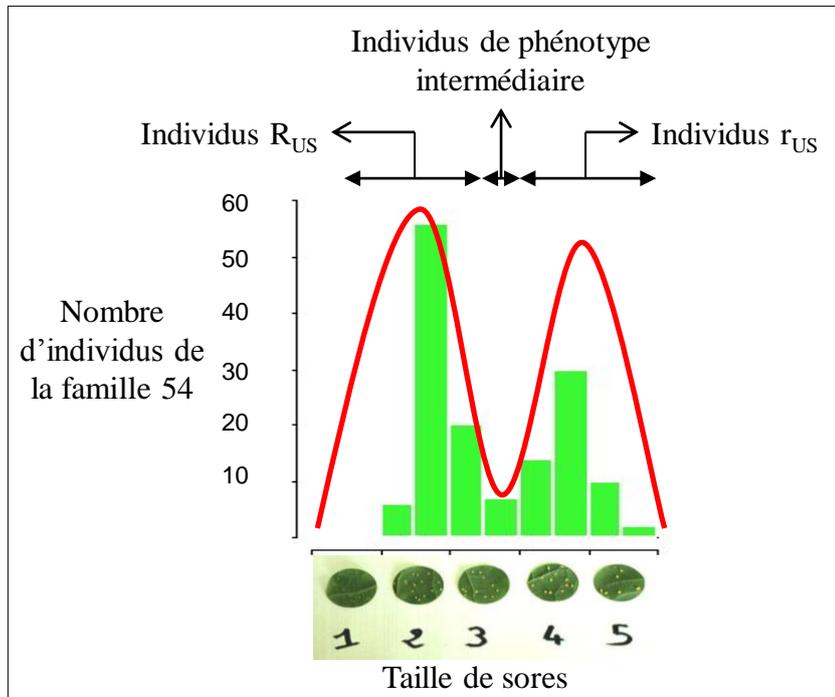
### 3.2.1 Inoculation

Le bouturage des individus de la famille 54 et des individus témoins est réalisé dans un pot de 3 l en serres, à l'abri des contaminations extérieures de *Mlp*.

Le dispositif expérimental est organisé en 5 blocs randomisés dans une chambre de culture climatisée (15 °C, 16 h de lumière par jour) (figure 23 A). Une expérience peut réunir jusqu'à 350 individus. Dans chaque bloc, chaque individu est représenté avec un disque foliaire de 3 cm de diamètre en flottaison sur de l'eau distillée, la face inférieure de la feuille vers le haut (figure 23 B). L'inoculation du champignon est faite par pulvérisation à l'aide d'un pulvérisateur à chromatographie.

Le clone de peuplier Robusta sert de témoin positif de l'inoculation et des cultivars discriminants permettent de vérifier la combinaison en virulence des souches utilisées. L'inoculation de boîtes de gélose permet une semi-quantification de l'inoculum et la vérification de l'homogénéité d'inoculation (figure 23 A). L'expérience est renouvelée pour chacune des souches de *Mlp*. La souche 93ID6 permet de suivre la ségrégation de la résistance qualitative  $R_1$  et de mesurer les paramètres de la résistance quantitative pour les individus sensibles et la souche 98AG69 permet de mesurer les paramètres de la résistance quantitative pour l'ensemble des individus.

Différents paramètres de la résistance sont mesurés (figure 23 C) :



**Figure 24** : Distribution des moyennes phénotypiques clonales d'une série d'individus de la famille 54 pour la taille des sores. Les individus porteurs de gros sores sont déterminés porteur de  $r_{US}$  et les individus porteurs de petits sores sont déterminés porteur de  $R_{US}$ , entre les deux les individus sont de phénotype intermédiaire.

- La période de latence, définie comme le temps écoulé entre l'inoculation et l'apparition du premier sore, est mesurée avec une précision d'1/2 journée entre le 5<sup>ème</sup> et le 13<sup>ème</sup> jour après inoculation.

- Le nombre de sores est compté 13 jours après inoculation.

- La taille des sores est évaluée 14 jours après inoculation. Cette évaluation est faite suivant une échelle à 5 classes, calibrée ensuite par analyse d'images en mm<sup>2</sup> (Dowkiw, 2003).

- La compatibilité (sensibilité) et l'incompatibilité (résistance) à la souche 93ID6 sont relevées 14 jours après inoculation.

### **3.2.2 Analyse de la taille des sores : le marqueur génétique $R_{US}$**

Pour analyser le phénotype taille des sores en cartographie génétique, le phénotype de la taille des sores a été transformé en un paramètre qualitatif. Pour chaque individu de la famille 54, il faut donc déterminer si l'individu est de phénotype gros sores  $r_{US}$  ou de phénotype petit sore  $R_{US}$  (figure 24). Des limites, inférieure et supérieure, sont fixées pour délimiter les individus porteurs de gros sores et de petits sores respectivement. Entre ces deux limites, les individus sont déterminés comme phénotype intermédiaire ; ils sont analysés en cartographie génétique comme des données manquantes.

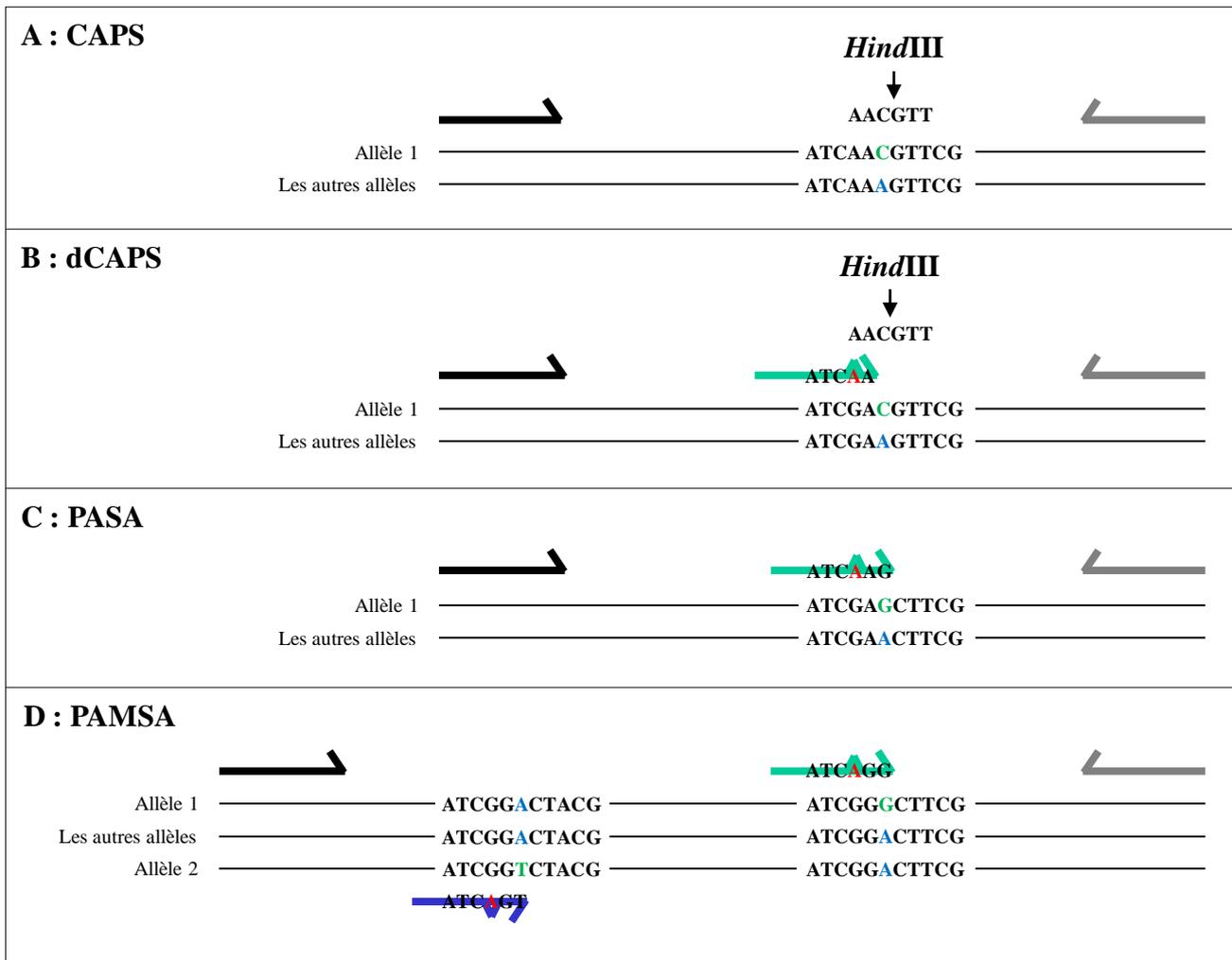
## **3.3 Développement de marqueurs**

Au cours de mon travail de nombreuses amorces PCR ont été dessinées pour développer de nouveaux marqueurs physiques et génétiques. Les amorces ont été dessinées à l'aide du logiciel Primer 3 (<http://fokker.wi.mit.edu/primer3>) avec les paramètres par défaut du logiciel sauf dans le cas indiqué ci dessous. Elles ont été dessinées avec une température d'hybridation de 59°C et une taille de 22 nucléotides dans l'idéal.

A de nombreuses reprises, plus de deux amorces ont été dessinées sur un même locus. Souvent, il y en a eu quatre : deux dans un sens et deux dans le sens inverse. Nous obtenons donc une combinaison de quatre couples d'amorces différents.

### **3.3.1 Vérification de la spécificité.**

La spécificité des couples d'amorces a été testée systématiquement avec les logiciels ePCR (Schuler, 1997) et BLASTn (Altschul *et al*, 1997). Cette vérification a été effectuée sur les deux assemblages du génome et les séquences des clones BAC. Lorsque les séquences n'étaient pas disponibles au moment du développement des couples d'amorces, la vérification a été faite *a posteriori*.



**Figure 26** : Marqueurs génétiques développés à partir de séquences de produits d'amplification PCR.

Les amorces noire et grise forment les couples d'amorces utilisés pour produire le produit d'amplification PCR séquencé. Les bases verte et bleue représentent les SNP découverts, la base verte est la version du SNP qui est utilisée pour observer la ségrégation dans la population  $F_1$ . (A) Marqueurs CAPS : le même couple d'amorces est utilisé pour produire l'amplification PCR, l'endonucléase de restriction (ici *HindIII*) est ensuite utilisée pour digérer l'allèle 1 spécifiquement. (B) Marqueur dCAPS : les amorces verte et grise sont utilisées pour produire l'amplification PCR, l'endonucléase de restriction est ensuite utilisée pour digérer l'allèle 1 spécifiquement. L'amorce verte a permis d'introduire une base différente (en rouge) nécessaire pour créer le site de digestion sur l'allèle 1. (C) Marqueur PASA : les amorces verte et grise sont utilisées pour amplifier spécifiquement l'allèle 1. La base rouge représente le mésappariement supplémentaire qui permet à l'amplification de ne pas fonctionner sur les autres allèles. (D) : Marqueurs PAMSA : Les trois amorces bleue, verte et grise sont utilisées dans une seule réaction PCR, les amorces verte et grise sont utilisées pour amplifier spécifiquement l'allèle 1, les amorces bleue et grise sont utilisées pour amplifier spécifiquement l'allèle 2. La base rouge représente le mésappariement supplémentaire qui permet à l'amplification de ne pas fonctionner sur les autres allèles.

- Le logiciel ePCR permet de déterminer les éventuelles localisations de produit d'amplification PCR pour un couple d'amorces donné dans une séquence donnée. Cette localisation est donnée suivant une taille d'amplification maximum (ici, 1500 nucléotides) et un nombre maximum de mésappariements par amorce, excepté les 3 premières bases en 3' de l'amorce (ici, avec 3 mésappariements maximum).

- Le logiciel BLASTn a été utilisé pour aligner les amorces avec les séquences (une e-value inférieure à 30 et un mot minimal de départ de 7 nucléotides). Dans ce cas, il a fallu regarder les alignements des deux amorces formant un couple pour vérifier qu'une amplification est possible (donc que l'orientation des amorces sur la séquence est bonne) et que la taille d'amplification est cohérente. De même que pour ePCR, les résultats enregistrés sont pour un maximum de 3 mésappariements par amorce.

### 3.3.2 Types de locus

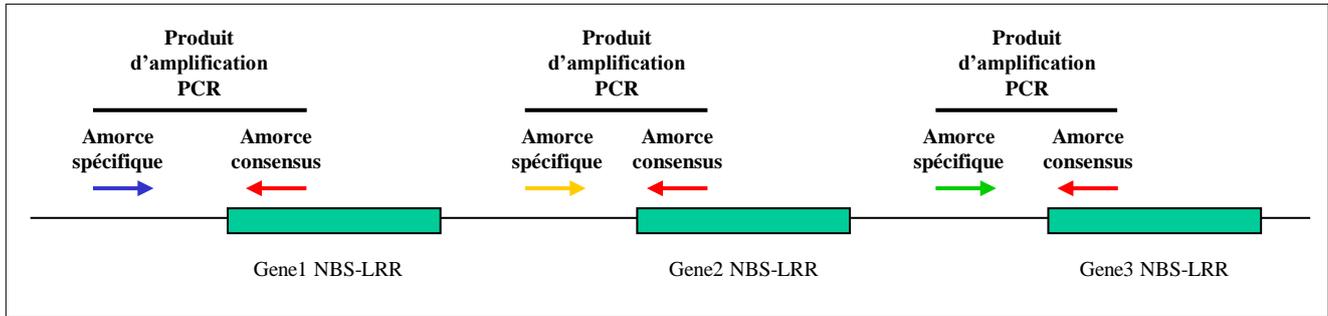
Différents types de locus ont été recherchés pour développer les couples d'amorces :

- Une recherche de SSR (simple séquence repeat) a été réalisée à l'aide du logiciel Sputnik (<http://www.cbib.u-bordeaux2.fr/pise/sputnik.html>) sur la séquence de l'extrémité du chromosome 19 ou avec RepBase (Cf. § II-3.7.3) pour les BES ou les séquences de clones BAC. Les couples d'amorces ont été dessinés avec une taille d'amplification très courte (75-150 pb) pour révéler au mieux le polymorphisme de taille éventuel entre les différents allèles.

- Des couples d'amorces ont été dessinés sur les gènes des séquences disponibles (STS génique (séquence tag site)). Les gènes NBS-LRR ont été sélectionnés en priorité. Des alignements des différents gènes redondants (MEGA4) ont été produits pour repérer les zones spécifiques du locus désiré. Pour les gènes les plus redondants, les couples d'amorces sont à cheval sur les gènes ; une amorce consensus aux différents gènes a été dessinée à l'intérieur du gène (qui est utilisée pour les différents couples) et des amorces spécifiques ont été dessinées à l'extérieur des gènes (figure 25).

- Des couples d'amorces ont été dessinés dans des zones inter-géniques (STS inter-génique). En majorité, ces couples d'amorces ont été dessinés d'après les BES, mais dans certains cas, des couples d'amorces ont aussi été dessinés sur la séquence du génome pour des zones importantes où il n'y avait pas de gène.

Pour les marqueurs STS, génique ou inter-génique, la taille recherchée est de 500 à 1 200 pb, ce qui donne plus de chance de trouver du polymorphisme dans le cas de séquençage de produit d'amplification PCR.



**Figure 25 :** Couples d'amorces développés dans les gènes TNL.

L'amorce rouge est une amorce fonctionnant à l'intérieur de plusieurs gènes TNL, la réaction PCR se fait avec cette amorce et une amorce spécifique d'un locus (bleue, orange ou verte) à l'extérieur du gène. Il en résulte une amplification PCR à cheval sur le gène et dont la spécificité provient de l'amorce spécifique.

### 3.3.3 Développement de marqueurs génétiques sur des séquences de produits d'amplification PCR

Des produits d'amplification PCR ont été séquencés (effectués chez MacroGen, Genoscreen ou ATGC) dans le but de découvrir du polymorphisme utilisable pour le développement de marqueurs génétiques. Ces produits d'amplification PCR ont été réalisés à partir d'ADNg, ou pour le 101-74, des clones BAC provenant des deux haplotypes. Dans la plupart des cas, le polymorphisme découvert était des SNP (single nucleotide polymorphism). Des couples d'amorces ont été développés par rapport à ce polymorphisme pour suivre la ségrégation de l'un des allèles d'un parent dans la descendance de la famille 54.

#### 3.3.3.1 Utilisation d'une endonucléase de restriction

Pour exploiter les SNP découverts, la première méthode a été d'utiliser une endonucléase de restriction pour marquer spécifiquement un allèle et d'en voir la ségrégation dans la population F<sub>1</sub>. Deux techniques ont été utilisées, le CAPS et le dCAPS :

- Les marqueurs CAPS (cleaved amplified polymorphic sequences) consistent à digérer un produit d'amplification PCR par une endonucléase de restriction au niveau d'un SNP ; la digestion se produit uniquement sur un des 4 allèles et produit des fragments de tailles différentes qui sont révélés par électrophorèse, ce qui permet de marquer spécifiquement cet allèle et d'en observer sa ségrégation dans la famille 54 (figure 26 A ; Konieczny & Ausubel, 1993).

- Les marqueurs dCAPS (derived CAPS) consistent à appliquer la même méthode que pour les marqueurs CAPS, mais le marqueur dCAPS utilise pour la PCR une amorce modifiée afin d'introduire un site spécifique d'une endonucléase de restriction à proximité d'un SNP (figure 23 B ; Michaels & Amasino, 1998).

Pour rechercher les endonucléases de restriction nécessaires aux marqueurs CAPS et dCAPS, le logiciel dCAPS finder 2.0 disponible sur Internet (<http://helix.wustl.edu/dcaps/dcaps.html>) a été utilisé (Neff *et al*, 2002).

Les digestions sont réalisées avec 5 µl de produit d'amplification PCR (Cf. § II-3.4.1), 1 unité d'endonucléase de restriction (Invitrogen), le tampon correspondant (1 X final) et de l'eau pour atteindre un volume final de 10 µl.



### 3.3.3.2 PCR spécifique d'un allèle

Pour exploiter les SNP découverts, la seconde méthode a été d'utiliser des amorces spécifiques d'un allèle pour marquer spécifiquement cet allèle et d'en voir la ségrégation dans la population F<sub>1</sub>. Deux techniques ont été utilisées, le PASA et le PAMSA :

- L'AS-PCR (allele specific PCR) aussi appelé PASA (PCR allele specific amplification) consiste à dessiner une amorce avec le SNP identifié à la première base en 3' de l'amorce (figure 26 C ; Gaudet *et al*, 2007 ; Bundock *et al*, 2006). Comme le SNP n'est pas suffisant pour que l'amplification se produise spécifiquement sur l'allèle possédant la version désirée du SNP, un mésappariement est introduit à la 3<sup>ème</sup> base en 3' de l'amorce ; l'amorce fonctionne correctement sur l'allèle possédant la version désirée du SNP (un mésappariement) mais ne fonctionne pas sur les autres allèles (deux mésappariements).

Les marqueurs AS-PCR offrent plus de possibilité que les marqueurs CAPS ou dCAPS ; en effet, on ne trouve pas toujours une endonucléase de restriction qui permet d'exploiter un SNP. De plus, les marqueurs AS-PCR sont très peu coûteux avec seulement une amorce supplémentaire.

Dans le cas où le polymorphisme découvert lors du séquençage de produit d'amplification était plus important qu'un simple SNP (plusieurs SNP côte à côte ou InDel (insertion/délétion), une amorce spécifique a été développée sur ce polymorphisme ce qui revient ainsi à un marqueur AS-PCR.

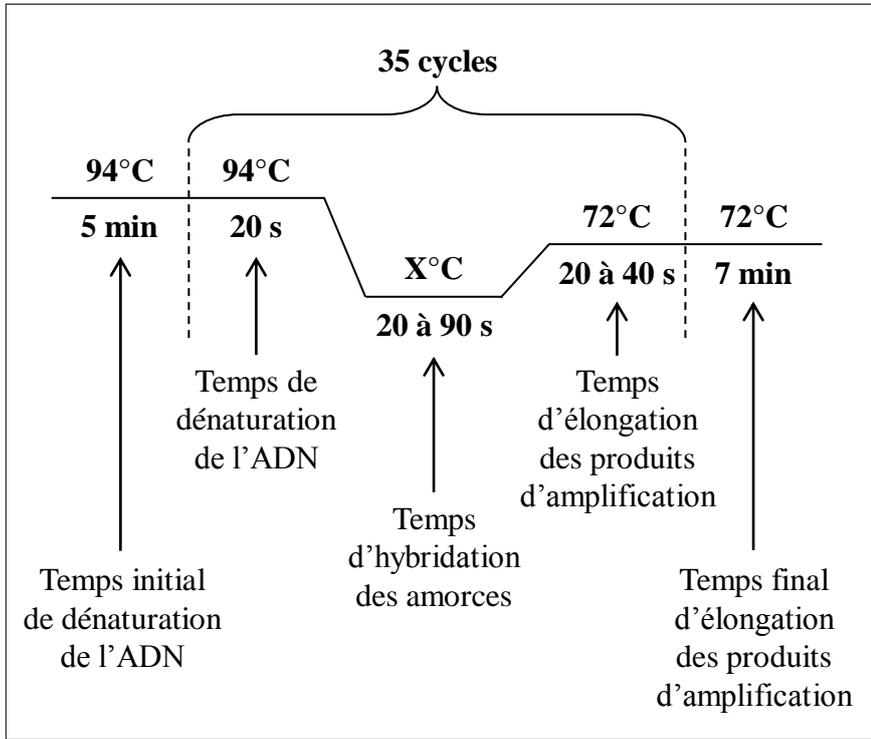
- Le PAMSA (PCR allele multiple specific amplification) consiste à dessiner plusieurs amorces spécifiques de différents allèles (figure 26 D). Chacune est dessinée pour être spécifique de l'un des deux allèles d'un des parents de la population. Pour développer ce type de marqueur, il était nécessaire de connaître l'appartenance à l'un ou l'autre des haplotypes pour chaque base polymorphe utilisée ; il était donc nécessaire de séquencer des produits d'amplification PCR provenant de clones BAC. Le marqueur génétique correspondant est donc co-dominant ; une amplification marque spécifiquement chacun des deux allèles d'un parent. Dans ces expériences, les réactions PCR sont faites simultanément avec 3 amorces.

### 3.3.3.3 Nomenclature des couples d'amorces

Une nomenclature des couples d'amorces a été faite et est décrite ci-dessous :

- SSR des publications : dans les tableaux et les écrits, ils sont notés GCPM\_Y, ORPM\_Y et PMGC\_Y et notés dans les figures G\_Y, O\_Y et P\_Y.

- SSR de l'INRA : dans les tableaux et les écrits, ils sont notés INRA\_X (ou X est un chiffre correspondant à la localisation sur la séquence utilisée) et notés dans les figures I\_X.



**Figure 27** : Protocole des réactions PCR.

- BES : ils sont notés avec la coordonnée du clone BAC puis –F ou –R suivant l'extrémité FP ou RP.

- STS dans les gènes R : ils sont notés RGAX

- STS dans les autres gènes : ils sont notés PTX

- STS inter-génique : ils sont notés ncX

Un “s” avant le chiffre “X” est ajouté pour les couples d'amorces dessinés sur un scaffold non assigné au chromosome 19.

Un “m” dénote les couples d'amorces dessinés sur le clone BAC provenant de l'individu S9-2 à proximité du locus *MER*.

Un “-AS” après le chiffre identifie les couples d'amorces dessinés par rapport à un produit d'amplification PCR et dont le but est d'avoir plus de spécificité.

Un “-dC” après le chiffre identifie les couples d'amorces développés pour un marqueur dCAPS

Un “-chiffre” apparait à la fin du nom dans le cas où il y a plusieurs couples d'amorces sur un même locus.

Un “R” avant le chiffre identifie les couples d'amorces dessinés sur un clone BAC de l'haplotype  $hR_{US}$ .

## 3.4 Réactions PCR et électrophorèse

### 3.4.1 Protocole PCR

Les réactions PCR sont réalisées dans le mélange de 9,6  $\mu$ l suivant : premix PCR 6  $\mu$ l (annexe 1), couple d'amorce 1  $\mu$ l (5  $\mu$ M pour chaque amorce, commandé chez Operon ou Invitrogen), Taq polymérase 0,1  $\mu$ l, ADN 2,5  $\mu$ l contenant 7 ng ou 2,5  $\mu$ l de culture bactérienne diluée (au 1/50<sup>ème</sup> lorsqu'il s'agit de pool ou au 1/100<sup>ème</sup> lorsqu'il s'agit de clone individuel). Les concentrations indiquées dans ce manuscrit sont celles initiales.

Les réactions PCR ont été effectuées suivant le protocole (figure 27) : 5 min à 94°C, suivi de 35 cycles de : 20 s à 94°C qui correspond à la dénaturation de l'ADN, 30 s à une température variable pour l'hybridation (Cf. ci dessous), 20 à 40 s à 72°C qui correspond à l'élongation des produits d'amplification, et enfin 7 min à 72°C pour le temps final d'élongation. Ces réactions PCR sont réalisées dans des thermocycleurs 96 ou 384 (Applied Biosystems, GeneAmp 9700).

Les températures d'hybridation des amorces ont été mises au point à partir d'une température de base de 60°C puis une baisse ou une augmentation progressive de 2°C jusqu'à trouver la température idéale. La température a été augmentée pour s'assurer d'une amplification spécifique du locus désiré, la température a été abaissée pour obtenir une amplification de bonne



qualité. Dans certains cas, le temps d'hybridation a été abaissé à 20 s pour avoir une meilleure spécificité (pour que les amorces s'hybrident moins facilement sur d'autre locus) et augmenté à 60 ou 90 s pour avoir une meilleure amplification (pour que les amorces s'hybrident mieux avec le ou les loci désirés). Les températures et le temps d'hybridation (s'il est différent de 30 s) sont donnés dans les annexes 3, 4, 5 et 6.

Le temps de l'élongation dépend de la taille du produit PCR : 20 s pour des amplifications inférieures à 500 pb, 30 s entre 500 et 1000 pb et 40 s pour 1000 pb et plus.

### **3.4.2 Electrophorèse**

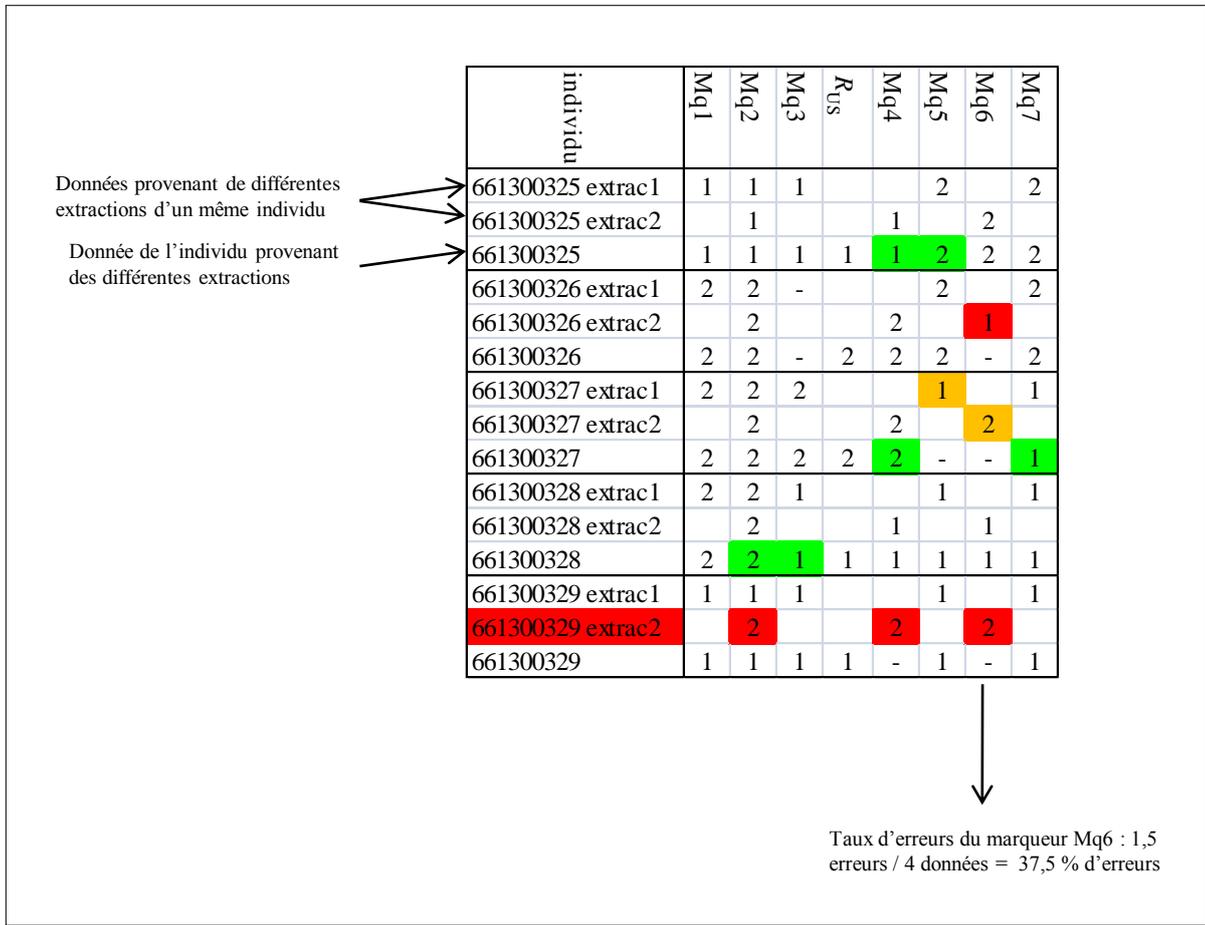
Les électrophorèses ont été réalisées dans des gels d'agarose de 1,5 % à 3,5 % (Lonza, SeaKem LE Agarose) avec du TAE 0,5 X (Lonza, AccuGENE TAE buffer : 2M Tris-Acetate, 0,05 EDTA, pH = 8,3) et quelques gouttes de bromure d'éthidium (0,625 mg/ml, une goutte pour 150 ml de gel). La concentration de l'agarose dépend de la taille des produits d'amplification attendus ; plus le produit d'amplification est grand, plus la concentration de l'agarose est faible. Ces gels peuvent contenir jusqu'à 384 puits de migration. Les migrations des produits d'amplification PCR ont été réalisées dans du TAE 0,5 X sous une intensité de 300 mA entre 20 min et 90 min. Le marqueur de taille 50 pb DNA Ladder (Invitrogen) a été utilisé pour mesurer les produits d'amplification PCR. L'ADN a été révélé avec un appareil photographique (BIO-RAD, Gel Doc 2000).

Certains marqueurs génétiques ont nécessité une migration sur gel d'agarose "métaphore" (Lonza, MetaPhor Agarose) ; cet agarose permet une meilleure résolution des produits d'amplification PCR, ce qui est utile quand il y a différents produits d'amplification de taille très proches à séparer.

### **3.4.3 Test d'amplification des couples d'amorces**

Les couples d'amorces développés sont testés sur l'ADNg des deux parents de la famille 54 : les individus 101-74 et 73028-62 et sur 6 descendants de la famille 54. Une partie des couples d'amorces ont aussi été testés sur l'ADNg de l'individu Nisqually-1 utilisé pour le séquençage du génome.

Les couples d'amorces issus de la bibliographie ont aussi été testés, il s'agit de SRR (Tuskan *et al*, 2004) et un STS (Yin *et al*, 2004b), ainsi que les différents couples d'amorces développés par rapport aux séquences de bandes AFLP (Masle, 2005).



**Figure 28** : Tableau des résultats de génotypage de la famille 54.  
Les recombinaisons sont indiquées en vert, les erreurs qui ont été éliminées sont indiquées en rouge.

## 3.5 Cartographie génétique

### 3.5.1 Stratégie pseudo-testcross

Le peuplier est une espèce dioïque et l'on ne possède pas de lignée fixée. De plus, le long temps de génération ne permet pas d'avoir facilement des populations  $F_2$  ou backcross. Les études de cartographie génétique se font donc généralement sur des populations  $F_1$ , créées à partir de deux individus hétérozygotes, comme c'est le cas dans cette thèse. La cartographie génétique se fait avec la stratégie du pseudo-testcross ; cette stratégie consiste à étudier la ségrégation dans la population  $F_1$  entre les deux allèles d'un parent, sans tenir compte des allèles apportés par l'autre parent. Grâce à la stratégie du pseudo-testcross, il est possible de construire deux cartes parentales lors d'une même expérience. La carte génétique du parent maternel *P. deltoides* (génotype  $[R_1/r_1]$ ) a été réalisée simultanément à celle du parent paternel *P. trichocarpa* (génotype  $[R_{US}/r_{US}]$ ).

### 3.5.2 Génotypage de la famille 54

L'ADNg des individus de la famille 54 a été extrait à l'aide de kit en plaque 96 (DNA easy kit, Quiagen). Les extractions ont été réalisées à l'INRA d'Orléans par Vanina Guérin.

La répartition de ces ADNg dans des plaques PCR de 384 puits a été effectuée à l'aide d'un robot (Beckman, Biomek 2000).

Les couples d'amorces qui ont révélés une ségrégation dans la famille 54, d'après les tests sur les parents et 6 descendants, sont ensuite utilisés en cartographie génétique de la famille 54. La plus grande partie de ces marqueurs génétiques a été utilisée sur les 336 premiers individus de la famille. Certains de ces marqueurs ont été sélectionnés, sur des critères de qualité et de localisation, pour génotyper les 1079 individus de la première extension de la famille 54. Une sélection différente a été faite pour la seconde extension de 729 individus de la famille 54.

### 3.5.3 Construction des cartes génétiques locales

#### 3.5.3.1 Ordonnement des marqueurs génétiques

Les électrophorèses de génotypage de la famille 54 ont été codées pour la construction des cartes génétiques ; un 2 est noté pour une donnée montrant la présence de l'allèle  $R_{US}$  pour un marqueur donné, et un 1 est noté pour une donnée montrant la présence de l'allèle  $r_{US}$ . De même (figure 28), un 2 est noté pour une donnée montrant la présence de l'allèle  $R_1$  pour un



marqueur donné, et un 1 est noté pour une donnée montrant la présence de l'allèle  $r_1$ . Chaque gel a été lu au moins deux fois indépendamment.

L'ordre des marqueurs est déterminé par rapport à l'ensemble des données de génotypage, c'est-à-dire tous les marqueurs et tous les individus. La place de chaque marqueur est donnée par rapport aux autres marqueurs afin de minimiser le nombre de recombinaisons.

Chaque individu a été analysé individuellement pour déterminer son génotype sur l'ensemble de l'extrémité du chromosome 19. La localisation des recombinaisons entre les deux haplotypes d'un des parents a été déterminée manuellement (en vert dans la figure 28).

Dans de nombreux cas, des erreurs se sont produites lors du génotypage de la famille 54 ou du codage des résultats. Ces erreurs de génotypage ont été déterminées manuellement. A l'origine, deux recombinaisons ne peuvent pas être détectées à des distances génétiques très proches. Plusieurs types d'erreurs ont été remarqués ; une erreur simple impliquant une double recombinaison autour d'un locus (cas de l'individu 661300326 pour le marqueur Mq6), une erreur au niveau d'une recombinaison, ce qui implique une triple recombinaison (cas de l'individu 661300327 pour les marqueurs Mq5 et Mq6). Dans ce dernier cas on ne peut pas savoir quelle est l'erreur parmi les deux possibles, et donc chacune est comptée comme une demi-erreur. Lorsqu'un individu a été extrait plusieurs fois, les données correspondant à chaque extraction ont été séparées pour voir d'éventuelles erreurs lors d'une extraction (cas de l'individu 661300329 pour la seconde extraction dans la figure 28).

Chaque erreur ou incohérence a été étudiée en particulier, notamment en revenant au gel de génotypage pour vérifier qu'il ne s'agissait pas d'une erreur de codage.

Pour chaque marqueur, le taux de données présentes, le taux d'erreur et le taux de distorsion ont été calculés.

### **3.5.3.2 Utilisation du logiciel MAPMAKER**

Les différentes cartes ont été construites avec le logiciel MAPMAKER (Lander *et al*, 1987). Les marqueurs validés pour cette construction présentent au moins 40% de données et doivent posséder moins de 3,5 % d'erreurs. Les erreurs restantes ont été éliminées et notées comme donnée manquante. Le logiciel est utilisé en fonction "Kosambi", sans détection d'erreur.

### **3.5.4 Carte génétique consensus**

La carte consensus entre les deux parents de la famille 54 a été produite en utilisant le logiciel Biomercator v2.1 (Arcade *et al*, 2004). Elle a été effectuée par Véronique Jorge en projetant la carte génétique maternelle *P. deltoides* sur la carte paternelle *P. trichocarpa*.



### **3.5.5 Carte génétique fine**

L'avancée du génotypage de la famille 54 a permis de travailler uniquement avec les individus porteurs d'une recombinaison proche des loci de résistance. Entre les limites fixées autour de ces loci, les distances génétiques sont calculées manuellement. On considère que le reste des individus est connu et non recombinant pour la zone étudiée. L'hypothèse est donc que ces individus ne possèdent pas de double recombinaison dans cette zone.

## **3.6 Cartographie physique**

### **3.6.1 Criblage de la banque BAC de l'individu 101-74**

Pour construire la carte physique, la première étape a été d'isoler, dans la banque BAC de l'individu 101-74, les clones BAC présents sur l'extrémité du chromosome 19.

Deux stratégies ont été utilisées : le criblage de la banque BAC avec des marqueurs PCR développés pour la zone d'intérêt, et l'utilisation des données de BES effectuées sur les 20 premières plaques de la banque en les alignant sur la séquence du génome.

#### **3.6.1.1 Criblage par PCR**

Certains marqueurs développés, qui fonctionnent sur l'ADNg de l'individu 101-74, sont utilisés pour trouver les clones BAC spécifiques de la région génomique désirée. Pour optimiser le criblage de ces clones BAC, celui-ci est d'abord conduit sur les pools plaques, puis les pools lignes correspondants aux pools plaques positifs, et enfin sur les clones individuels d'une ligne correspondante.

En premier lieu, ce criblage a été effectué sur 88 plaques de la banque (équivalent 10 X). La répartition de ces 88 pools plaque dans des plaques PCR de 384 puits a été effectuée à l'aide d'un robot (Beckman, Biomek 2000) pour obtenir quatre fois ces pools plaques dans une plaque PCR 384 puits et permet d'utiliser quatre marqueurs différent à la fois.

Quand la couverture n'est pas suffisante ou si la zone étudiée est particulièrement intéressante, le criblage par PCR a été poursuivi sur les 48 autres plaques avec le même mode expérimental.

Lorsque plusieurs couples d'amorces proches physiquement ont permis d'identifier un même pool plaque, un seul couple d'amorces a été utilisé pour effectuer le criblage des pools lignes et des clones individuels. Les clones BAC ainsi trouvés sont nommés par leurs coordonnées dans la banque BAC.

Certains de ces clones BAC ont été envoyés en séquençage pour leurs extrémités.



### 3.6.1.2 Criblage avec les BES des 20 premières plaques de la banque BAC

Les BESs des 20 premières plaques ont été alignées sur la séquence du génome de Nisqually-1 pour la version V1, puis après pour la version V2 (BLASTn avec une e-value inférieure à  $10^{-50}$ ). Les clones BAC sélectionnés sont ceux qui possèdent un des deux BES s'alignant correctement avec la zone d'intérêt, cet alignement devant être le meilleur de tous les résultats d'alignements sur le génome.

Pour valider ces clones BAC, ils ont été testés avec des marqueurs qui correspondent à la zone génomique que les deux BES délimitaient.

### 3.6.1.3 Culture des clones

Lorsqu'un clone BAC individuel a été identifié, une très petite quantité de la culture de ce clone a été étalée sur une boîte de Pétri contenant 20 ml de milieu de culture LB agar (37 g/l) avec l'antibiotique chloramphénicol (12,5 µg/ml).

A chaque fois, une colonie isolée dans la boîte de Pétri a été prélevée avec un cure-dent, qui a permis d'ensemencer une culture de 1 ml de milieu de culture FM glycérolé (annexe 1) et contenant l'antibiotique chloramphénicol (12,5 µg/ml). La culture a été menée à 37 °C pendant 15 h avec une agitation de 250 rpm. Ces cultures sont conservées au congélateur à -80 °C.

Ces cultures ont été diluées au 1/100<sup>ème</sup> dans de l'eau pour effectuer des PCR. Le marqueur utilisé pour le criblage du clone BAC a été utilisé pour vérifier que la colonie isolée correspond bien au clone BAC désiré.

## 3.6.2 Contigage des cartes physiques

Les différents clones BAC identifiés individuellement sont ensuite testés avec tous les autres marqueurs qui fonctionnent à proximité. Les différents clones BAC qui ont permis l'amplification d'un même marqueur sont chevauchants, et de proche en proche, il a été possible de construire des contigs de clones BAC. Les marqueurs qui ont été développés dans les BES ont permis de délimiter les clones BAC dans les contigs et d'orienter les 2 BES d'un clone BAC.

Les marqueurs génétiques ont permis de différencier les clones BAC provenant de chaque haplotype. Les contigs de clones BAC ont été assemblés indépendamment pour chaque haplotype. Il en résultait deux cartes physiques indépendantes, une pour chaque haplotype. L'haplotype porteur de l'allèle  $R_{US}$  a été appelé l'haplotype "h $R_{US}$ ", l'haplotype porteur de l'allèle  $r_{US}$  a été appelé l'haplotype "h $r_{US}$ ".



### **3.6.3 Assemblage des contigs**

En premier, l'assemblage des contigs a été construit en utilisant leurs alignements avec la carte génétique. En second, pour les finitions, les cartes physiques ont été alignées sur le génome de Nisqually-1 (version V1 et V2) à l'aide des marqueurs PCR et des BES.

#### **3.6.3.1 Alignement *via* les marqueurs PCR**

L'alignement avec les marqueurs PCR a été établi avec les résultats des logiciels ePCR et BLASTn (Cf. § II-3.3.1). Seuls les résultats avec un maximum de 2 mésappariements ont été pris en compte dans ces alignements.

#### **3.6.3.2 Alignement *via* les BES**

Les séquences des clones BAC, dont la présence sur l'extrémité du groupe de liaison a été prouvée, ont été alignées sur le génome (BLASTn, e-value inférieure à  $10^{e-20}$ ). Les meilleurs résultats d'alignement ont été enregistrés pour chacun des BES.

Ces résultats ont été confirmés avec l'alignement entre le BES et la ou les séquences correspondantes du génome à l'aide du logiciel MEGA4. Cet alignement manuel a permis de déterminer si ce résultat était significatif (> 70 % d'alignement et > 80 % d'identité). De plus, sur certains BES, il y avait des séquences de faible complexité (par exemple, les microsatellites) qui ont été filtrées par le logiciel BLASTn et qui découpent les alignements. L'alignement manuel par MEGA4 a permis de contourner ce problème et de valider le raboutement des alignements.

Ces séquences ont aussi été soumises sur RepBase pour détecter les éléments répétés, notamment les éléments transposables (Cf. § II-3.7.3).

L'orientation des 2 BES d'un clone BAC a été effectuée en premier lieu avec les marqueurs PCR dessinés sur les BES ; le marqueur qui fonctionne sur les clones BAC chevauchant d'un côté du clone BAC, constitue une vérification expérimentale. En second lieu, l'orientation a été effectuée avec l'alignement sur la séquence du génome ; mais il s'agissait là d'un travail *in silico*, sans preuve expérimentale.

### **3.6.4 Extraction de BAC**

Une pré-culture a été réalisée à partir de 10 µl de la culture d'un clone BAC dans 0,5 ml de milieu 2YT (annexe 1) avec 12,5 µg/ml de chloramphénicol. Elle a été mise en culture à 37°C pendant 24 h avec une agitation de 250 rpm. Puis une seconde culture,ensemencée avec 5 µl de



la pré-culture, a été menée dans 1,5 ml de milieu de culture 2YT avec 12,5 µg/ml de chloramphénicol à 37 °C pendant 15 h avec une agitation de 250 rpm.

L'extraction plasmidique a été menée par la méthode de lyse alcaline modifiée ; le protocole de cette méthode est résumé ci dessous.

La culture a été centrifugée (1500 g, 30 min, T° ambiante) pour sédimenter les bactéries. Le culot a été repris dans 100 µl de GET (annexe 1) et de la RNase (100 µg/ml). Un ajout de 200 µl de NaOH/SDS (annexe 1) a été effectué, suivi d'un repos pendant 10 min. Un ajout de 1500 µl d'AcK 5/3 M (annexe 1) a été effectué, suivi d'un repos pendant 10 min dans la glace puis d'une centrifugation (1500 g, 40 min, T° 4°C). Le surnageant est repris dans un volume équivalent (450 µl) d'isopropanol à -20°C, et la précipitation a été effectuée pendant 40 min à -20°C. Une centrifugation a été effectuée (1500 g, 30 min, T° ambiante) et le culot a été lavé deux fois avec 500 µl d'éthanol à 70° à -20°C. Les culots sont repris dans 20 µl d'eau pendant une nuit à 4°C.

Les extractions plasmidiques ont été quantifiées par migration sur gel d'agarose d'un µl de l'extraction plasmidique avec une gamme d'ADN λ digéré par *HindIII*, dont la concentration était connue.

### **3.6.5 Evaluation de la taille des cartes physiques**

#### **3.6.5.1 Evaluation de la taille**

Après l'extraction du vecteur, une digestion a été effectuée avec l'endonucléase de restriction *NotI* pour libérer l'insert. Elle a été effectuée avec 10 µl de l'extraction plasmidique, 1 unité de l'endonucléase de restriction *NotI* (Invitrogen), le tampon correspondant (1X final) et de l'eau pour atteindre un volume final de 20µl.

La migration du produit de digestion a été effectuée dans un gel d'agarose 1%, TBE 0,5 X (annexe 1) et une électrophorèse en champs pulsé (BIO-RAD, CHEF MAPPER, angle de 120°, gradient de 6V/cm). La migration est menée pendant 14 h avec un marqueur de taille (échelle de 50 Kb, BioLabs, Lambda Ladder PFG Marker).

#### **3.6.5.2 Prise en compte des mesure des clones BAC dans les cartes physiques**

Une coordonnée fictive pour chaque marqueur et de chaque BES a été donnée pour les deux cartes physiques. Cette opération a été principalement réalisée premièrement en utilisant les coordonnées du génome de Nisqually-1 V2, ensuite ces coordonnées ont été corrigées avec les données de longueur des clones BAC ; ceci a permis de construire les figures des cartes



physiques et aussi d'aligner ces cartes avec les autres données telles que la séquence du génome, ou avec les cartes génétiques, tout en gardant une échelle physique.

### **3.6.6 Fingerprints de clone BAC**

Avant d'envoyer des clones BAC en séquençage, le fingerprint de ces clones a été effectué pour vérifier qu'il n'y avait pas eu de réarrangement dans l'insert avant séquençage. Après l'extraction du vecteur (Cf. § II-3.6.4), une digestion a été effectuée avec l'endonucléase *HindIII*. Elle a été effectuée avec 20 µl de l'extraction plasmidique, 1 unité de l'endonucléase de restriction *HindIII* (Invitrogen), le tampon correspondant (1X final) et de l'eau pour atteindre un volume final de 40µl.

La migration du produit de digestion a été effectuée dans un gel d'agarose 1%, TBE 0,5 X (annexe 1) avec une électrophorèse en champs pulsé (angle de 120°, gradient de 6V/cm). La migration a été menée pendant 16 h avec un marqueur de taille 1 Kb+ (Invitrogen).

## **3.7 Analyse de séquences**

### **3.7.1 Annotation automatique.**

Le logiciel EUGENE a été utilisé pour l'annotation des clones BAC. Eugene est un réconciliateur de différents logiciels avec ses propres algorithmes (effectué par Cécile Guichard) :

- splice machine afin de déterminer les sites d'épissage.
- BLASTx sur deux banques protéiques : 1<sup>er</sup> alignement contre swissprot, 2<sup>ème</sup> alignement contre les protéomes d'*A. thaliana*, *Medicago*, le peuplier et le riz.
- Genome threader (équivalent de sim4) qui fait de l'alignement global de séquences EST/cDNA sur une banque peuplier de 342 475 séquences.

### **3.7.2 Identification des gènes NBS-LRR**

#### **3.7.2.1 Motif MEME**

Les motifs MEME développés par Kolher *et al*, (2008) et Meyers *et al*, (2003) ont été utilisés pour trouver les gènes NBS-LRRs du chromosome 19 (Annexe 2). Les motifs MEME de Meyers ont été développés sur la séquence des gènes NBS-LRR du génome d'*A. thaliana* et ceux de Kolher sur la séquence des gènes NBS-LRR du génome de *P. trichocarpa* version V1. Il s'agit de séquences protéiques qui ont été alignés sur les séquences à l'aide du logiciel tBLASTn



(avec une e-value inférieure à 5 sauf pour les différents domaines LRR où la e-value est inférieure à 10). Les différents alignements des différents motifs MEME ont été raboutés suivant leur sens sur la séquence, leur ordre et la taille qui les sépare afin d'identifier des gènes.

### **3.7.2.2 Phylogénie de gènes NBS-LRR**

Pour une phylogénie des gènes NBS-LRR, les domaines NBS ont été alignés entre eux avec le logiciel MEGA4 ; ce domaine est présent dans un seul exon quels que soient les gènes.

La phylogénie a été construite avec le logiciel MEGA4 suivant la méthode du « Neighbor-joining » et avec un bootstrap de 500 réplifications.

### **3.7.3 RepBase**

La soumission des séquences des clones BAC et des BES a été faite sur le site de GIRI (<http://www.girinst.org>) en utilisant le programme CENSOR et la base de données d'ETs RepBase (release 15.10). Cela a permis de connaître et de filtrer les séquences de faible complexité. Cela a surtout permis d'aligner les séquences aux éléments transposables (ETs) présents dans la base de données. Le raboutement des différents alignements a été effectué en regardant la taille, le score, l'orientation des différents alignements, la catégorie des éléments transposables et la distance qui sépare les différents alignements.

### **3.7.4 Alignement de séquence de grande taille**

Les séquences de grande taille ont été alignées avec deux logiciels différents : le logiciel Artemis Comparaison Tool et le logiciel Dot Plot.

- Le logiciel Artemis Comparaison Tool (Release 9) utilise les résultats d'alignements obtenus avec le logiciel megaBLAST, sans filtre de faible complexité pour qu'il n'y ait pas de coupure dans les alignements et une e-value faible (inférieure à  $10^{-100}$ ) pour qu'il n'y ait pas trop de résultats parasites. Le logiciel Artemis Comparaison Tool effectue une représentation de ces résultats, et il est possible de sélectionner les alignements représentés avec le score donné par megaBLAST.

- Le logiciel Dot Plot produit directement des alignements entre deux séquences, la première en abscisse et la seconde en ordonnée. La présentation des résultats permet une analyse plus exhaustive et réelle que celle que fournissait le logiciel Artemis Comparaison Tool.



**Partie III**  
**ORGANISATION STRUCTURALE**  
**DE L'EXTREMITÉ DU CHROMOSOME 19**



# Partie III

## ORGANISATION STRUCTURALE DE L'EXTREMITE DU CHROMOSOME 19

<b>1</b>	<b>Introduction.....</b>	<b>65</b>
1.1	Vers le clonage positionnel de $R_{US}$ .....	65
1.2	Vers le clonage positionnel de $R_1$ .....	65
<b>2</b>	<b>Résultats.....</b>	<b>66</b>
2.1	Séquence du génome de Nisqually-1.....	66
2.1.1	D'après la version V1 du génome .....	66
2.1.2	D'après la version V2 du génome .....	66
2.1.3	Les différences entre les deux versions de l'assemblage du génome .....	67
2.1.4	Les gènes de résistance.....	67
2.2	Développement de marqueurs .....	70
2.2.1	Sur la séquence du génome de Nisqually-1 .....	70
2.2.2	Sur les BES.....	71
2.2.3	Clone BAC <i>P. deltoides</i> proche du locus <i>MER</i> .....	71
2.3	Test des couples d'amorces .....	71
2.3.1	Test des parents de la famille 54 .....	71
2.3.2	Test sur un échantillon de la famille 54.....	72
2.3.3	Développement de marqueurs sur des produits d'amplification PCR .....	73
2.4	Cartographie génétique de la famille 54.....	74
2.4.1	Phénotypage de la famille 54.....	74
2.4.2	Existence d'individus atypiques.....	75
2.4.3	Cartographie génétique du père 101-74 .....	75
2.4.4	Cartographie génétique de la mère 73028-62 .....	78
2.4.5	Comparaison des deux cartes génétiques parentales.....	79
2.4.6	Alignement des cartes génétiques avec le génome de Nisqually-1.....	80
2.5	Cartographie physique de l'individu 101-74.....	80
2.5.1	Criblage de la banque BAC .....	81
2.5.2	Utilisation des BES en cartographie physique.....	81
2.5.3	Assemblage des contigs de clone BAC .....	83
2.5.4	Alignement des cartes physiques avec la carte génétique .....	84
2.5.5	Alignement de la carte physique $hR_{US}$ avec le génome de Nisqually-1 .....	84
2.5.6	Taux de recombinaison.....	85
2.6	Caractéristiques de la zone génomique .....	86
2.7	Vers le clonage positionnel de $R_{US}$ et $R_1$ .....	88



# 1 Introduction

Pour réaliser le clonage positionnel du locus  $R_{US}$  (principalement) et le clonage du locus  $R_1$  (secondairement), la première étape de cette thèse passait par l'étude de l'extrémité du chromosome 19 du peuplier où se trouvent ces 2 loci.

## 1.1 Vers le clonage positionnel de $R_{US}$

La séquence de référence du génome de Nisqually-1 *P. trichocarpa* contient beaucoup de trous ; de nombreux scaffolds non encore assignés à des chromosomes peuvent combler ces trous. La construction de cartes génétiques et physiques de cette région génomique permettra de vérifier l'ordre exact des séquences de l'extrémité du chromosome 19.

- La séquence du génome sera utile au développement rapide de nouveaux marqueurs qui permettent l'élaboration des cartes physique et génétique.
- Les cartes génétiques seront utiles pour l'ordonnement des marqueurs génétiques développés sur la séquence du génome, et ainsi d'en vérifier le bon ordre. Elles permettront surtout de localiser génétiquement le locus  $R_{US}$  avec plus de précision. L'alignement des marqueurs génétiques sur le génome permettra d'étudier les gènes candidats pour ce locus.
- Les cartes physiques, complémentaires de la carte génétique pour l'ordonnement des marqueurs, seront utiles aussi pour développer de nouveaux marqueurs *via* les BES, qui à leur tour permettront de combler les trous de la séquence du génome. Elles permettent surtout de déterminer le ou les clones BAC porteurs de l'allèle  $R_{US}$  qui après leur séquençage permettra l'identification de cet allèle.

## 1.2 Vers le clonage positionnel de $R_1$

L'exploitation de la synténie entre *P. trichocarpa* et *P. deltoides* aidera à la cartographie génétique du locus  $R_1$ . Les marqueurs dessinés sur les individus Nisqually-1 et 101-74 de l'espèce *P. trichocarpa*, permettront l'élaboration de la carte génétique du 73028-62 *P. deltoides*. Cette cartographie comparée permettra d'obtenir trois résultats :

- Préciser la localisation du locus  $R_1$  sur la carte génétique.
- Connaître le ou les types de gènes candidat en comparant au génome de référence.
- Connaître la relation entre les loci  $R_{US}$  et  $R_1$  des parents de la famille 54 : s'agit-il de 2 formes alléliques d'un seul gène ou de deux gènes distincts ?



## 2 Résultats

### 2.1 Séquence du génome de Nisqually-1

#### 2.1.1 D'après la version V1 du génome

La zone principale étudiée porte sur les premiers 3 millions de paires de base (pb) du chromosome 19 de la séquence du génome de Nisqually-1 V1. Elle porte aussi sur le scaffold 117 de 1 000 030 pb supplémentaires indiqué par les séquences de produits d'amplification AFLP (figure 21, page I-41). Le choix de cette zone génomique repose sur la présence de nombreux gènes NBS-LRR révélés par l'annotation automatique disponible de la version V1 du génome, Ils sont moins présents après ces 3 premiers Mb. il s'agit potentiellement de gènes impliqués dans la résistance d'une plante à un pathogène.

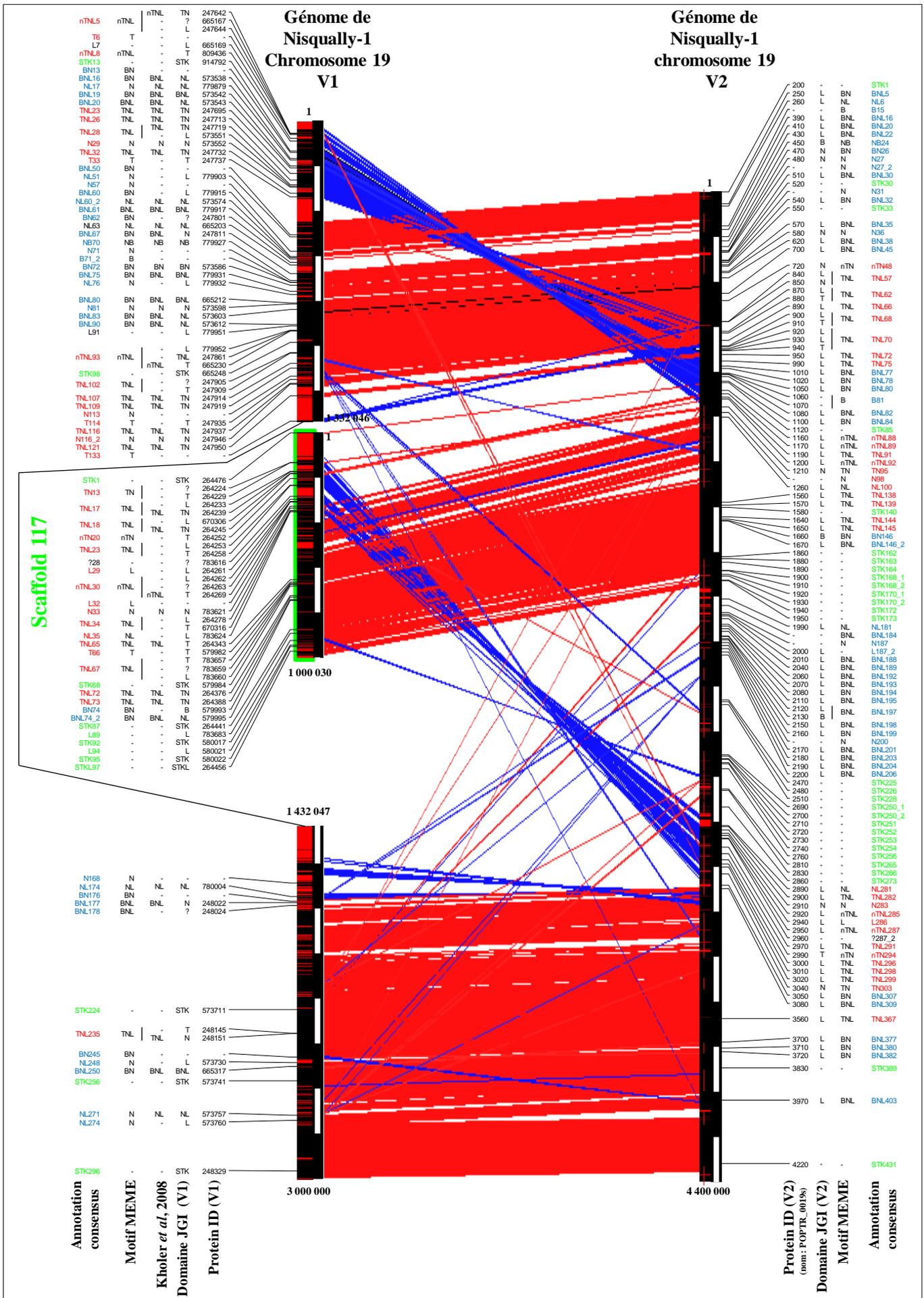
Ces séquences comportent beaucoup de trous : sur un total de 4 000 030 pb, 943 150 pb sont réparties en 308 trous (représentés dans la séquence par des suites de N), ce qui représente 23,58% de ces séquences dont cinq trous font 100 Kb. Il reste donc 3 056 880 pb de séquence travaillable.

L'annotation automatique du JGI a été examinée gène à gène à partir du « genome browser » du JGI ([http://genome.jgi-psf.org/Poptr1\\_1/Poptr1\\_1.home.html](http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html)). Sur cette zone d'intérêt, 303 gènes différents, codant des protéines, sont annotés, ce qui représente environ un gène tous les 10 Kb.

#### 2.1.2 D'après la version V2 du génome

La zone équivalente aux 3 premiers Mb de la V1 correspond aux 4 382 460 pb de la séquence du second assemblage V2 (déterminé avec le logiciel BLASTn avec les 2 derniers Kb). Dans cette seconde version, l'ancien scaffold 117 de la version V1 a été intégré au chromosome 19. Il y a 240 769 pb de trous répartis en 219 segments, ce qui représente 5,49% de ces séquences (taille moyenne d'un trou : 1100 pb). Il reste 4 141 691 pb de séquence travaillable. Il y a donc plus de 1 Mb de séquence supplémentaire travaillable dans cette seconde version.

Sur cette zone d'intérêt, 409 gènes différents sont annotés sur le "genome browser" du JGI (<http://www.phytozome.net/poplar.php>). La proportion de gène est identique à la version V1 : environ un gène tous les 10 Kb.



**Figure 29** : Alignement entre les 3 premiers Mb du chromosome 19 de la version V1 et la région équivalente de la version V2 du génome de Nisqually-1.

Les échelles physiques sont indiquées en pb. Dans la barre de gauche représentant les séquences : l'absence de séquence est indiquée en rouge, la présence de séquence est indiquée en noir. La barre de droite représente l'échelle tous les 200 kb. Les alignements en rouge sont dans le même sens et en bleu ils sont en sens inverse. Le scaffold 117 encadré en vert a été intégré au chromosome 19 de la V1. Les gènes de résistance sont indiqués sur les séquences suivant les motifs qu'ils portent pour l'annotation automatique du JGI, l'annotation de Kholer et al, (2008), et celle produite avec les motifs MEME. Le n° Protein ID est indiqué suivant le "genome browser" du JGI. Pour l'annotation consensus les gènes de la catégorie des TNLs sont indiqués en rouge, les gènes de la catégorie des BNLs sont indiqués en bleu, et les gènes STK sont indiqués en vert.

### 2.1.3 Les différences entre les deux versions de l'assemblage du génome

L'alignement entre les deux assemblages a été effectué avec le logiciel Artemis Comparaison Tool qui représente les résultats de megaBLAST (e-value inférieure à  $10^{e-100}$  sans filtre de faible complexité), seulement pour ceux qui ont un score de plus de 2 000. Les alignements présentés en rouge montrent les alignements dans le même sens et en bleu les alignements en sens inverse. Cet alignement montre deux différences importantes dans l'ordre des séquences entre les deux assemblages : (i) les 350 premiers Kb de la version V1 sont placés en sens inversé entre les 600 et 950 Kb de la version V2, (ii) les 250 Kb entre 100 et 300,5 Kb du scaffold 117 de la version V1 sont placés en sens inverse de la version V2 entre les 2 800 et 3 050 Kb (figure 29).

Une correspondance entre les annotations automatiques des deux assemblages du génome est donnée par le JGI. La correspondance est donnée lorsque le résultat de BLAT a plus de 90% d'identité et plus de 80% de la couverture du CDS :

- Quarante deux des 303 gènes de la version V1 n'ont pas de correspondance avec ceux de la version V2 (13,9%), un gène possède une correspondance sur le nouveau chromosome 12, 28 gènes possèdent des correspondances sur des nouveaux scaffolds non assignés à un chromosome (9,2%). Les 232 gènes restants ont une correspondance sur le nouveau chromosome 19, dont 5 ont deux correspondances.

- Cent six des 409 gènes de la version V2 n'ont pas de correspondance avec ceux la version V1 (25,9%). Les 303 gènes restant ont une correspondance dans l'ancien assemblage, dont 76 ont plusieurs correspondances. Ces observations sont cohérentes avec la meilleure couverture en séquence atteinte dans la version V2 du génome.

### 2.1.4 Les gènes de résistance

#### 2.1.4.1 Les gènes NBS-LRR

##### 2.1.4.1.1 D'après le génome de Nisqually-1 V1

Pour la version V1 du génome, trois différents paramètres ont été suivis : l'annotation automatique donnée sur le site du JGI, l'annotation faite par Kolher *et al*, (2008) et l'alignement des motifs MEME des gènes NBS-LRR (annexe 2) sur la séquence (figure 29 ; tableau 4) :

- Par l'annotation automatique V1 (JGI), 81 gènes ont été annotés comme RGA, appartenant à la catégorie des NBS-LRR. Huit d'entre eux n'ont pas de domaine prédit (HMMPfam et FPrintScan) mais ont leur meilleur score avec un gène NBS-LRR (comme tous

**Tableau 4** : Inventaire des gènes NBS-LRR sur l'extrémité du chromosome 19.

Le nombre de gènes est donné suivant les motifs qu'il porte et de la catégorie auxquels ils appartiennent. Cette inventaire est donné pour l'annotation automatique du JGI, l'annotation de Kolher *et al*, 2008, l'annotation produite avec les motifs MEME, et l'annotation consensus produite à partir de ces différentes annotations.

Motif Predit	JGI V1	Kolher <i>et al</i> , 2008 V1	motifs MEME V1	Consensus V1		JGI V2	motifs MEME V2	Consensus V2
Sans motif	8	-	-	1		-	-	1
NBS (N)	7	4	14	8		8	8	8
NBS-LRR (NL)	10	5	4	10		-	4	4
LRR (L)	20	-	2	4		64	1	2
catégorie des BNL								
BED-NBS-LRR (BNL)	5	12	4	15		-	24	35
BED (B)	1	-	1	1		3	2	2
BED-NBS (BN)	1	1	17	6		-	13	2
NBS-BED (NB)	1	1	1	1		-	1	1
catégorie de TNL								
TIR-NBS-LRR (TNL)	1	14	18	18		-	18	18
NLS-TIR-NBS-LRR (nTNL)	-	3	4	4		-	5	5
TIR-NBS (TN)	13	-	1	1		-	2	2
NLS-TIR-NBS (nTN)	-	-	1	1		-	2	2
TIR (T)	14	-	5	5		4	-	-
Total	81	40	72	75		79	80	82

les 79 gènes). Il y a un gène ne possédant que le domaine B, un gène BN, un gène NB, 5 gènes BNL, 20 gènes L, 7 gènes N, 10 gènes NL, 14 gènes T, 13 gènes TN, un gène TNL.

- L'annotation donnée par Kolher *et al*, (2008) prédit 40 gènes. Il y a un gène BN, un gène NB, 12 gènes BNL, 4 gènes N, 5 gènes NL, 14 gènes TNL, 3 gènes nTNL. Ces annotations ont été indexées au browser du JGI, elles sont donc identiques en terme de structure mais différentes sur la prédiction des domaines.

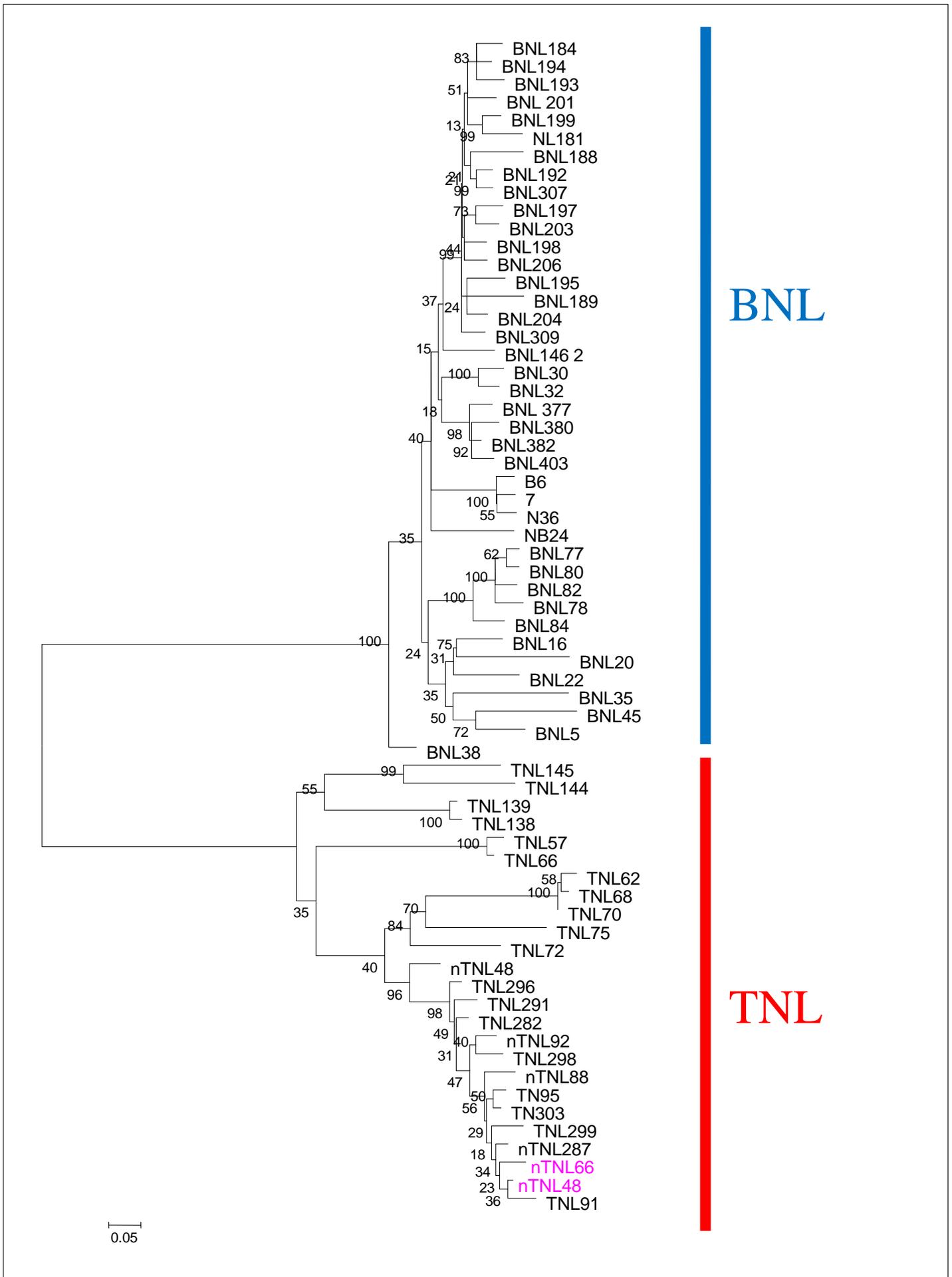
- Les alignements des motifs MEME donnés par Kolher *et al*, (2008) et Meyers *et al*, (2003) (annexe 2) ont été produits avec la séquence du chromosome 19. Les différents alignements des différents domaines ont été raboutés suivant leur orientation et leur proximité pour en déduire 72 gènes NBS-LRR putatifs (ces gènes sont qualifiés de putatifs car un simple alignement par BLAST ne permet pas de vérifier qu'il y a bien les CDS : en effet, des codons stop peuvent être présents entre les différents motifs utilisés). Il y a un gène B, 17 gènes BN, 4 gènes BNL, un gène NB, deux gènes L, 14 gènes N, 4 gènes NL, 5 gènes T, 18 gènes TNL, un gène TN, un gène nTN et 4 gènes nTNL.

- Une annotation consensus a été réalisée entre ces trois démarches (figure 29), à l'aide des alignements des trois paramètres observés pour un même gène (MEGA4). L'alignement des gènes entre eux a permis d'assigner les gènes N, NL et L à l'une des deux catégories BNL ou TNL. Douze des gènes putatifs caractérisés par les motifs MEME regroupent deux ou trois gènes prédits par l'annotation automatique du JGI, et neuf des gènes putatifs n'ont pas de gènes prédits par l'annotation automatique. Les domaines LRR n'ont pas toujours été caractérisés par les alignements des motifs MEME. Dans le consensus, ils proviennent le plus souvent des annotations automatiques du JGI et de celles produites par Kolher *et al*, (2008), mais également des alignements des gènes entre eux. Au total, 75 gènes sont annotés : il y a un gène B, 6 gènes BN, 15 gènes BNL, un gène NB, 4 gènes L, 8 gènes N, 10 gènes NL, 5 gènes T, 1 gène TN, 18 gènes TNL, un gène nTN et 4 gènes nTNL et un gène sans motif.

#### **2.1.4.1.2 D'après Nisqually-1 V2**

Pour la version V2 du génome, deux différents paramètres ont été suivis : l'annotation automatique donnée sur le site du JGI et l'alignement des motifs MEME (annexe 2) sur la séquence (figure 29 ; tableau 4) :

- L'annotation automatique du second assemblage du génome (V2) ne donne qu'un seul domaine Pfam par gène, il est donc impossible de connaître tous les domaines que portent les gènes NBS-LRR. Il y 79 gènes qui possèdent donc un domaine Pfam des gènes NBS-LRR : 4



**Figure 30** : Phylogénie des gènes NBS-LRR présents sur le chromosome 19. Les gènes provenant du clone BAC *P. deltoides* proche du locus MER sont indiqués en rose.

pour un domaine TIR, 3 pour un domaine BED, 8 pour un domaine NBS et 64 pour un domaine LRR.

- De même que pour la version V1, les motifs MEME donnés par Kolher *et al*, (2008) et Meyers *et al*, (2003) ont été alignés sur la région d'intérêt, et ont montré la présence de 82 gènes putatifs. Il y a deux gènes B, 13 gènes BN, 24 gènes BNL, un gène NB, un gène L, 9 gènes N, 4 gènes NL, deux gènes TN, 18 gènes TNL, deux gènes nTN et 5 gènes nTNL et un gène n (NLS).

- Une annotation consensus a été produite entre les deux paramètres observés, avec l'aide des alignements des gènes entre eux : 82 gènes NBS-LRR ont été déterminés : deux gènes B, 2 gènes BN, 35 gènes BNL, un gène NB, deux gènes L, 8 gènes N, 4 gènes NL, deux gènes TN, 18 gènes TNL, deux gènes nTN, 5 gènes nTNL, et un gène sans motif. Il y a aussi deux gènes annoté comme RGA dans l'annotation du JGI qui n'ont pas de motif.

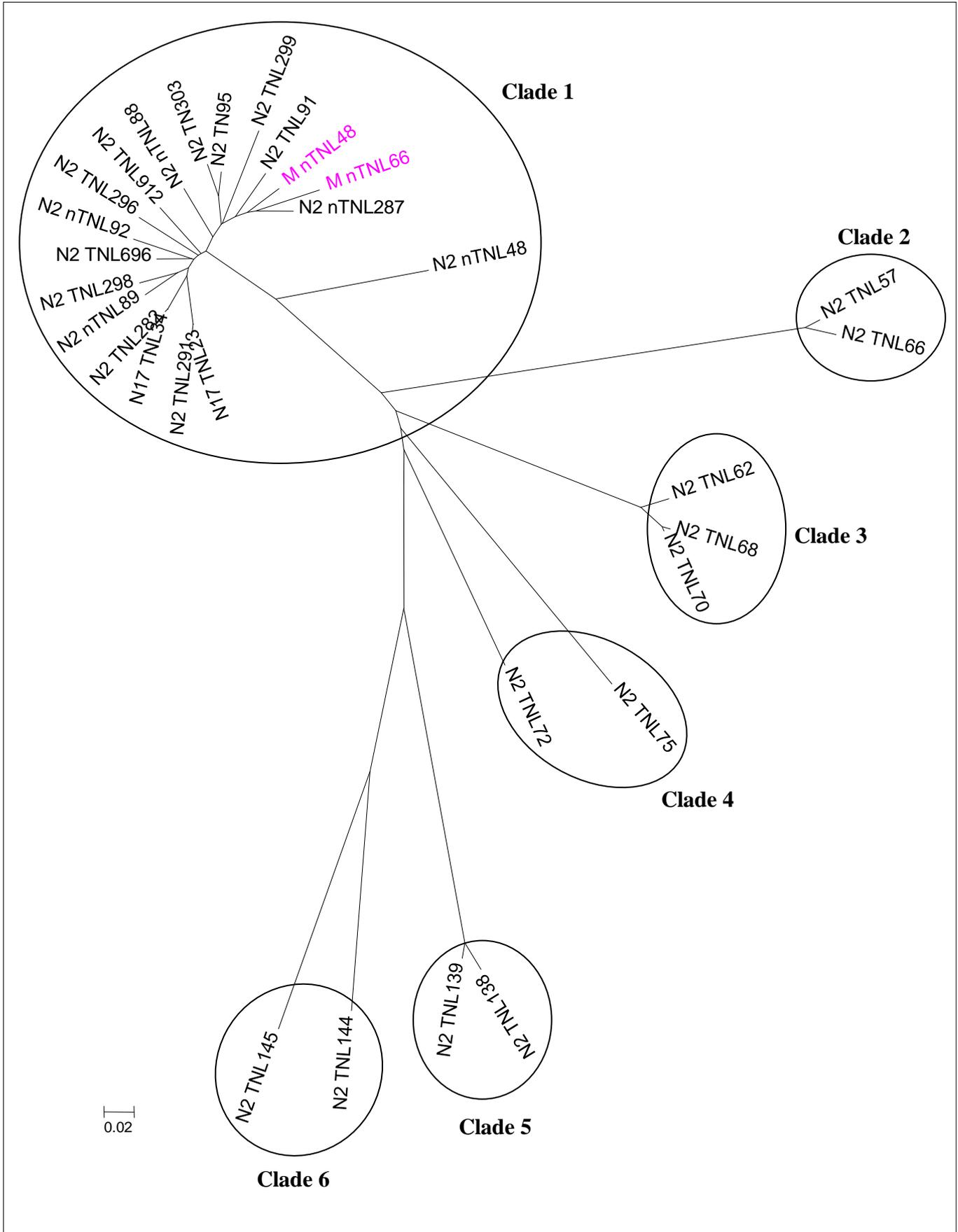
Dans la seconde version de l'assemblage, il y a plus de gènes NBS-LRR que dans la version V1, ce qui est logique puisque il y a plus de séquence et plus de gènes.

#### **2.1.4.2 Phylogénie des gènes NBS-LRR**

L'alignement des domaines NBS des gènes NBS-LRR a été produit avec le logiciel MEGA4 pour établir la phylogénie de ces gènes (figure 30). Ce domaine est présent chez les deux catégories de gène (BNL ou TNL), il permet donc de faire une phylogénie complète quelle que soit la catégorie des gènes. Le domaine LRR est présent chez ces deux catégories, mais il est difficile de les aligner. Comme un bon alignement était nécessaire pour une bonne phylogénie, le domaine LRR n'a pas été utilisé dans cette phylogénie. La phylogénie sépare clairement les deux types de gènes, les TNLs et les BNLs. Il est observé une plus grande diversité dans les gènes TNLs que dans les gènes BNLs. Il n'est pas possible de définir correctement des clades pour les BNL (figure 30).

#### **2.1.4.3 Définition de clade pour les gènes TNL**

La phylogénie des gènes TNLs (figure 31) présents sur cette extrémité du chromosome 19 a permis de définir 6 clades pour les TNLs. Des alignements avec les gènes des différents clades de TNL ont été produits sur le génome entier (V1) dans le but d'identifier des gènes de ces clades présents sur d'autres chromosomes ou sur des scaffolds non assignés à ces chromosomes. Comme il y a une forte redondance sur ces gènes TNLs, les gènes identifiés ont été alignés aux gènes présents sur la zone d'intérêt pour vérifier qu'ils provenaient bien des mêmes clades. Seuls les gènes du clade 1 sont présents sur d'autres chromosomes (chromosome 1 et 7) et de nombreux scaffolds de la version V1 (scaffolds 2701, 272, 283, 2139, 4600 et 1268),



**Figure 31** : Phylogénie des gènes TNL de l'extrémité du chromosome 19 de Nisqually-1 V2. Les gènes provenant du clone BAC *P. deltoides* proche du locus MER sont indiqués en rose.

ainsi que sur le clone BAC *P. deltoides* proche du locus *MER* (Lescot *et al*, 2004). Les autres clades n'ont que des gènes présents sur l'extrémité du chromosome 19.

#### **2.1.4.4 Les gènes STK**

Pour les gènes STK, seules les annotations automatiques du génome (V1 et V2) ont été regardées : 13 STKs sur la version V1 et 30 STKs sur la version V2 répartis en différents clusters (figure 29, page III-67).

## **2.2 Développement de marqueurs**

### **2.2.1 Sur la séquence du génome de Nisqually-1**

Cent trente-six couples d'amorces ont été dessinés sur la séquence du génome de Nisqually-1 (134 sur la version V1 de l'assemblage du génome et deux sur la version V2). Il s'agit d'une recherche positionnelle sur la séquence du chromosome 19.

Soixante et onze couples sont dessinés sur un locus unique du chromosome 19, et 44 sur un locus unique du scaffold 117. Dix couples ont été dessinés sur deux loci différents à cause de la redondance très élevée entre différentes parties de l'extrémité du chromosome 19.

Les huit derniers couples sont dessinés sur des scaffolds non assignés aux chromosomes du premier assemblage du génome V1 ; nous n'avons donc aucune donnée expérimentale permettant de les intégrer à ce chromosome. Ces scaffolds ont été sélectionnés comme portant des gènes TNL qui appartiennent au clade 1 présent sur l'extrémité du chromosome 19 étudiée (Scaffolds 2701, 272, 283, 2139, 4600 et 1268).

#### **2.2.1.1 Marqueurs microsatellites**

Parmi ces 136 couples d'amorces, 52 sont dessinés autour de répétitions de faible complexité (SSR).

#### **2.2.1.2 STS**

##### **2.2.1.2.1 STS génique**

Quatre-vingt couples d'amorces ont été dessinés sur la séquence du génome, et 68 couples d'amorces dessinés sur des gènes ou à cheval sur des gènes. Quarante six d'entre eux ont été dessinés sur des gènes NBS-LRR. Six sont à cheval sur l'extrémité 3' de gènes TNL avec une amorce consensus dans le domaine LRR, 16 sont dessinés sur l'extrémité 5' du gène avec une amorce consensus dans le domaine TIR.



Six couples d'amorces sont dessinés sur des gènes NBS-LRR provenant des six scaffolds non assignés.

#### **2.2.1.2.2 STS non génique**

Douze couples d'amorces ont été dessinés sur des zones inter-géniques du chromosome 19 et deux couples sur le scaffold 282.

### **2.2.2 Sur les BESs**

Au fur et à mesure de l'avancée des travaux, 42 couples d'amorces supplémentaires ont été dessinés sur les BESs (dont 2 fois 2 couples sur le même BES) issues de 34 clones BAC (dont 9 fois 2 couples ont été dessinées sur les 2 BESs du même clone BAC) (annexes 3 et 4). Les éléments transposables ont pu être évités dans une bonne proportion, mais dans plusieurs cas, aucune alternative n'a été possible.

### **2.2.3 Clone BAC *P. deltoides* proche du locus *MER***

Trois couples d'amorces ont été dessinés sur la séquence du clone BAC *P. deltoides* proche du locus *MER* (Lescot *et al*, 2004). Ils ont été dessinés sur les 3 gènes TNL de ce clone BAC. Ces couples ont été dessinés en utilisant l'amorce consensus développée dans le domaine TIR du clade 1.

## **2.3 Test des couples d'amorces**

Deux cent cinq couples d'amorces ont été testés pour travailler sur les 3 premiers millions de pb de Nisqually-1 V1 (y compris le scaffold 117).

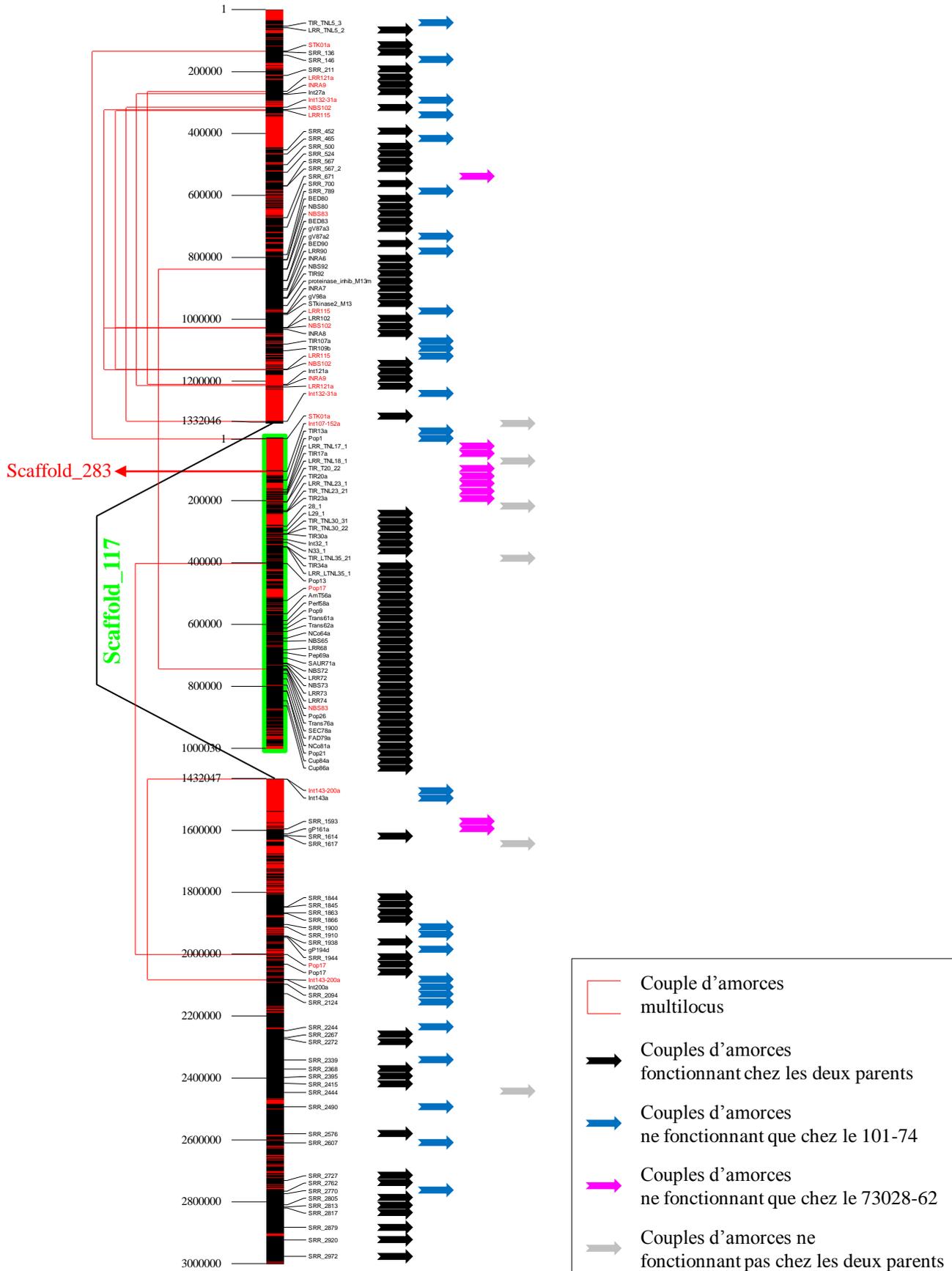
En plus des 183 couples d'amorces nouvellement développés, neuf couples provenaient de SCAR issus de la BSA-AFLP et un couple SCAR issu de RAPD (Masle, 2005 ; données personnelles Véronique Jorge, Villar *et al*, 1996). Douze autres couples sont issus de la bibliographie : 11 SSR (issus de BES de l'individu Nisqually-1 ; Tuskan *et al*, 2004 ; [http://www.ornl.gov/sci/ipgc/ssr\\_resource.htm](http://www.ornl.gov/sci/ipgc/ssr_resource.htm)) et un STS (R7 développé par Yin *et al*, (2004b) sur le clone BAC proche du locus *MER*).

### **2.3.1 Test des parents de la famille 54**

#### **2.3.1.1 Matrice Nisqually-1**

Parmi les 136 couples d'amorces dessinés sur Nisqually-1, 21 ne donnent pas de produit d'amplification chez le 101-74, il en est de même pour 47 couples chez le 73028-62. Dix couples

## Génome de Nisqually-1 V1 Extrémité du chromosome 19



**Figure 32 :** Couples d'amorces développés sur le génome sur Nisqually-1 V1.

Dans la barre représentant la séquence du génome, l'absence de séquence est indiquée en rouge, la présence de séquence est indiquée en noir. Les flèches représentent le résultats des test des couples d'amorces, en noir quand ils fonctionnent chez les deux parents de la famille 54, en bleu quand ils fonctionnent que chez le 101-74, en rose quand ils fonctionnent que chez le 73028-62 et en gris quand ils ne fonctionnent pas chez les deux parents de la famille 54. Les couples d'amorces en rouge sont dessinés sur plusieurs loci.

ne marchent pas chez les deux parents de la famille 54. Pour ces derniers couples, ils ont montré une amplification sur ADN<sub>g</sub> de Nisqually-1, ce qui a permis de vérifier le bon développement des couples d'amorces et donc la cohérence de l'approche (figure 32).

### 2.3.1.2 Matrice BES

Parmi les 45 couples dessinés sur les BES, trois couples ne marchent pas chez le 101-74 : il s'agit sûrement de séquences de mauvaise qualité ; les amorces ont donc été dessinées sur des séquences qui ne sont pas présentes chez le 101-74. Ces 3 couples d'amorces ne fonctionnent pas non plus chez le 73028-62. Sur les 42 couples d'amorces qui marchent chez le 101-74, 18 ne fonctionnent pas chez le 73028-62.

### 2.3.1.3 Matrice du clone BAC proche du locus *MER*

Quatre couples d'amorces ont été développés sur la séquence du clone BAC *P. deltoides* proche du locus *MER*. Des 3 couples sur les gènes NBS-LRR, un ne fonctionnait que sur le père 101-74, un fonctionnait chez les deux parents, et le dernier ne fonctionnait pas chez les deux parents de la famille 54. Le marqueur STS, développé par Yin *et al.*, (2004b), ne fonctionnait que chez la mère 73028-62.

## 2.3.2 Test sur un échantillon de la famille 54

Chaque marqueur qui montrait un polymorphisme chez l'un des parents, ou qui ne marchait pas chez l'un des parents, a été testé sur 6 individus de la famille 54. Après avoir testé plusieurs centaines de marqueurs, j'ai pu vérifier que parmi ces 6 individus, il y avait les quatre possibilités pour l'ensemble de la zone testée :  $[R_{US} R_1]$ ,  $[r_{US} R_1]$ ,  $[R_{US} r_1]$ , et  $[r_{US} r_1]$ .

Sur 216 marqueurs testés, 78 montraient naturellement une ségrégation. La figure 33 montre les différents cas où une ségrégation a pu être observée dans la famille 54 :

- Les marqueurs dominants pour l'un des parents (figure 33, cas 1) ségrège 1:1 pour l'un des parents ; il marque spécifiquement sur l'un des quatre allèles présents dans la descendance.

- Les marqueurs dominants pour l'un des parents et co-dominants pour l'autre parent (figure 33, cas 2). La ségrégation observée est de 1:2:1. Ces marqueurs permettent de faire un pont entre les cartes génétiques des deux parents, par contre pour le parent où c'est un marqueur dominant, il n'est possible de déterminer que la moitié des génotypes pour ces marqueurs.

- Les marqueurs dont la ségrégation est révélée par une digestion d'une endonuclease (figure 33, cas 3). Même si ceux-ci fonctionnent sur les deux allèles d'un individu, la lecture des génotypes se fait juste par présence ou absence de la digestion. En effet l'allèle non digéré est

<p>Cas 1 : Marqueur dominant</p>	<p>♀    ♂</p> <p>a</p> <p><u>nul</u>    <u>a</u></p> <p><u>nul</u>    <u>nul</u></p>	<p>1/2    1/2</p> <p><u>nul</u>    <u>a</u></p> <p><u>nul</u>    <u>nul</u></p>	<p>Lecture de la ségrégation ♂ : présence/absence de a</p>
<p>Cas 2 : Double hétérozygote Marqueur dominant + marqueur co-dominant</p>	<p>♀    ♂</p> <p>a</p> <p>b</p> <p><u>b</u>    <u>a</u></p> <p><u>nul</u>    <u>b</u></p>	<p>1/2    1/4    1/4</p> <p><u>b</u>    <u>a</u>    <u>a</u></p> <p><u>nul</u>    <u>b</u>    <u>nul</u></p> <p><u>b</u></p> <p><u>b</u></p>	<p>lecture de la ségrégation ♀ : 1 génotype /2 présence/absence b si a est présent</p> <p>lecture de la ségrégation ♂ : présence/absence a</p>
<p>Cas 3 : Marqueur CAPS et dCAPS</p>	<p>♀    ♂</p> <p>a</p> <p>Digestion</p> <p><u>b</u>    <u>a</u></p> <p><u>b</u>    <u>b</u></p>	<p>1/2    1/2</p> <p><u>b</u>    <u>a</u></p> <p><u>b</u>    <u>b</u></p>	<p>Lecture de la ségrégation ♂ : présence digestion de a / absence digestion de a</p>
<p>Cas 4 : Marqueur co-dominant</p>	<p>♀    ♂</p> <p>a</p> <p>b</p> <p><u>nul</u>    <u>a</u></p> <p><u>nul</u>    <u>b</u></p>	<p>1/2    1/2</p> <p><u>b</u>    <u>a</u></p> <p><u>nul</u>    <u>nul</u></p>	<p>lecture de la ségrégation ♂ : présence a/absence b</p>
<p>Cas 5 : Marqueur double dominant différents</p>	<p>♀    ♂</p> <p>a</p> <p>b</p> <p><u>b</u>    <u>a</u></p> <p><u>nul</u>    <u>nul</u></p>	<p>1/4    1/4    1/4    1/4</p> <p><u>b</u>    <u>a</u>    <u>a</u>    <u>nul</u></p> <p><u>nul</u>    <u>b</u>    <u>nul</u>    <u>nul</u></p>	<p>lecture de la ségrégation ♀ : présence/absence b</p> <p>lecture de la ségrégation ♂ : présence/absence a</p>
<p>Cas 6 : Marqueur double dominant mais avec compétition entre locus</p>	<p>♀    ♂</p> <p>a</p> <p>b</p> <p><u>b</u>    <u>a</u></p> <p><u>nul</u>    <u>nul</u></p>	<p>1/4    1/2    1/4</p> <p><u>b</u>    <u>a</u>    <u>nul</u></p> <p><u>nul</u>    <u>b</u>    <u>nul</u></p> <p>ou</p> <p><u>a</u></p> <p><u>nul</u></p>	<p>lecture de la ségrégation ♀ : 1 génotype /2 présence/absence b si a est absent</p> <p>lecture de la ségrégation ♂ : présence/absence a</p>
<p>Cas 7 : Marqueur double dominant identique</p>	<p>♀    ♂</p> <p>a</p> <p><u>a</u>    <u>a</u></p> <p><u>nul</u>    <u>nul</u></p>	<p>1/4    3/4</p> <p><u>nul</u>    <u>a</u></p> <p><u>nul</u>    <u>a</u></p> <p><u>a</u></p> <p><u>nul</u></p>	<p>lecture de la ségrégation ♀ : présence/absence b</p> <p>lecture de la ségrégation ♂ : présence/absence a</p>

**Figure 33 :** Génotype du 101-74 et du 73028-62 et des descendants de la famille 54 pour les différents cas de marqueurs.

généralement masqué par l'autre parent ; la digestion ne marque pas spécifiquement le second allèle du parent cartographié.

- Les marqueurs co-dominants pour l'un des parents (figure 33, cas 4) ségrègent 1:1 pour l'un des parents, ils marquent spécifiquement les deux allèles d'un parent parmi les 4 présents dans la descendance.

- Les marqueurs double-dominants (figure 33, cas 5) ségrègent 1:1:1:1, il s'agit d'un pont entre les cartes génétiques des deux parents.

- Les marqueurs doubles dominants dont la taille d'amplification est identique aux deux parents (figure 33, cas 7) ségrègent 1:3.

Bien que la plupart des SSR ait été développés chez *P. trichocarpa*, il y a quelques SSR qui ne marchent pas, et la plupart ne fonctionne pas chez les deux allèles du père *P. trichocarpa*. La même observation a été faite chez la mère *P. deltoïdes* mais avec un rendement bien inférieur (figure 32, page III-72).

#### **2.3.2.1 Compétition entre locus**

Il y a avec certains marqueurs génétiques, des phénomènes de compétition entre locus (figure 33, cas 6). En effet, dans le cas où il y a des amplifications de différentes tailles provenant de différents haplotypes, certaines amplifications ne se font pas quand il y a un mélange de ces différents haplotypes. Dans ce cas, la ségrégation pour l'un des parents n'est pas visible pour l'ensemble des individus, mais seulement pour la moitié des génotypes.

#### **2.3.3 Développement de marqueurs sur des produits d'amplification PCR**

Quinze marqueurs ne montrant pas de ségrégation visible sur agarose, ont été séquencés pour déterminer des SNP et des petites insertions/délétion. Les polymorphismes ainsi découverts ont été utilisés pour développer de nouveaux marqueurs génétiques. Pour déterminer au mieux le polymorphisme présent entre les deux allèles de chaque locus, les PCR ont été effectuées sur des clones BAC appartenant aux deux haplotypes de l'individu 101-74 : les 2 allèles des marqueurs ont pu être déterminés. Pour déterminer les haplotypes de la mère 73026-62, les PCR ont été réalisées directement sur l'ADNg cet individu, donc des 2 haplotypes en même temps, mais aussi sur un des deux individus « aneuploïdes » qui ne présente qu'un seul haplotype pour la zone du chromosome 19 étudié (Cf. § III-2.4.2).

Sur onze couples d'amorces qui ont été dessinés sur la séquence de produits PCR issus de l'individu 101-74, dix d'entre eux étaient dessinés par rapport à des SNPs. Ces nouveaux couples



d'amorces ont été dessinés pour produire des marqueurs dCAPS ou PASA ou PAMSA. Ces 11 couples d'amorces ont pu être utilisés en carte génétique.

## 2.4 Cartographie génétique de la famille 54

### 2.4.1 Phénotypage de la famille 54

#### 2.4.1.1 Les 336 premiers individus

Les premiers individus ont été phénotypés lors de précédentes études avec 7 souches différentes, dont 93ID6 et 98AG69 (Jorge *et al*, 2005), qui sont les deux souches utilisées pour la définition des loci  $R_1$  et  $R_{US}$  respectivement.

- L'analyse du phénotype de la taille des sores a été faite sur une échelle de 1 à 5, il y a 179 individus déterminés  $R_{US}$ , 149 individus déterminés  $r_{US}$  et 8 individus qui ne possèdent pas de phénotype.

- Pour le locus  $R_1$ , il y a 190 individus déterminés  $R_1$  et 145 individus déterminés  $r_1$  et un individu qui ne possède pas de phénotype.

#### 2.4.1.2 La première extension

Les 1079 individus de la première extension de la famille 54 en 2003, ont tous été phénotypés pour les deux souches 93ID6 et 98AG69, expériences menées en 2004 et 2005 à l'INRA d'Orléans par Arnaud Dowkiw et Catherine Bastien.

- L'analyse du phénotype de la taille des sores a été faite avec une analyse informatique photographique plus précise, mais aussi à partir d'une lecture manuelle de la taille des sores sur une échelle de 1 à 5. Il y a 512 individus déterminés  $R_{US}$  et 515 individus déterminés  $r_{US}$ , 22 individus n'ont pas pu être déterminés (taille des sores intermédiaire) et 30 individus n'ont pas été phénotypés.

- Pour le locus  $R_1$ , il y a 513 individus déterminés  $R_1$ , 501 individus déterminés  $r_1$  et 65 individus qui n'ont pas été phénotypés.

#### 2.4.1.3 La seconde extension

Les 729 individus de la seconde extension de la famille 54 en 2007, ont été phénotypés avec les deux mêmes souches de *Mlp* : seuls les individus possédant une recombinaison à proximité des loci  $R_1$  et  $R_{US}$  et les individus de génotypes  $[r_1]$  pour l'haplotype provenant de la mère 73028-62 ont été utilisés. En effet, le phénotypage de ces individus a été fait après le génotypage (Cf. § III-2.4.3.3 et § III-2.4.4.3). Le but est d'étudier particulièrement ces individus

**Tableau 5** : Les individus atypiques de la famille 54.

? indique une phénotype non déterminé ou intermédiaire. En rose sont indiqués le génotype ou le phénotype hérité de la mère 73028-62, en bleu sont indiqués le génotype ou le phénotype hérité du père 101-74.

Individu de la famille 54	génotype	nombre d'extraction vérifié	phénotype	série
661300356	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	R <sub>I</sub> , R <sub>US</sub>	série de base
661300419	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	3	R <sub>I</sub> , R <sub>US</sub>	série de base
661300552	[r <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	r <sub>I</sub> , R <sub>US</sub>	série de base
661300362	[R <sub>I</sub> , r <sub>I</sub> , r <sub>US</sub> ]	3	R <sub>I</sub> , r <sub>US</sub>	série de base
661302477	[R <sub>I</sub> , r <sub>I</sub> , r <sub>US</sub> ]	2	R <sub>I</sub> , ?	première extension
661302592	[r <sub>I</sub> , -]	2	?, r <sub>US</sub>	première extension
661302832	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	R <sub>I</sub> , R <sub>US</sub>	première extension
661303031	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	R <sub>I</sub> , R <sub>US</sub>	première extension
661302916	[r <sub>I</sub> , -]	2	?, r <sub>US</sub>	première extension
661303093	[r <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	r <sub>I</sub> , R <sub>US</sub>	première extension
661303167	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	R <sub>I</sub> , R <sub>US</sub>	première extension
661303221	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	R <sub>I</sub> , R <sub>US</sub>	première extension
661303254	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	?, R <sub>US</sub>	première extension
661303275	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	r <sub>I</sub> , R <sub>US</sub>	première extension
661303282	[r <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	?, ?	première extension
661303294	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	R <sub>I</sub> , R <sub>US</sub>	première extension
661303412	[r <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	r <sub>I</sub> , R <sub>US</sub>	première extension
661303698	[r <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	r <sub>I</sub> , R <sub>US</sub>	seconde extension
661303743	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	R <sub>I</sub> , R <sub>US</sub>	seconde extension
661303782	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	2	R <sub>I</sub> , R <sub>US</sub>	seconde extension
661303800	[r <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	3	r <sub>I</sub> , R <sub>US</sub>	seconde extension
661303818	[R <sub>I</sub> , R <sub>US</sub> , r <sub>US</sub> ]	1	?, ?	seconde extension
661303719	[R <sub>I</sub> , r <sub>I</sub> , r <sub>US</sub> ]	2	R <sub>I</sub> , r <sub>US</sub>	seconde extension
661303783	[R <sub>I</sub> , r <sub>I</sub> , r <sub>US</sub> ]	2	R <sub>I</sub> , r <sub>US</sub>	seconde extension

en QTL vis-à-vis des deux souches pour les paramètres quantitatifs ; en effet, quand les individus sont  $r_1$ , il n'y a pas de résistance totale qui masque la mesure des paramètres quantitatifs pour la souche 93ID6 de *Mlp*.

- Il y a 154 individus déterminés  $R_{US}$  et 164 individus déterminés  $r_{US}$ .
- Pour le locus  $R_1$ , il y a 50 individus déterminés  $R_1$  et 267 individus déterminés  $r_1$ .

## 2.4.2 Existence d'individus atypiques

L'analyse spécifique du génotype de chaque individu de la famille 54 a permis d'écartier certains individus particuliers qui ne pouvaient pas être analysés en cartographie génétique (tableau 5). Ces individus se rangent en 2 catégories :

1°) les individus qui portent les deux haplotypes d'un des parents et un haplotype pour l'autre parent. Il existe 18 individus qui possèdent les deux haplotypes du père 101-74 pour le chromosome 19, 16 de ces individus qui ont été phénotypés sont  $R_{US}$ . Ces 18 individus ont été exclus de l'analyse de cartographie génétique pour le père 101-74 ; en effet, nous ne pouvons pas observer les éventuelles recombinaisons entre les deux haplotypes, par contre ils ont été pris en compte pour l'établissement de la carte génétique de la mère 73028-62. Il existe aussi quatre individus qui possèdent deux haplotypes de la mère 73028-62, et sont de phénotype  $R_1$ . Au total 22 individus des 2144 sont donc trisomiques ou triploïdes (soit plus de 1 %). Comme ces individus n'ont pas été étudiés sur le reste du génome, nous ne pouvons pas savoir s'il s'agit de trisomiques ou de triploïdes.

2°) deux individus n'ont pas hérité d'un haplotype du père 101-74, mais ils ont bien hérité d'un haplotype de la mère 73028-62. Ces individus ont été pris en compte pour la cartographie génétique de la mère 73028-62, mais pas pour celle du père 101-74.

Vingt trois des 24 individus atypiques ont été vérifiés sur plusieurs extractions, car il pouvait s'agir de contamination des ADNg.

## 2.4.3 Cartographie génétique du père 101-74

### 2.4.3.1 Carte génétique locale utilisant les 336 premiers individus

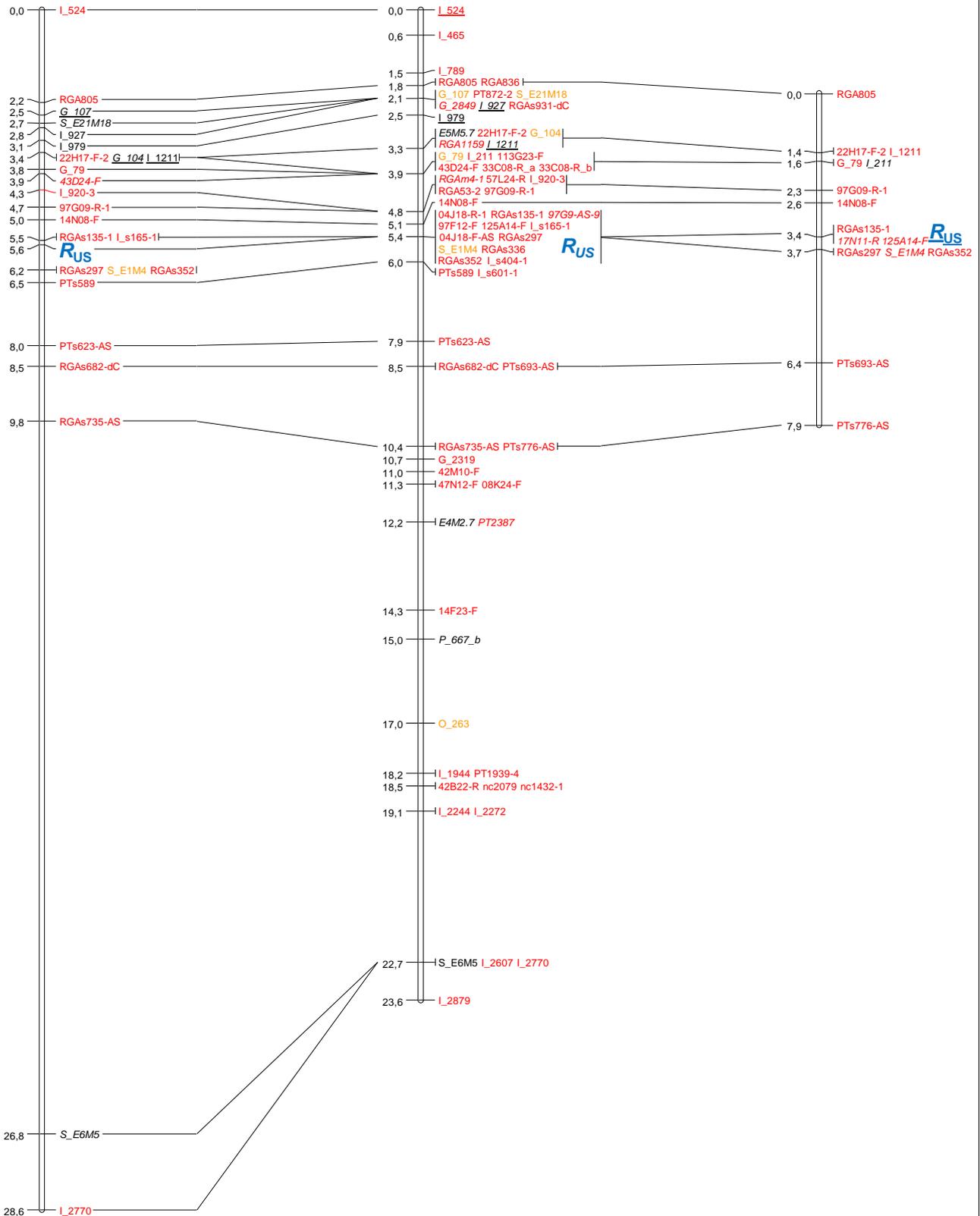
La première carte génétique a été établie à partir de l'analyse des 336 premiers individus. Dans cette série, il y a 3 individus atypiques qui ne peuvent être analysés dans la carte génétique (tableau 5) ; l'analyse a donc porté sur 333 individus (figure 34 B).

Au total, 68 couples d'amorces ont été utilisés pour génotyper les 336 premiers individus, qui ont permis d'obtenir 69 marqueurs génétiques différents (annexe 3), dont deux marqueurs

**A : Carte génétique locale**  
**101-74 *P. trichocarpa***  
**LG XIX**  
**1<sup>ère</sup> extension**  
**1067 individus**

**B : Carte génétique locale**  
**101-74 *P. trichocarpa***  
**LG XIX**  
**333 premiers individus**

**C : Carte génétique locale**  
**101-74 *P. trichocarpa***  
**LG XIX**  
**2<sup>ème</sup> extension**  
**724 individus**



**Figure 34** : Comparaison entre les cartes génétiques du père 101-74 *P. trichocarpa*.

A : Carte génétique élaborée avec les individus de la première extension de la famille 54. B : Carte génétique élaborée avec les 336 premiers individus. C : Carte génétique élaborée avec les individus de la seconde extension.

Les distances génétiques indiquées sont en cM. Les marqueurs possédants entre 1,5 et 3,5 % d'erreurs sont indiqués en italique. Les marqueurs possédants entre 40 et 80% de données présentes sont indiqués en souligné. Les marqueurs qui ont été utilisés à l'URGV sont indiqués en rouge. Les marqueurs qui ont été utilisés à l'INRA d'Orléans et à l'URGV sont indiqués en orange.

AFLP (E5M5.7 et E4M2.7) déjà inclus dans l'étude de Jorge *et al.*, (2005) et 11 marqueurs dans l'étude de Masle, (2005) dont 7 ont été réutilisés à l'URGV pour fiabiliser ces résultats. Un marqueur a été éliminé car il possède plus de 3,5% d'erreurs (57L24-F) ; la carte génétique regroupe et ordonne donc 68 marqueurs génétiques (sans compter le locus  $R_{US}$ ).

Dix marqueurs génétiques ont été développés par rapport aux SNPs caractérisés par le séquençage de produit d'amplification PCR : 2 CAPS (RGA836, PTs589), 2 dCAPS (RGA931-dC, RGAs682-dC), 3 AS-PCR (97G9-AS-9, PTs623-AS, RGAs735-AS), 3 PAMSA (04J18-F-AS, PTs693-AS, PTs776-AS).

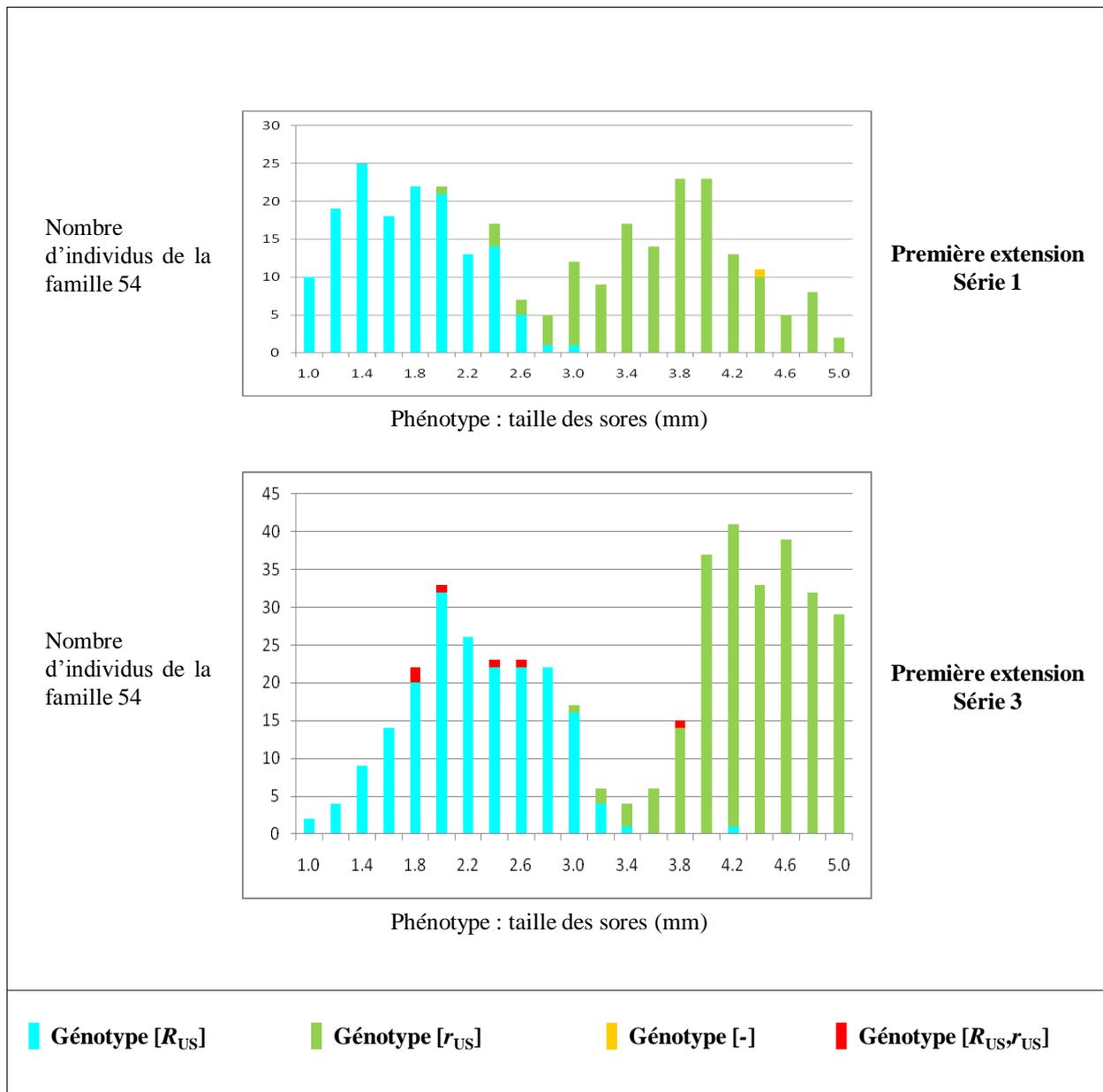
Parmi ces 68 marqueurs génétiques, il y avait 11 marqueurs qui possédaient entre 1,5 et 3,5 % d'erreurs (en italique dans la figure 34), et 4 marqueurs qui avaient entre 40 et 80 % de données présentes (en souligné dans la figure 34). Dans la globalité, les données qui ont servi à construire cette carte génétique portent 94,6 % de données présentes avec un taux d'erreur moyen de 0,57 %.

Le locus  $R_{US}$  a été introduit comme un marqueur génétique avec 97,6% de données présentes, il co-localise avec 12 marqueurs génétiques, et est encadré par le marqueur 14N08-F et le groupe de marqueurs PTs589 et INRA\_s601-1 qui sont distants de 0,9 cM. La ségrégation au niveau des marqueurs qui co-localisent avec le locus  $R_{US}$  est de 52,25 % pour l'allèle [ $R_{US}$ ] (sans compter les individus atypiques).

#### **2.4.3.2 Carte génétique locale utilisant la première extension**

Certains marqueurs génétiques ont été sélectionnés pour génotyper les 1079 individus de la première extension de la famille 54. Ces marqueurs ont été sélectionnés par rapport à leur qualité et leur proximité avec le locus  $R_{US}$ . Sur les 1079 individus, 12 individus atypiques ont été éliminés de l'analyse (tableau 5), la carte génétique a donc été faite avec 1067 individus (figure 34 A). Sur les 25 marqueurs PCR qui ont été passés sur cette série, un marqueur a été éliminé car il possédait plus de 3,5 % d'erreurs (RGA1159). La carte regroupe et ordonne 24 marqueurs génétiques, parmi lesquels il y a 5 marqueurs qui possèdent entre 1,5 et 3,5 % d'erreurs, et 3 marqueurs qui ont entre 40 et 80 % de données présentes. Dans la globalité, les données qui ont servi à construire cette carte génétique portent 91,9 % de données présentes avec un taux d'erreur moyen de 0,85 %.

Le locus  $R_{US}$  a été introduit comme un marqueur génétique avec 95,2 % de données présentes. Sur les 12 marqueurs qui co-localisent avec le locus  $R_{US}$  dans l'analyse des 336 premiers individus, 5 marqueurs ont été passés sur cette seconde série, et encadrent le locus dans un intervalle de 0,7 cM par un groupe de 2 marqueurs (RGAs135-1 et INRA\_s165-1) et un



**Figure 35** : Distribution de la famille 54 lors de deux expériences de phénotypage pour la taille des sores.

Ces deux exemples correspondent aux expériences de deux séries de la première extension de la famille 54 avec la souche de *Mlp* 98AG69. Les couleurs introduites dans l'histogramme correspondent aux génotypes des individus de la famille.

groupe de 3 marqueurs (RGAs297, SCAR\_E1M4 et RGAs352). Cette carte génétique permet donc de préciser la localisation du locus  $R_{US}$ . La ségrégation au niveau du locus  $R_{US}$  est de 48,59 % pour l'allèle [ $R_{US}$ ].

La figure 35 montre un exemple des courbes bimodales obtenues avec les données de deux expériences de phénotypage. Le génotype au locus  $R_{US}$  est comparé aux résultats de la taille des sores pour les descendants de la famille 54. De rares individus montrent un génotype [ $R_{US}$ ] et un phénotype gros sore  $r_{US}$  et vice versa. De nombreuses vérifications ont permis d'en corriger certaines (certainement des erreurs humaines) et en confirmer d'autres. Les vérifications ont aussi porté sur des individus de phénotype intermédiaire qui se sont confirmés.

#### **2.4.3.3 Carte génétique locale utilisant la seconde extension**

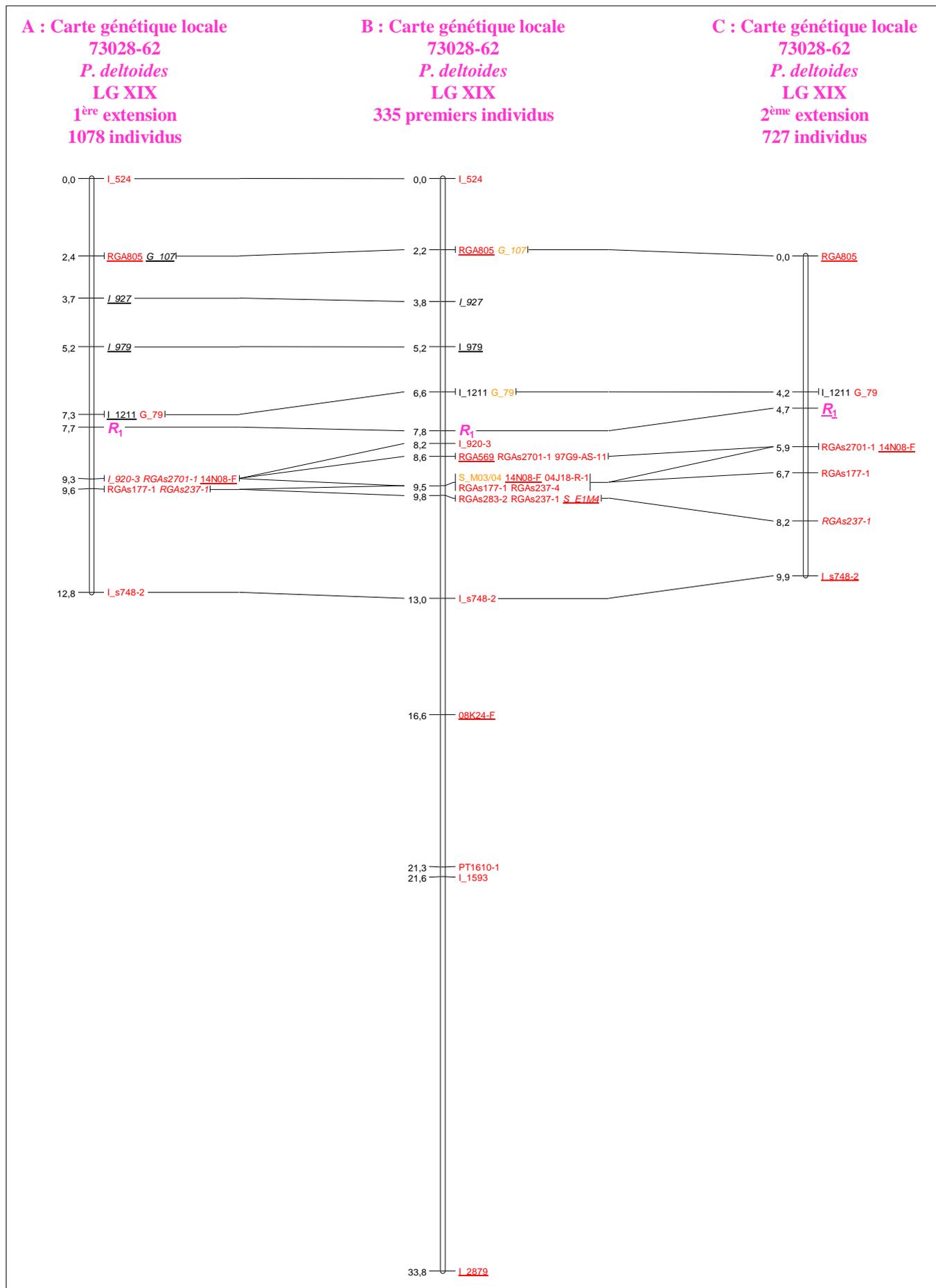
Une dernière carte génétique a été développée à partir de la seconde extension. Sur les 729 individus, 5 individus atypiques ont été éliminés de l'analyse (tableau 5), la carte génétique est donc faite avec 724 individus (figure 34 C).

Une sélection de 15 marqueurs génétiques est utilisée pour déterminer le génotype des individus supplémentaires. Cette sélection n'est pas exactement la même que pour la première extension : l'avancée du travail a permis de sélectionner des marqueurs de meilleure qualité, et la sélection a été recentrée autour des deux loci  $R_1$  et  $R_{US}$ . Parmi les 15 marqueurs, aucun n'a plus de 3,5 % d'erreurs, il y a 4 marqueurs qui possèdent entre 1,5 et 3,5 % d'erreurs, et tous les marqueurs ont plus de 80 % de données présentes. Dans la globalité, les données qui ont servi à construire cette carte génétique portent 93,1 % de données présentes et un taux d'erreur moyen de 1 %.

Le locus  $R_{US}$  a été introduit comme un marqueur génétique avec 43,6 % de données présentes. Dans cette dernière carte génétique, le locus  $R_{US}$  co-localise avec 3 marqueurs (RGAs135-1, 17N11-R et 125A14-F), il est encadré dans un intervalle de 1,1 cM par le marqueur 14N08-F et un groupe de 3 marqueurs (RGAs297, SCAR\_E1M4 et RGAs352). La ségrégation au niveau du locus  $R_{US}$  est de 49,52 % pour l'allèle [ $R_{US}$ ].

#### **2.4.3.4 Comparaison entre les cartes génétiques des trois séries de la famille 54**

Cette nouvelle carte génétique a été alignée avec la carte génétique plus complète faite à partir des 336 premiers individus de la famille 54 ; les marqueurs sont colinéaires entre les deux cartes et les distances génétiques sont du même ordre.



**Figure 36** : Comparaison entre les cartes génétiques de la mère 73028-62 *P. deltoides*.

A : Carte génétique élaborée avec les individus de la première extension de la famille 54. B : Carte génétique élaborée avec les 336 premiers individus. C : Carte génétique élaborée avec les individus de la seconde extension.

Les distances génétiques indiquées sont en cM. Les marqueurs possédants entre 1,5 et 3,5 % d'erreurs sont indiqués en italique. Les marqueurs possédants entre 40 et 80% de données présentes sont indiqués en souligné. Les marqueurs qui ont été utilisés à l'URGV sont indiqués en rouge. Les marqueurs qui ont été utilisés à l'INRA d'Orléans et à l'URGV sont indiqués en orange.

## 2.4.4 Cartographie génétique de la mère 73028-62

### 2.4.4.1 Carte génétique locale utilisant les 336 premiers individus

Pour la mère 73028-62, un individu atypique a été éliminé de l'analyse et a donc conduit à une carte génétique avec 335 individus. Parmi les 27 marqueurs génétiques, 3 ont plus de 3 % d'erreurs (Nc799-1, GCPM\_2849, INRA\_1944). Ces 3 marqueurs ont été éliminés de l'analyse, la carte regroupe et ordonne 24 marqueurs génétiques (sans compter  $R_1$ ) sur l'extrémité du chromosome 19 dans lequel se trouve le locus  $R_1$ . Parmi ces 24 marqueurs génétiques, il y a 3 marqueurs qui possèdent entre 1,5 et 3,5 % d'erreurs, et 7 marqueurs qui ont entre 40 et 80 % de données présentes (figure 36 B). Dans la globalité, les données qui ont servi à construire cette carte génétique portent un taux d'erreur moyen de 0,66 %, et 82,6 % de données présentes.

Le locus  $R_1$  a été introduit comme un marqueur génétique avec 99,7% de données présentes. Dans cette première carte génétique,  $R_1$  est encadré par un groupe de marqueurs (INRA\_1211 et GCPM\_79) et le marqueur INRA\_920-3 distants de 1,6 cM. La ségrégation au niveau du locus  $R_1$  est de 55,82 % pour l'allèle [ $R_1$ ].

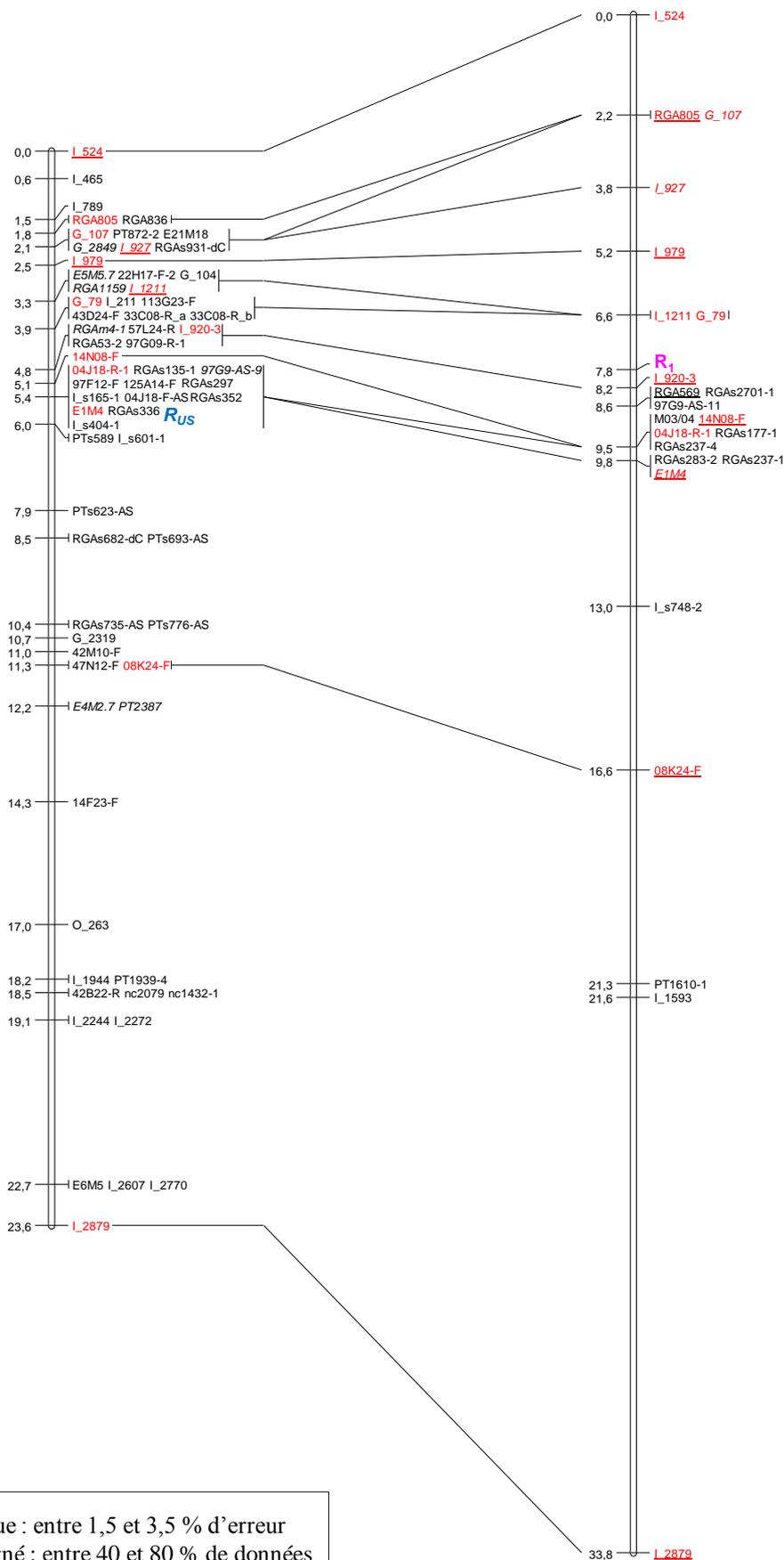
### 2.4.4.2 Carte génétique locale utilisant la première extension

Certains marqueurs génétiques ont été ensuite sélectionnés pour génotyper les 1079 individus de la première extension. Ces marqueurs ont été sélectionnés par rapport à leur qualité et leur proximité avec le locus  $R_1$ . Sur les 1079 individus, un individu atypique a été éliminé, l'analyse a donc été faite avec 1078 individus. Sur les 13 marqueurs génétiques qui ont été passés sur cette série (sans compter le locus  $R_1$ ), un marqueur a été éliminé car il possédait plus de 3,5 % d'erreurs (SCAR\_E1M4), la carte regroupe et ordonne donc 12 marqueurs génétiques. Parmi ceux-ci, il y a 5 marqueurs qui possèdent entre 1,5 et 3,5 % d'erreurs, et 6 marqueurs qui ont entre 40 et 80 % de données présentes (figure 36 A). Dans la globalité, les données qui ont servi à construire cette carte génétique portent 78,6 % de données présentes et un taux d'erreur moyen de 1,23 %.

Le locus  $R_1$  a été introduit comme un marqueur génétique avec 94 % de données présentes. Dans cette seconde carte génétique,  $R_1$  est encadré par deux groupes de marqueurs génétiques distants de 2 cM (INRA\_1211 et GCPM\_79 pour le premier groupe, INRA\_920-3, RGAs2701-1 et 14N08-F pour le second). La ségrégation au niveau du locus  $R_1$  est de 51,39 % pour l'allèle [ $R_1$ ].

**A : Carte génétique locale  
(101-74)  
*P. trichocarpa*  
LG XIX  
333 individus**

**B : Carte génétique locale  
(73028-62)  
*P. deltooides*  
LG XIX  
335 individus**



En italique : entre 1,5 et 3,5 % d'erreur  
En souligné : entre 40 et 80 % de données

**Figure 37** : Alignement entre les cartes génétiques du 101-74 et du 73028-62.  
A : Carte génétique du 101-74 élaborée avec les 336 premiers individus de la famille 54.  
B : Carte génétique du 73028-62 élaborée avec les 336 premiers individus de la famille 54.  
Les marqueurs communs aux deux cartes génétiques sont indiqués en rouge.

#### **2.4.4.3 Carte génétique locale utilisant la seconde extension**

Neuf marqueurs génétiques ont été sélectionnés pour géotyper cette dernière extension de 729 individus supplémentaires. Deux individus atypiques ont été éliminés (tableau 5) : l'analyse est donc faite avec 727 individus (figure 36 C). Un marqueur a été éliminé car il possédait plus de 3,5 % d'erreurs (SCAR\_E1M4), la carte ordonne 8 marqueurs génétiques dont un marqueur possédant entre 1,5 et 3,5 % d'erreur, et 4 possédant entre 40 et 80 % de données présentes. Dans la globalité, les données qui ont servi à construire cette carte génétique portent 75,3 % de données présentes et un taux d'erreur moyen de 0,52 %.

Le locus  $R_1$  a été introduit comme un marqueur génétique avec 43,5 % de données présentes. Dans cette dernière carte génétique,  $R_1$  est encadré par deux groupes de marqueurs génétiques distants de 1,7 cM (INRA\_1211 et GCPM\_79 pour le premier groupe, RGAs2701-1 et 14N08-F pour le second). La ségrégation au niveau du locus  $R_1$  est de 54,12 % pour l'allèle [ $R_1$ ].

#### **2.4.4.4 Comparaison des 3 cartes génétiques du 73028-62**

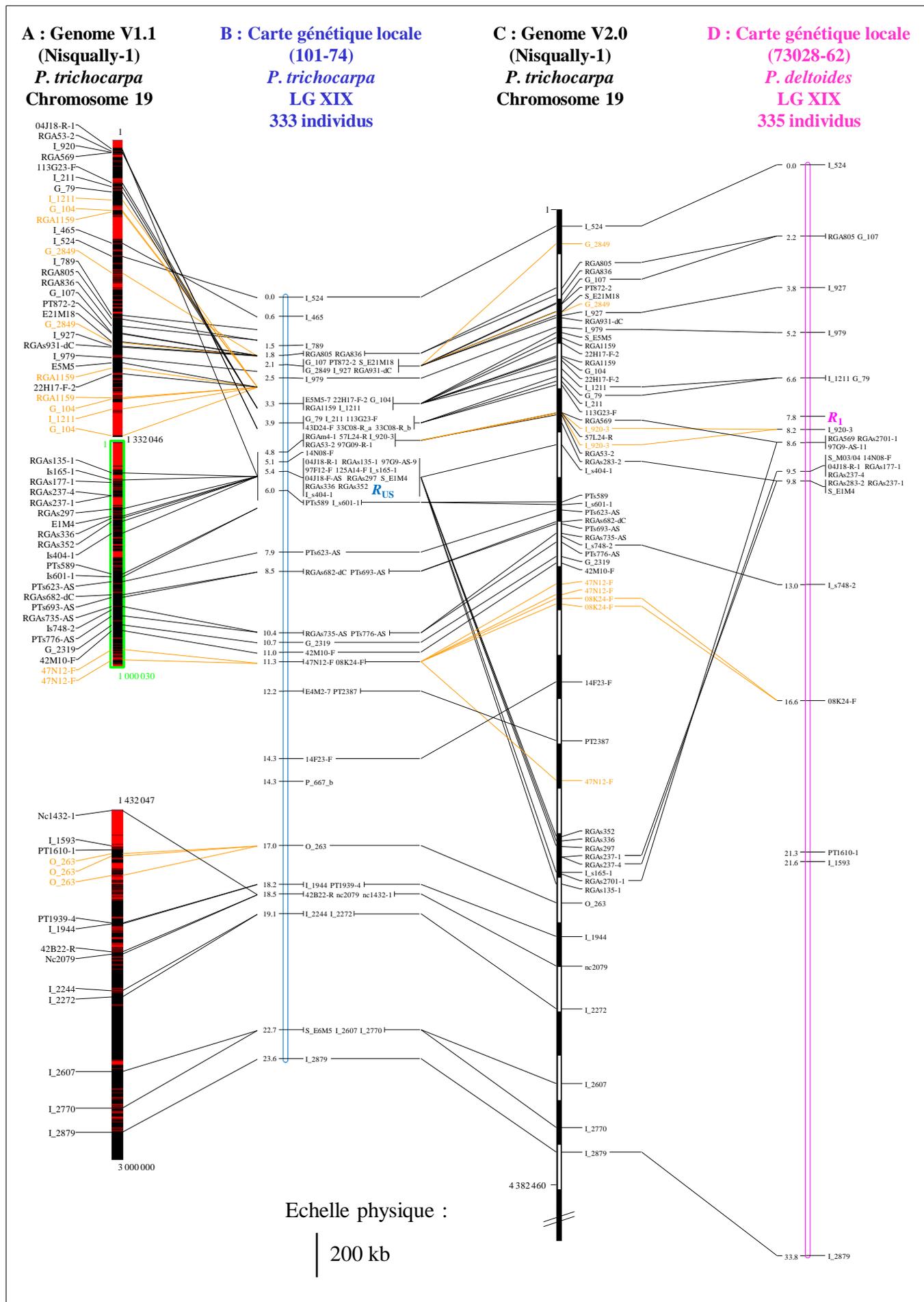
De même que pour les cartes génétiques du 101-74, les marqueurs génétiques sont colinéaires et les distances génétiques sont de même ordre entre les différentes cartes génétiques du 73028-62.

#### **2.4.5 Comparaison des deux cartes génétiques parentales**

Il y a 15 marqueurs génétiques dans la carte issue du géotypage des 336 premiers individus, qui ségrègent pour les deux parents en même temps. Le marqueur INRA\_s404-1 ségrège chez les deux parents, mais il n'a pas été placé sur le LG XIX pour le 73028-62, ce marqueur n'a pu être placé dans la carte génétique du génome entier (Véronique Jorge, donnée personnelle). Le marqueur GCPM\_2849 ségrège chez les deux parents, mais il a été écarté de l'analyse de l'individu 73028-62 car il possède plus de 3,5% d'erreur.

Il y a donc 13 marqueurs communs entre les cartes génétiques des parents pour les premiers individus (figure 37). Il y a une parfaite colinéarité entre les deux cartes génétiques parentales.

Le taux de recombinaison est plus élevé chez la mère que chez le père pour cette région génomique ; entre les marqueurs INRA\_524 et INRA\_2879, la distance génétique chez la mère est de 33,8 cM alors que chez le mâle elle est de 23,6 cM.



**Figure 38** : Alignement des cartes génétiques des deux parents et des séquences des deux assemblages du génome de Nisqually-1. Les cartes génétiques du 101-74 (B) et du 73028-62 (D) sont élaborées avec les 336 premiers individus de la famille 54. Les séquences du génomes V1 (A) et V2 (C). Dans la barre de gauche représentant les séquences : l'absence de séquence est indiquée en rouge, la présence de séquence est indiquée en noir. La barre de droite représente l'échelle tous les 200 kb. Les marqueurs génétiques possédants plusieurs locus sur le génome de Nisqually-1 sont indiqués en orange.

La relation entre les deux loci de résistance  $R_1$  et  $R_{US}$  était inconnue : il y a deux marqueurs, 14N08-F et I\_920-3, qui séparent les deux loci :  $R_1$  et  $R_{US}$  ne constituent donc pas un seul et même gène avec différents allèles, il s'agit de deux gènes différents.

#### **2.4.6 Alignement des cartes génétiques avec le génome de Nisqually-1**

Nous manquons de marqueurs qui permettraient de comparer les cartes génétiques entre elles. Il est donc plus intéressant de comparer les cartes génétiques avec l'intermédiaire de la séquence du génome.

L'alignement des cartes génétiques avec le chromosome 19 de la version V1 du génome montre une large inversion entre les cartes génétiques des deux parents et le début de ce chromosome. Le scaffold 117 s'intègre dans un trou précis de la séquence ; deux marqueurs génétiques (GCPM\_104 et nc1432-1) encadrent les marqueurs développés sur scaffold 117, et sont placés aux premiers contigs encadrant ce trou de séquence (figure 38).

Dans cette inversion, plusieurs marqueurs montrent un locus unique dans la carte génétique alors qu'il y a plusieurs locus sur la séquence du génome (GCPM\_104, RGA1159, 22H17-F-2). Une partie de cette inversion est donc vraisemblablement déjà représentée dans le génome.

L'alignement sur le chromosome 19 de la version V2 du génome montre une région mal assemblée entre les cartes génétiques et ce chromosome entre 2 800 et 3 050 Kb (figure 38).

Ces différents alignements renforcent la colinéarité entre les deux cartes génétiques ; en effet, en comparant les cartes génétiques *via* la séquence du génome, on peut aussi voir la colinéarité des marqueurs génétiques qui ne ségrègent pas en même temps chez les deux parents.

Le fait que les deux cartes génétiques parentales soient cohérentes entre elles et montrent les mêmes différences avec le génome renforce le fait qu'il s'agit d'erreurs d'assemblage. Ces erreurs d'assemblage ne sont pas des duplications, que ce soient pour l'une ou l'autre version ; en effet, tous les couples d'amorces dessinés sur ces zones retombent dans une et même zone quel que soit l'individu.

## **2.5 Cartographie physique de l'individu 101-74**

La zone étudiée pour l'élaboration de la carte physique a été réduite par rapport à la zone étudiée en cartographie génétique. Ces limites sont les marqueurs génétiques RGA805 du côté télomérique du chromosome 19 et 47N12-F à l'extrémité du scaffold 117 du génome de Nisqually-1 V1. Ces deux marqueurs sont espacés par 1 380 Kb sur la séquence du génome V2. Cette zone a été recentrée sur les loci  $R_{US}$  et  $R_1$ .



## **2.5.1 Criblage de la banque BAC**

### **2.5.1.1 Criblage par PCR**

Cent cinq couples d'amorces ont été utilisés pour le criblage de 88 pools plaques de la banque BAC, ce qui représente 10 X. Trente cinq d'entre eux ont servi à cribler les 48 autres pools plaques restants de la banque BAC. Ces marqueurs utilisés correspondent aux zones où la couverture de la banque n'était pas suffisante avec les 88 premiers pools plaques.

Les pools lignes correspondant au pool plaque positif et les clones individuels d'une ligne correspondante au pool ligne positif ont été criblés avec le même couple d'amorces.

Le criblage de la banque BAC a permis l'identification de 95 clones BAC correspondants à la zone d'intérêt.

### **2.5.1.2 Criblage avec les BES des 20 premières plaques de la banque BAC**

Dans le criblage des 88 premiers pools plaque, il y avait seulement 4 plaques communes ; il peut donc y avoir de nouveaux clones BAC caractérisés par le criblage des 20 premières plaques de la banque BAC. De plus, ces clones BAC possèdent des BES utilisables pour redessiner de nouveaux couples d'amorces et donc permettent d'améliorer l'élaboration des cartes physiques.

L'alignement des 13 536 séquences de BES provenant des 20 premières plaques de la banque BAC avec la séquence du génome a permis d'identifier 30 clones BAC supplémentaires potentiellement présents dans la zone d'intérêt.

Les clones BAC sélectionnés ont été récupérés par rapport à leurs coordonnées dans les plaques de la banque. La confirmation par PCR avec un couple d'amorces concordantes avec la localisation des BES a permis de valider 20 de ces clones BAC. Neuf des 10 clones BAC non confirmés montrent une forte similarité avec des éléments transposables (Cf. § III-2.5.2.2) et 4 d'entre eux correspondent à un élément particulièrement redondant à proximité du locus  $R_{US}$ . Un autre clone BAC non confirmé (le clone 04P24) possède une de ces extrémités sur un gène NBS-LRR qui correspond au clade 1 des gènes TNL.

## **2.5.2 Utilisation des BES en cartographie physique**

Il y a 20 clones BAC qui possèdent des BES provenant des 20 premières plaques de la banque BAC. En plus, parmi les 95 clones BAC criblés par PCR, 51 clones BAC ont été séquencés à leurs extrémités. Au total, 71 clones BAC possèdent un ou deux BES, ce qui a permis d'obtenir 134 BES (8 clones BAC ne possèdent qu'un seul BES). Sur ces 134 BES, 4



couples de BES provenant de clones BAC différents correspondent à un même locus, il y a donc 130 séquences différentes.

### **2.5.2.1 Alignement des BES sur la séquence du génome**

#### **2.5.2.1.1 Génome V1**

L'alignement de ces 130 BES sur le génome de Nisqually-1 (V1) a permis de caractériser 215 alignements sur le génome, sachant qu'au maximum quatre alignements ont été pris en compte. Dans de nombreux cas, ces alignements sont le raboutement de plusieurs alignements qui ont été coupés par des séquences de faible complexité ou de courtes zones différentes. Six BES n'ont pas d'alignement sur le génome, 10 BES n'ont pas été pris en compte car ils possédaient trop d'alignements ; pour 7 d'entre eux, il s'agit d'éléments transposables (Cf. § III-2.5.2.2). La confirmation de ces alignements avec le logiciel MEGA4 a donc permis de valider 176 alignements, dont 126 correspondent à la zone d'intérêt ; 39 alignements n'ont pas été validés.

#### **2.5.2.1.2 Génome V2**

L'alignement de ces 130 BES sur le génome de Nisqually-1 (V2) a permis de caractériser également 215 alignements sur le génome, sachant qu'au maximum 4 alignements ont été pris en compte. Cinq BES n'ont pas d'alignement sur le génome, quatre BES n'ont pas été pris en compte car ils possédaient trop d'alignements ; pour 3 d'entre eux, il s'agit d'éléments transposables. La confirmation des ces alignements avec le logiciel MEGA4 a permis de valider 161 alignements, dont 127 correspondent à la zone d'intérêt ; 54 alignements n'ont pas été validés.

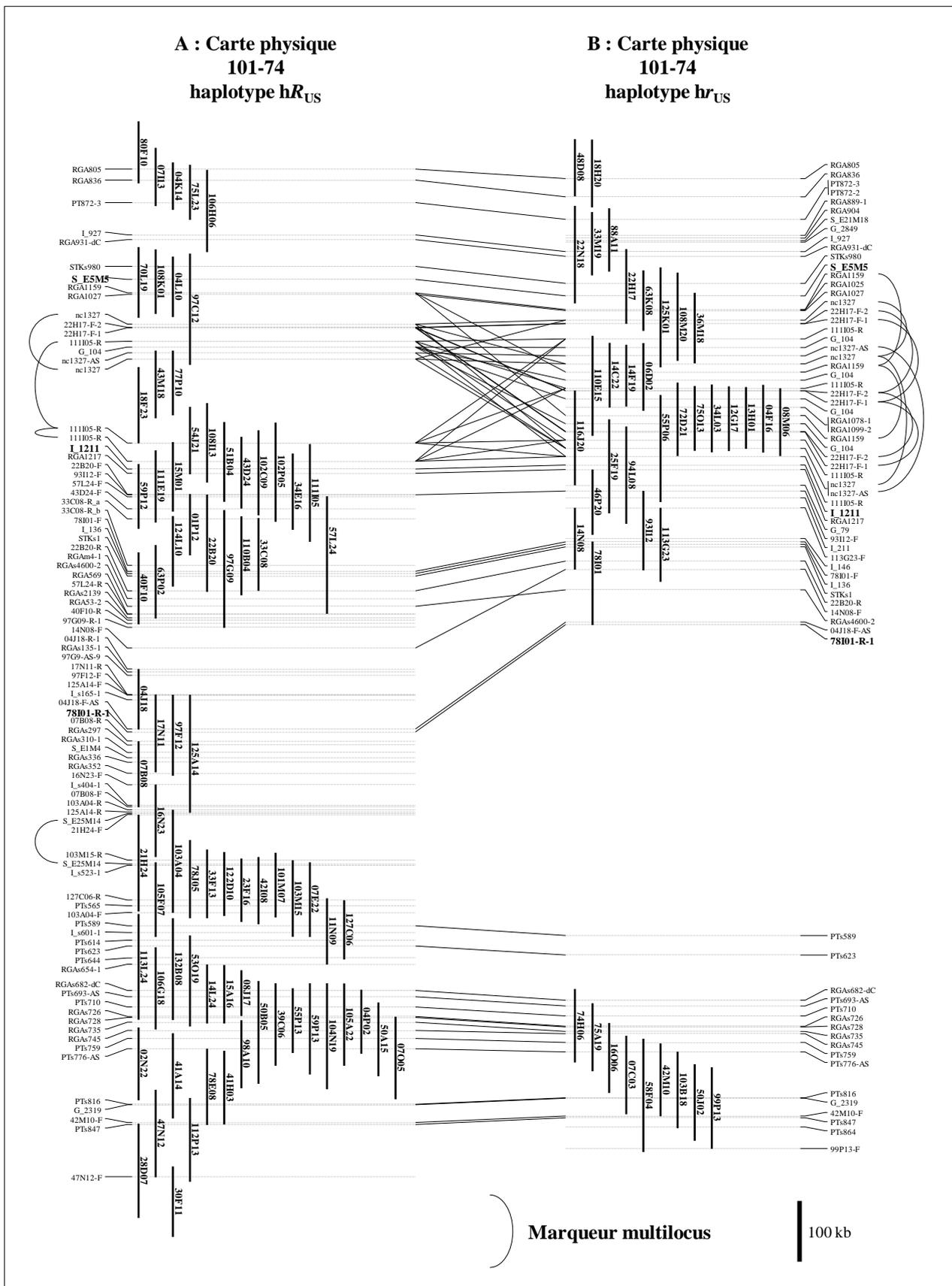
### **2.5.2.2 REP BASE**

Une analyse des BES a été effectuée par REP BASE. Trente quatre des 130 BES possèdent un élément transposable, 11 d'entre eux montrent un alignement de moins de 200pb.

### **2.5.2.3 Orientation des BES**

L'orientation de ces BES a pu se faire de deux façons :

- Des marqueurs dessinés sur un ou les deux BES d'un clone BAC ont été utilisés ; le criblage des clones BAC individuels avec ces marqueurs a permis de voir de quel côté l'extrémité se situe. En effet ces marqueurs ne marchent que chez les clones BAC chevauchant pour l'un de ces deux côtés. Cette méthode expérimentale a permis d'orienter les BES de 25 clones BAC.



**Figure 39** : Construction des cartes physiques pour chacun des deux haplotypes du 101-74. A : carte physique de l'haplotype hR<sub>US</sub>, B : carte physique de l'haplotype hr<sub>US</sub>. L'alignement entre les deux cartes physiques est faite avec les marqueurs PCR communs. A droite de chaque clone BAC (représenté par un trait noir vertical) est indiquée la coordonnée de ce clone dans la banque BAC. Les marqueurs multi-locus sont indiqués par une courbe regroupant les différents loci.

- L'utilisation des alignements des BES sur la séquence du génome a permis aussi de les orienter. Il fallait que les deux BES soient en orientation inverse entre eux sur la séquence du génome et espacés d'une longueur cohérente avec celle d'un clone BAC. Les BES de 36 clones BAC ont été orientés par ces alignements.

Dix clones BAC possédant des BES n'ont pas pu être orientés. Ces clones BAC ne possèdent pas de couple d'amorces sur l'un des BES et ils ont des alignements incohérents sur le génome, ce qui ne permet pas de les orienter.

### **2.5.3 Assemblage des contigs de clone BAC**

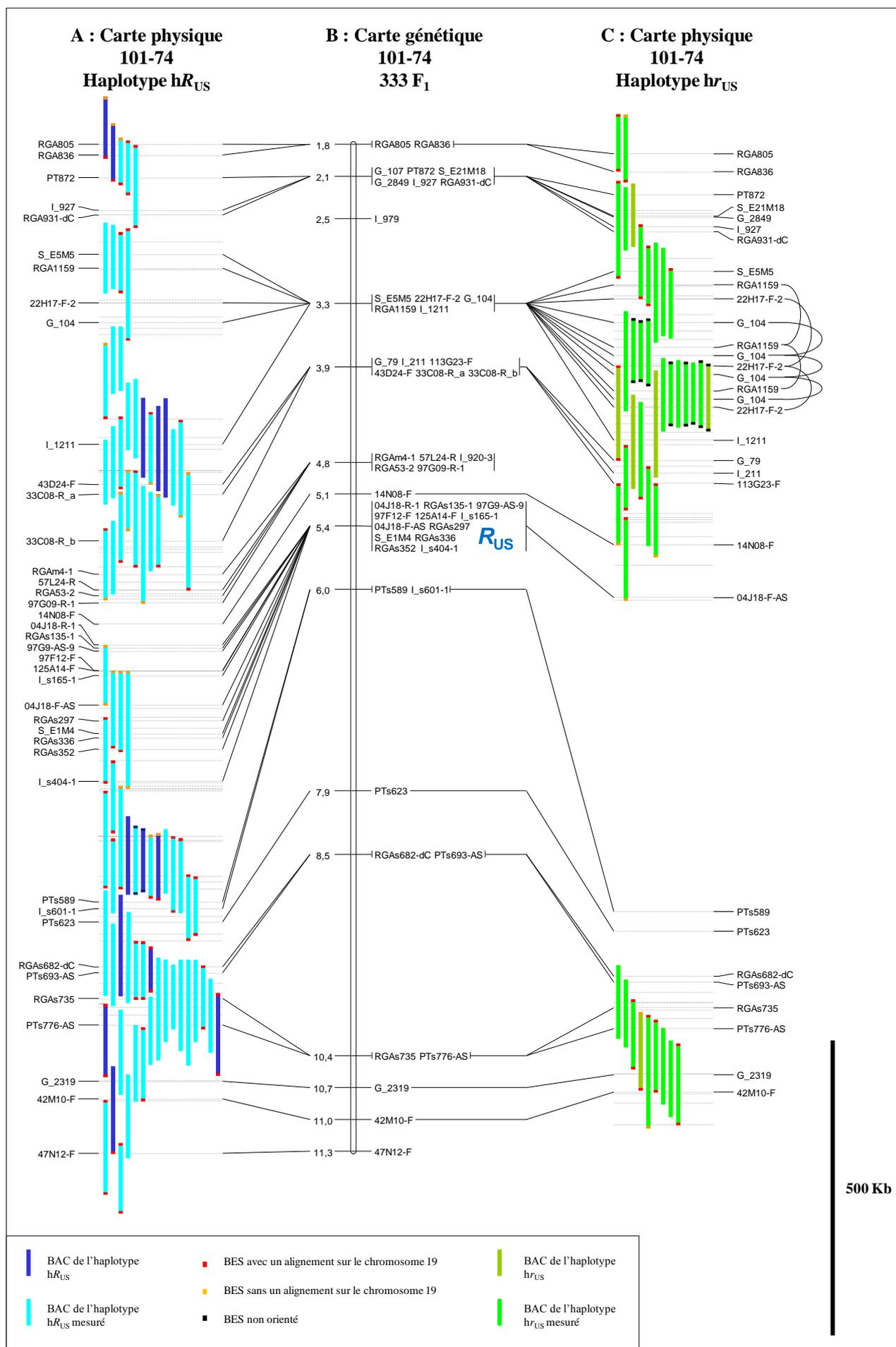
Le criblage des clones BAC individuels avec 96 couples d'amorces a permis de construire les contigs. L'appartenance à l'un ou l'autre haplotype pour chacun des clones BAC a pu être déterminée à l'aide des marqueurs génétiques (figure 39). Les deux cartes physiques ont donc pu être construites indépendamment pour chaque haplotype.

Cent quinze clones BAC ont été intégrés aux deux cartes physiques, 76 pour l'haplotype  $hR_{US}$  et 39 pour l'haplotype  $hr_{US}$ . La carte physique de l'haplotype  $hR_{US}$  a donné deux contigs ; le premier de 32 clones BAC, couvre 850 Kb, et est délimité par les marqueurs RGA805 et 97G09-R-1; et le second de 44 clones BAC, couvre 900 Kb, et est délimité par les marqueurs 04J18-1-R et 47N12-F. La carte physique de l'haplotype  $hr_{US}$  a donné deux contigs ; le premier, de 39 clones BAC, couvre 1 000 Kb, et est délimité par les marqueurs RGA805 et 78I01-R-1 ; le second, de 9 clones BAC, couvre 350 Kb, et est délimité par les marqueurs RGA682-dC et 99P13-R.

#### **2.5.3.1 Les marqueurs multi-locus**

Avec le criblage des pools plaques, de nombreux marqueurs se sont révélés multi-locus, que ce soit à l'extérieur de la zone ou à l'intérieur de la zone.

Avec le criblage des clones BAC individuels, certains couples d'amorces se sont révélés multi-locus à l'intérieur de la zone, sans pouvoir forcément différencier ces locus. Une partie de ceux-ci a été éliminée, mais une zone, entre les marqueurs E5M5 et I\_1211, renferme des marqueurs qui sont tous multi-locus (figure 39). La redondance dans cette zone est très importante car certains de ces marqueurs sont dessinés dans des régions inter-géniques. Seul le couple d'amorce GCPM\_104 permet de construire correctement la carte physique de l'haplotype  $hr_{US}$  car il amplifie 4 locus avec 3 tailles d'amplification différentes. La carte physique de l'haplotype  $hR_{US}$  n'est pas correctement caractérisée pour la même zone, en raison de l'absence de marqueur avec un polymorphisme de taille ou de marqueur mono-locus.



**Figure 40** : Alignement entre les cartes physiques et la carte génétique du 101-74.

A : carte physique de l'haplotype  $hR_{US}$ . B : carte génétique du 101-74 élaborée avec les 336 premiers individus de la famille 54. C : carte physique de l'haplotype  $hr_{US}$ . Seuls les marqueurs génétiques sont indiqués sur les cartes physiques.

## 2.5.4 Alignement des cartes physiques avec la carte génétique

L'alignement des deux cartes physiques avec la carte génétique du même individu a permis de confirmer l'assemblage interne de chaque contig et l'assemblage des contigs entre eux (figure 40). La cohérence entre ces deux cartes physiques et la carte génétique renforce la bonne construction de chacune des cartes.

Pour les 96 couples d'amorces utilisés pour la cartographie physique, seuls 42 d'entre eux fonctionnent sur les deux haplotypes soit moins de la moitié.

## 2.5.5 Alignement de la carte physique hR<sub>US</sub> avec le génome de Nisqually-1

### 2.5.5.1 Alignement *via* les couples d'amorces

Les coordonnées des couples d'amorces sur le génome de Nisqually-1 ont été déterminées par e-PCR et par BLASTn. La limite prise en compte a été d'au maximum deux SNP entre la séquence de l'amorce et la séquence du génome. L'alignement des marqueurs entre la carte physique de l'haplotype hR<sub>US</sub> et le génome est montré en noir dans la figure 41.

### 2.5.5.2 Alignement *via* les BES

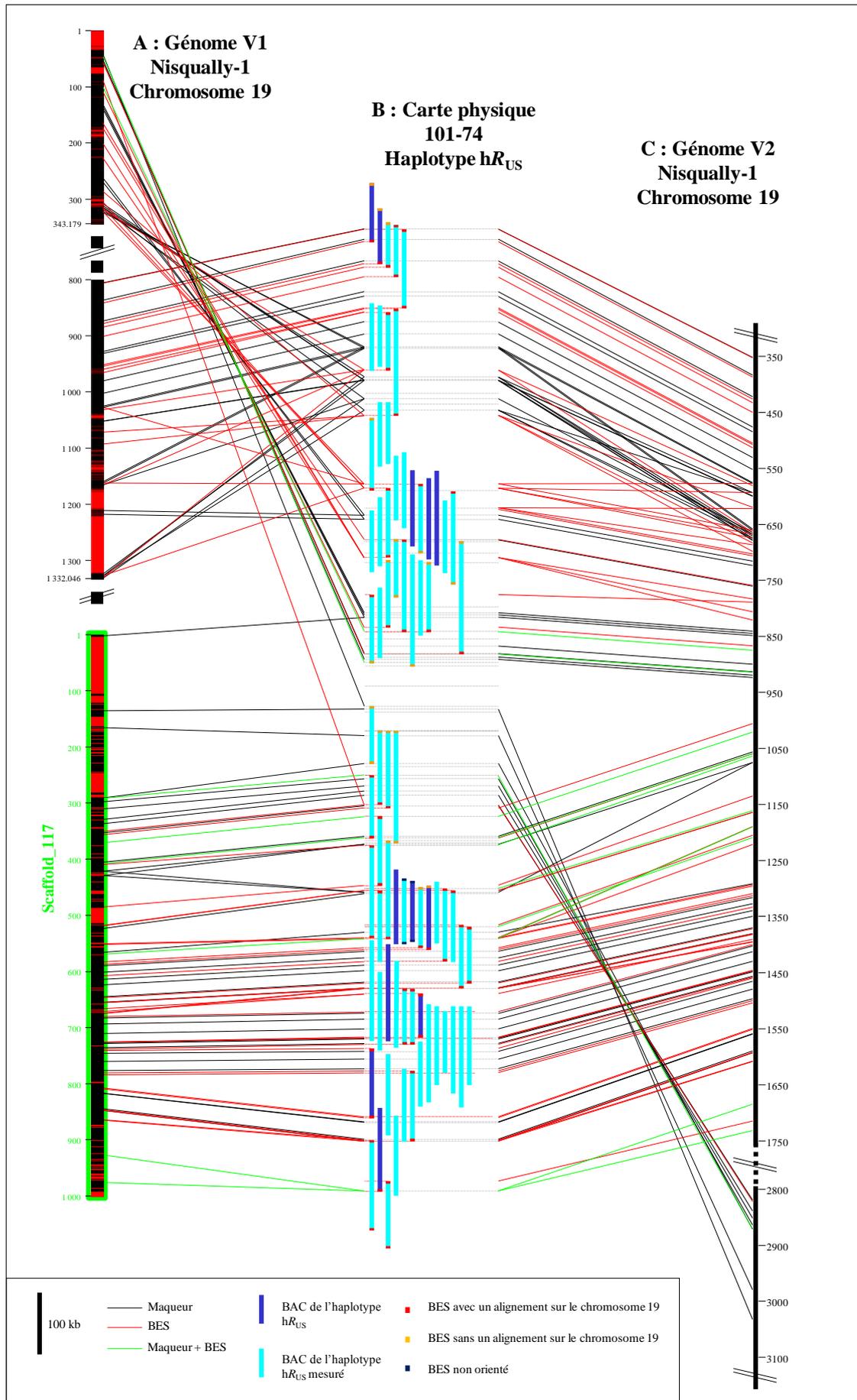
Les cartes physiques ont été comparées à la séquence disponible de *P. trichocarpa* V1 (figure 41 A) et V2 (figure 41 C). La séquence du génome comporte de nombreuses discontinuités dans la zone d'intérêt et leur longueur est mal définie. Par exemple, d'après les données de cartographie, le scaffold 117 s'intègre au groupe de liaison XIX dans un trou de 100 Kb, alors que ce scaffold a une longueur de 1 Mb.

Par ailleurs, une région de 350 Kb environ apparaît avoir été mal placée et inversée dans l'assemblage V1 du JGI ce qui confirme ce qui a été déjà remarqué avec les cartes génétiques (Cf. § III-2.4.6). De même, l'erreur de 250 Kb dans la version V2 de l'assemblage est aussi confirmée.

L'alignement des BES sur le génome est montré en rouge dans la figure 41 et en vert quand il y a un marqueur développé sur ce BES.

### 2.5.5.3 Estimation de la taille des cartes physiques

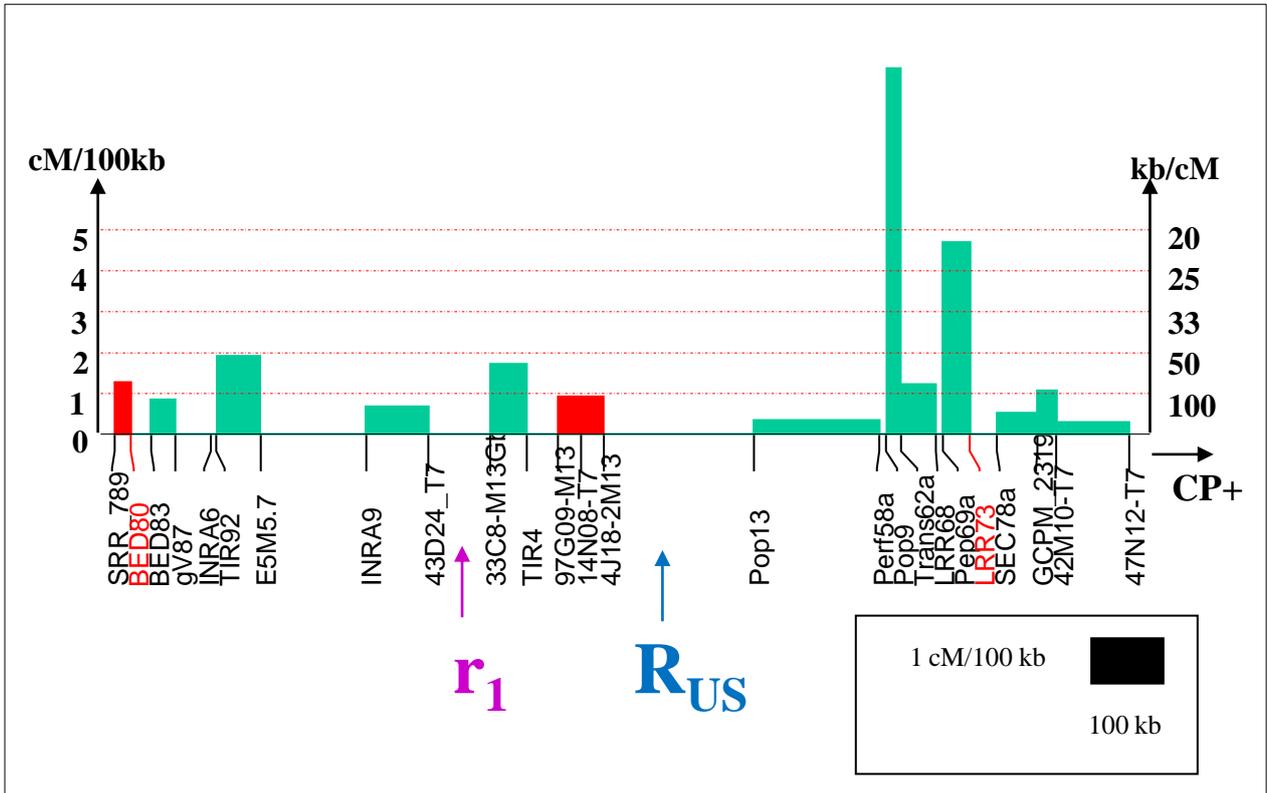
La bonne gestion des deux cartes physiques a été assurée en donnant une coordonnée pour les différents paramètres des cartes physiques. Ces paramètres sont les marqueurs et les BES et par conséquent les clones BAC.



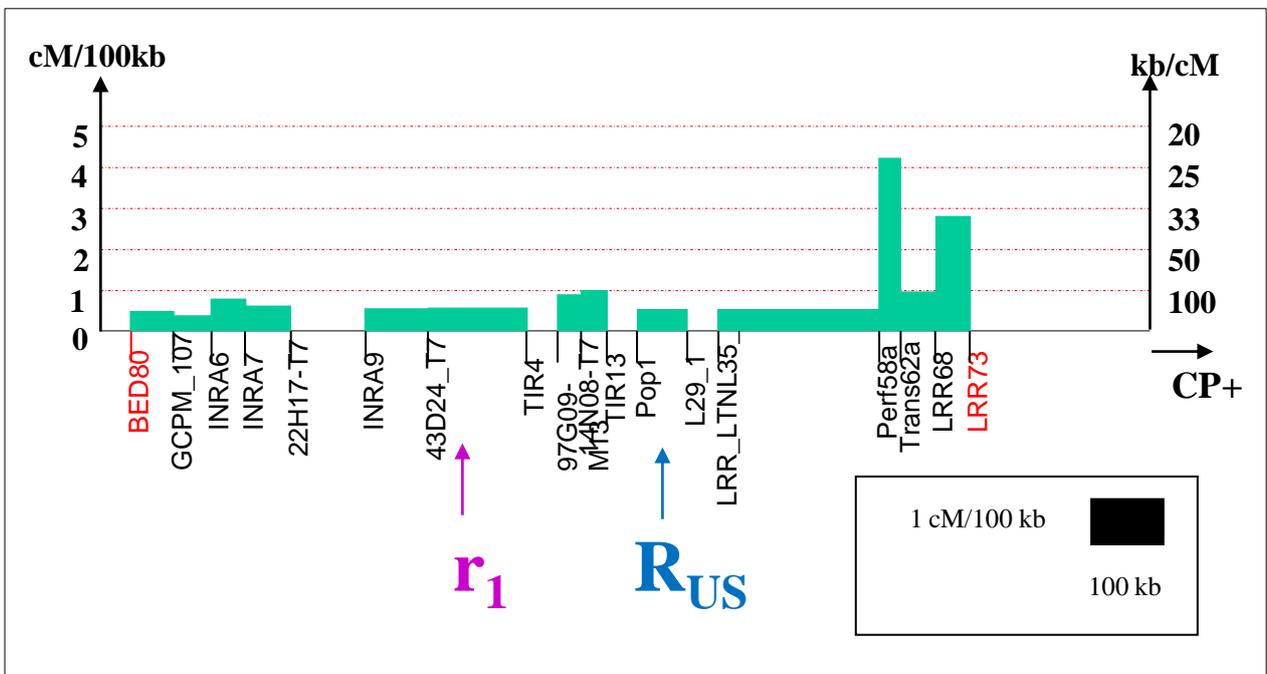
**Figure 41** : Alignement entre la carte physique de l'haplotype  $hR_{US}$  et les deux assemblages du génome de Nisqually-1.

B : carte physique de l'haplotype  $hR_{US}$ . Les séquences des génomes V1 (A) et V2 (C).





**Figure 42 :** Taux de recombinaison entre la carte génétique de 333 individus et la carte physique de l'haplotype  $hR_{US}$ .



**Figure 43 :** Taux de recombinaison entre la carte génétique de 1399 individus et la carte physique de l'haplotype  $hR_{US}$ .

Quatre vingt treize clones BAC ont été mesurés parmi les 115 des deux cartes physiques (la taille moyenne est de 129 Kb).

Ces tailles mesurées expérimentalement sont comparées aux tailles théoriques sur la séquence du génome. Ces tailles théoriques sont prises par rapport aux alignements des BES sur le génome ou par rapport à l'alignement des couples d'amorces. Dans le cas des BES, la taille théorique est précise mais dans le cas des couples d'amorces la taille est un intervalle.

La comparaison entre ces mesures a permis de définir les différences qui peuvent intervenir entre le génome de Nisqually-1 et les cartes physiques du 101-74. Cette étape est conduite indépendamment entre les deux cartes physiques. La taille de la carte physique de l'haplotype  $hR_{US}$  est d'environ 1 800 Kb. Par contre, la taille de l'haplotype  $hr_{US}$  est difficile à estimer du fait de l'important trou qu'il possède.

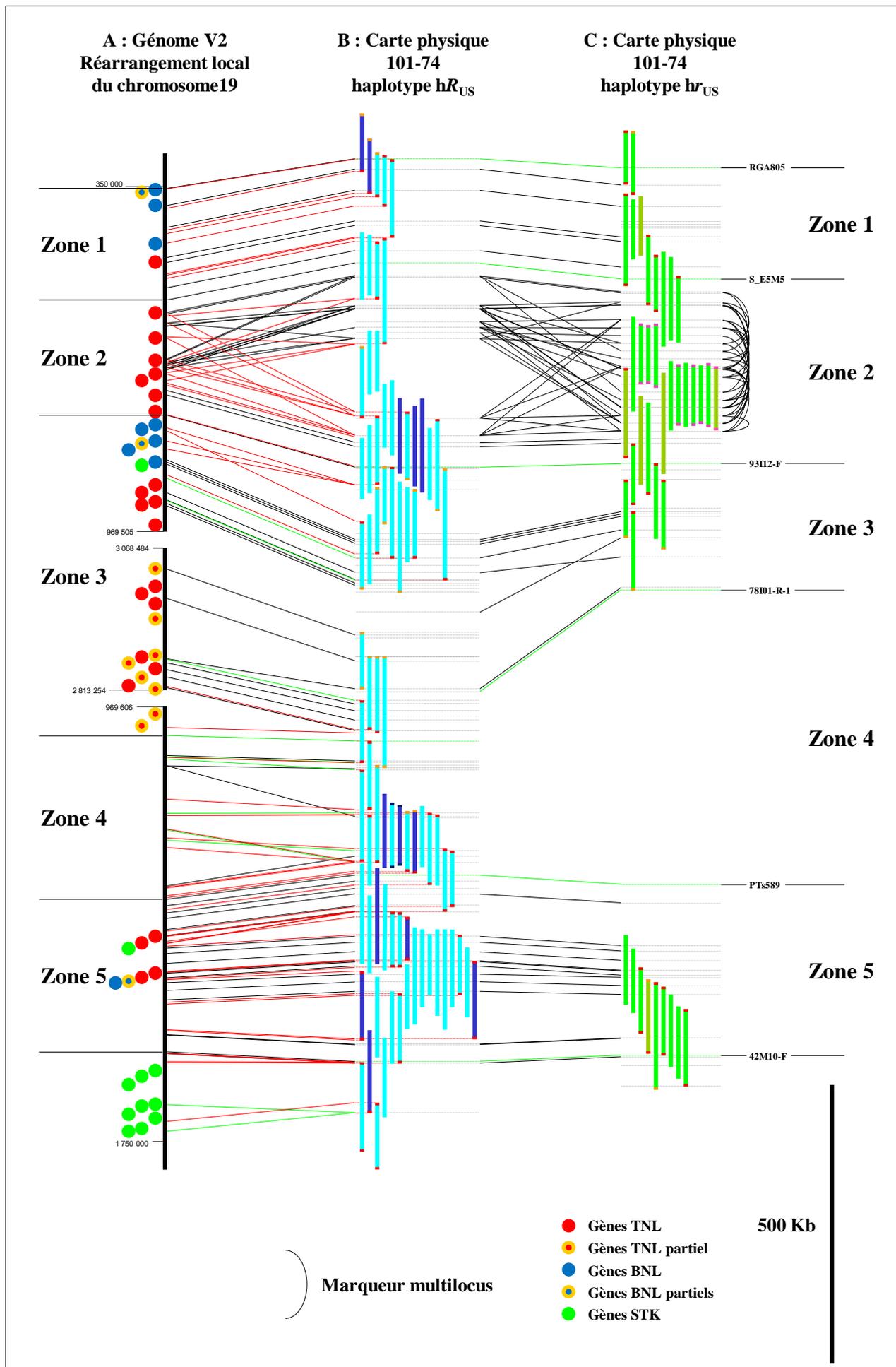
Ceci a permis de caractériser les différences de taille qui existent entre les deux haplotypes de l'individu 101-74 : la figure 39 (Cf. page III-83) montre une différence de distance entre les marqueurs I\_1211 et 4J18-F-AS : entre ces deux marqueurs, la distance sur l'haplotype  $hr_{US}$  est d'environ 270 Kb alors que elle est au minimum de 370 Kb sur l'haplotype  $hR_{US}$ . Pour cette même zone les différences éventuelles de distances entre le 101-74 et Nisqually-1 ne peuvent pas être observées ; en effet, il s'agit de la zone où les deux versions sont mal assemblées.

### **2.5.6 Taux de recombinaison**

La moyenne du ratio distance génétique/distance physique pour la zone est d'environ 1cM pour 200 Kb. Il existe cependant une certaine variabilité de cette valeur à l'intérieur de la zone étudiée.

La figure 42 montre les taux de recombinaison par rapport à la carte génétique 333 et la carte physique de l'haplotype  $hR_{US}$ . La figure 43 montre les taux de recombinaison avec la carte génétique faite à partir de 1399 individus (carte portant sur l'analyse des 336 individus et la première extension de la famille 54).

Le taux de recombinaison est équivalent entre les deux cartes génétiques. Les figures montrent deux zones qui sont particulièrement recombinantes : les différences du taux de recombinaison entre les deux cartes génétiques sont probablement dues au fait que les intervalles ne sont pas exactement les mêmes.



**Figure 44** : Alignement entre le génome et les cartes physiques : définition de cinq zones différentes.

Seuls les marqueurs utilisés dans le texte sont représentés et délimitent les 5 zones.

## 2.6 Caractéristiques de la zone génomique

Dans ce premier travail, l'évaluation du polymorphisme et de la redondance est faite par l'intermédiaire des marqueurs PCR, des BES, et de la taille des régions génomiques.

- Les BES localisés et orientés permettent de montrer le polymorphisme et la redondance avec le génome de Nisqually-1.

- La mesure des clones BAC a permis de mesurer la taille physique des haplotypes du 101-74. Ceci a permis de comparer les tailles des deux haplotypes entre eux et avec le génome de Nisqually-1.

- Les marqueurs PCR permettent d'aligner et donc de comparer les cartes physiques et de montrer le polymorphisme lorsqu'il ne fonctionne pas et la redondance lorsqu'ils sont multi-locus.

Par l'intermédiaire des comparaisons des deux haplotypes de l'individu 101-74, mais aussi entre un des deux haplotypes du 101-74 et Nisqually-1, il est possible de délimiter cinq zones différentes suivant leur degré de polymorphisme et de leur redondance dans la région des cartes physiques d'environ 1,8 Mb (figure 44). Il est donc proposé un réarrangement local des contigs de séquence pour les deux assemblages du génome de Nisqually-1.

Zone 1) La première zone est peu complexe, depuis le marqueur RGA805 jusqu'au marqueur SCAR\_E5M5 (figure 44) ; les marqueurs fonctionnent généralement sur les deux allèles d'un seul et même locus, les alignements des marqueurs et des BES sur le génome sont cohérents, et la taille du segment chromosomique est peu variable : 200 Kb pour le génome de Nisqually-1 pour l'haplotype  $h_{R_{US}}$ , et 187 Kb pour l'haplotype  $h_{R_{US}}$ .

Zone 2) La zone suivante est particulièrement redondante entre les marqueurs SCAR\_E5M5 et INRA\_1211 pour une longueur d'environ 300 Kb. Lorsque le clone BAC est chevauchant à cette zone, l'orientation des BES est faite en utilisant les données du BES extérieur. Si le clone BAC est entièrement dans la zone, les BES ne peuvent pas être localisés ; ils ont différents alignements multiples à l'intérieur même de cette zone. Plusieurs d'entre eux possèdent deux BES qui s'alignent à quelque Kb en orientation inverse et qui sont représentés plusieurs fois. Les marqueurs PCR sont quasiment tous multi-locus au sein du même haplotype et ont parfois plusieurs localisations sur le génome de Nisqually-1, même pour les marqueurs dessinés dans les régions inter-géniques. L'hypothèse est qu'il y a 3 répétitions d'environ 80 Kb qui s'enchainent dans la même orientation mais que seules 2 répétitions sont présentes dans l'haplotype. Cette zone montre des duplications récentes qu'il est quasi impossible à assembler



correctement avec le séquençage d'un génome en WGS et reste très difficile à travailler en cartes génétiques et physiques produites.

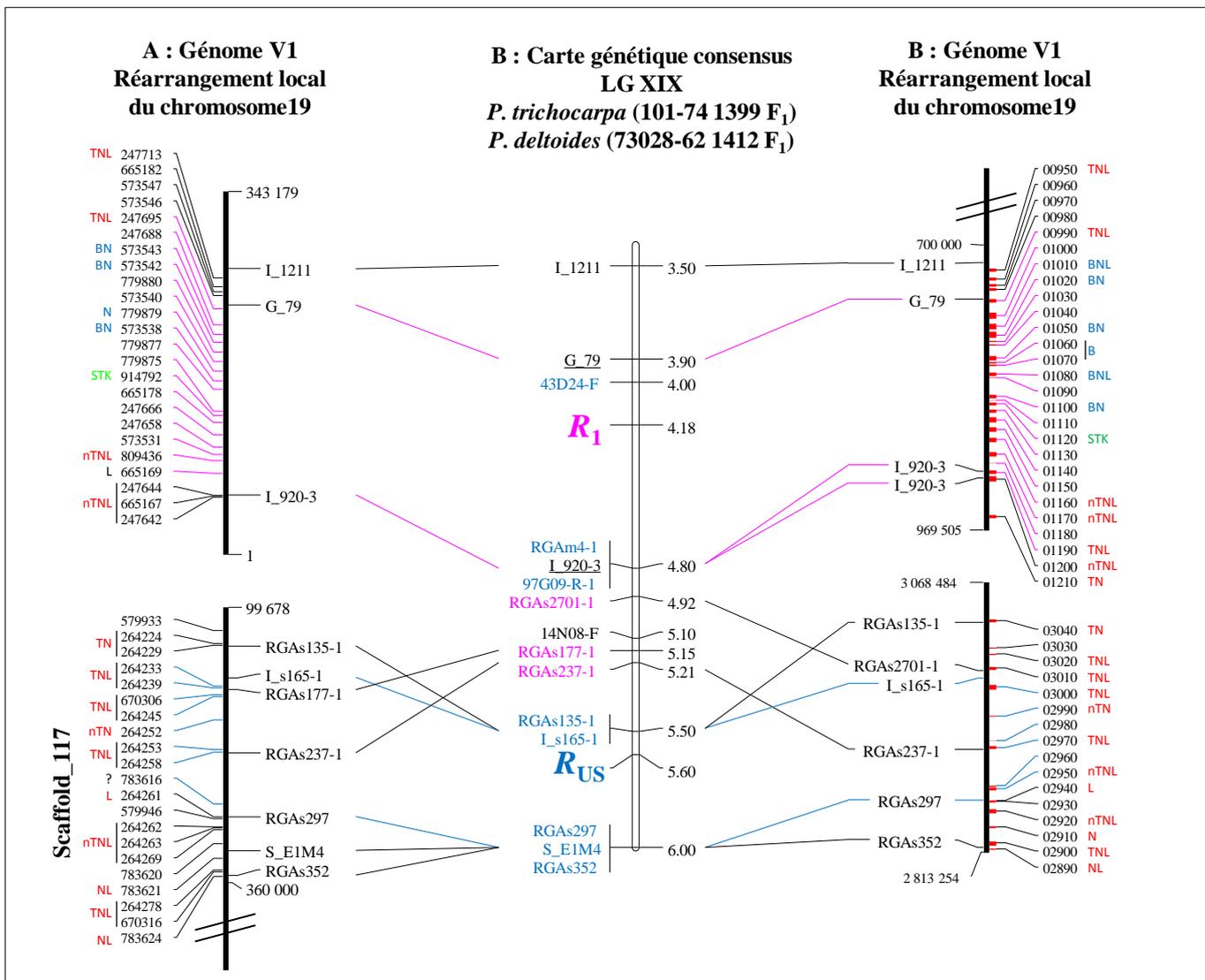
Seul l'haplotype  $hr_{US}$  pour cette zone a pu être correctement caractérisé car le marqueur multi-locus GCPM\_104 montre trois amplifications de taille différente, ce qui a permis d'assembler correctement les clones BAC en contig. C'est pourquoi il y a une partie de l'haplotype  $hR_{US}$  sans marqueur ; il y a certainement des marqueurs qui fonctionnent sans savoir sur combien de locus.

Dans les zones très dupliquées, l'assemblage des clones BAC en contig a été problématique : les marqueurs étaient répétés à très courte distance, il était impossible de différencier les loci. Cela pouvait créer des incohérences dans les cartes physiques avec des mauvaises interprétations du chevauchement des clones BAC. Cette zone de 300 Kb environ est concernée par ces difficultés qui n'ont pas pu être complètement contournées.

Zone 3) La zone délimitée par les marqueurs INRA\_1211 et RGAs352 (pour l'haplotype  $hR_{US}$ ) est particulièrement hétérozygote (figure 44). Sur les 21 marqueurs fonctionnant sur l'haplotype  $hR_{US}$ , seulement huit fonctionnent sur l'autre haplotype du 101-74. La taille entre les marqueurs INRA\_1211 et 04J18-F-AS est très différente, sur l'haplotype  $hr_{US}$  elle fait environ 270 Kb alors que sur l'haplotype  $hR_{US}$ , elle fait au minimum 370 Kb et au minimum 450 Kb sur le génome version V2. Les marqueurs dessinés sur le génome ne fonctionnent pas toujours sur le 101-74, pourtant de la même espèce *P. trichocarpa* et aussi rarement sur l'individu 73028-62 *P. deltoides* (Cf. figure 32, page III-72). De plus, beaucoup de BES n'ont pas un bon alignement avec le génome.

Cette zone contient un cluster de BNL et un cluster de TNL qui font tous partie du clade 1 des TNL. Ces derniers sont particulièrement redondants à cause de leur nombre important mais surtout pour la proximité de séquence qu'ils ont entre eux, même dans leurs introns. C'est pourquoi dans le clade, les couples d'amorces ont été dessinés à cheval sur les gènes afin de trouver une zone spécifique extra-génique où dessiner l'amorce sur laquelle repose la spécificité du couple.

Zone 4) Une autre zone entre les marqueurs RGAs352 et PTs589 montre une forte redondance, mais cette fois-ci, hors du chromosome 19. Le marqueur INRA\_s404-1 ségrège chez les deux parents, mais il n'est pas placé sur le LG XIX pour le 73028-62, ce marqueur n'a pu être placé dans la carte génétique du génome entier (Véronique Jorge, donnée personnelle). D'autres marqueurs ont montré aussi être multi-locus ; en effet, des clones BAC ont été identifiés avec ces marqueurs mais ils ne proviennent pas de cette région. Cette zone correspond à environ 260 Kb chez l'haplotype  $hR_{US}$  et au minimum à 350 Kb chez le génome version V2.



**Figure 45** : Localisation génétique des loci de résistance sur les deux assemblages du génome de Nisqually-1.

Dans la carte génétique consensus, élaborée avec les 336 premiers individus et la première extension de la famille 54, les marqueurs génétiques en bleu sont spécifiques du père 101-74, en violet ils sont spécifiques de la mère 73028-62 et en noir ils sont communs aux deux parents.

Elle n'a pu être caractérisée pour l'haplotype  $hr_{US}$  à cause d'un manque de couverture de la banque BAC.

Zone 5) De nouveau, il y a une zone peu complexe, entre les marqueurs PTs589 et 42M10-F. Il y a si peu de polymorphisme que la plupart des marqueurs fonctionne chez les deux parents de la famille 54 et ne montre pas de ségrégation dans cette famille (Cf. figure 32, page III-72). Pour la cartographie génétique, il était nécessaire de séquencer des produits d'amplification PCR pour caractériser et utiliser du polymorphisme. Les alignements des marqueurs et des BES sur le génome sont cohérents (Cf. figure 41, page III-84-85), et la taille du 101-74 est proche à celle de Nisqually-1.

Finalement, après le marqueur 42M10-F, c'est le début d'une zone complexe avec de l'hétérozygotie et de la redondance dans un cluster de gènes STK. Cette zone est limitée par un trou dans la séquence du génome V1, la fin du scaffold 117, et vraisemblablement, il y a une erreur d'assemblage dans la V2 avec une inversion de certains contigs montrée par quelques marqueurs génétiques (Cf. figure 38, page III-80).

## 2.7 Vers le clonage positionnel de $R_{US}$ et $R_1$

Il fallait déjà remettre de l'ordre dans la globalité de cette zone génomique de 1,8 Mb avant de pouvoir élucider les localisations précises des loci  $R_1$  et  $R_{US}$  individuellement (figure 41, page III-84-85). La construction de cartes physiques et génétiques précises était donc importante pour retrouver un ordre cohérent dans cette région complexe.

Ce travail a permis de connaître la relation entre les loci  $R_1$  et  $R_{US}$  : ils sont deux loci différents. Le locus  $R_{US}$  est placé dans le cluster du clade 1 des TNL (figure 45). L'intervalle de  $R_1$  possède encore plusieurs catégories de gènes ; des TNL, 3 gènes du clade 1 dont  $R_{US}$  fait probablement partie et un gène d'un autre clade, 6 gènes de la catégorie des BNL et un gène STK.

Les loci  $R_1$  et  $R_{US}$  sont placés dans une zone particulièrement polymorphe mais avec des gènes très redondants (zone 3 dans la figure 44, page III-86). Les deux zones génomiques dans lesquelles se trouvent les deux loci sont mal assemblées dans le génome de Nisqually-1, l'intervalle de  $R_1$  correspond à l'erreur du premier assemblage et qui est correctement placé dans le second. L'intervalle de  $R_{US}$ , quant à lui, est placé correctement dans un scaffold non assigné à un chromosome du premier assemblage et est mal placé dans le second assemblage.

Maintenant, le séquençage de clone BAC du 101-74, pour le clonage positionnel du locus  $R_{US}$ , permettra de connaître les réels gènes qui sont dans l'intervalle de ce locus. Les nouveaux



marqueurs dessinés dans la séquence des clones BAC permettront de restreindre au maximum l'intervalle dans lequel se trouve  $R_{US}$ .

La cartographie comparée a permis de montrer la conservation de l'ordre des séquences entre *P. trichocarpa* et *P. deltoides*. Le séquençage de clone BAC du 101-74 pour la région du locus  $R_1$  et la séquence du génome aideront à dessiner de nouveaux marqueurs pour mieux localiser ce locus.

De façon intéressante, le taux de recombinaison entre les marqueurs GCPM\_79 et INRA\_920-3 est 2,3 fois plus important chez le 73028-62 que chez le 101-74. La distance physique du génome de Nisqually-1 V2 entre ces deux marqueurs est de 170 Kb, le locus  $R_1$  se trouve dans une région avec un ratio de 90 Kb/cM. Avec 2140 individus de la famille 54, on peut donc espérer localiser  $R_1$  dans un intervalle de 0,1 cM correspondant à 9 Kb.



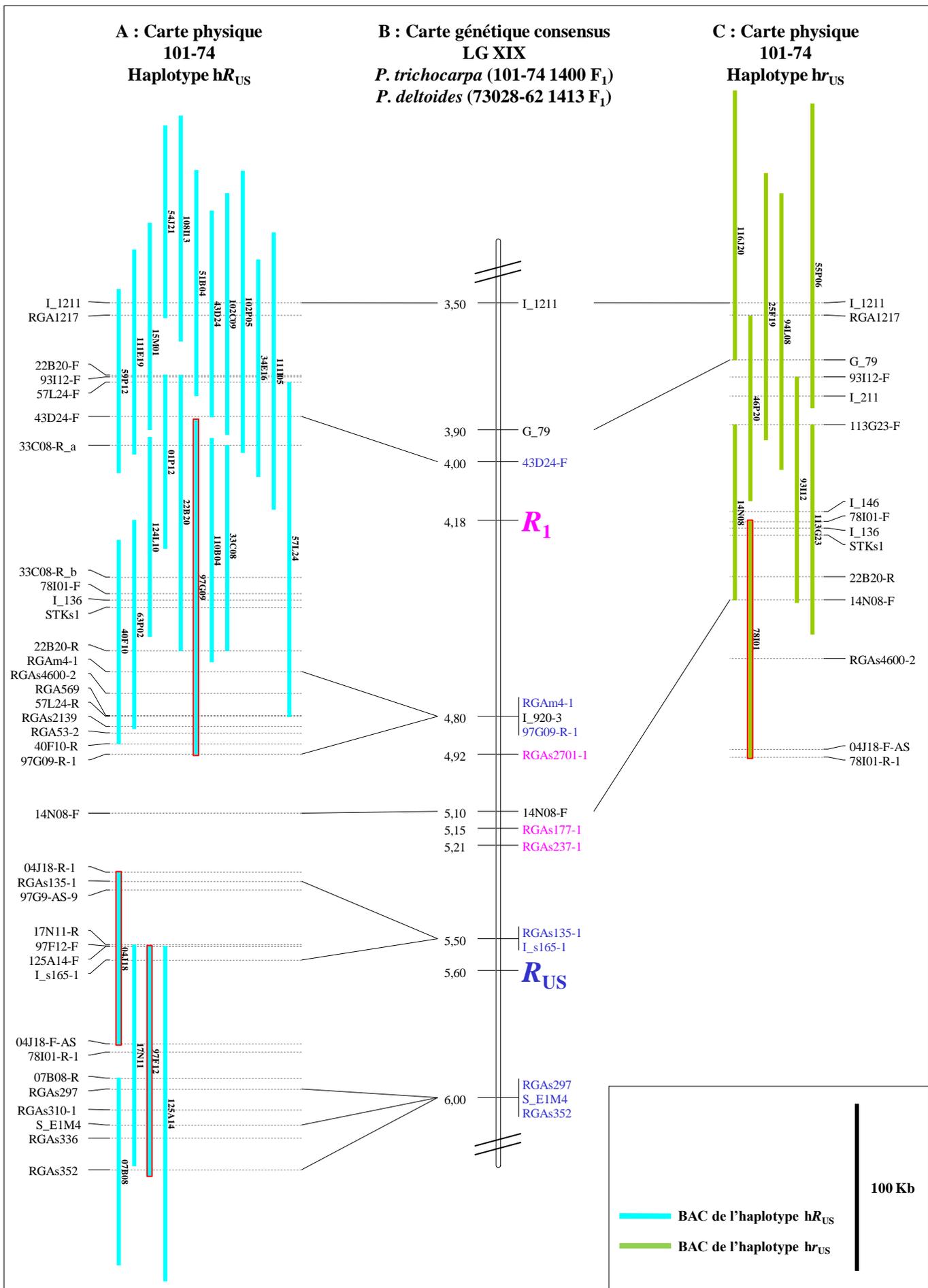
**Partie IV**  
**CLONAGE POSITIONNEL**  
**DES LOCI DE RESISTANCE**



# Partie IV

## CLONAGE POSITIONNEL DES LOCI DE RESISTANCE

<b>1</b>	<b>Le clonage positionnel du locus <math>R_{US}</math></b>	<b>91</b>
1.1	Séquençage des clones BAC	91
1.1.1	Annotation des séquences génomiques	92
1.1.2	Comparaison des séquences des clones BAC et du génome de référence	94
1.2	Cartographie génétique fine de $R_{US}$	96
1.2.1	Sélection des recombinants proches du locus $R_{US}$	96
1.2.2	Développement de nouveaux marqueurs génétiques	96
1.2.3	Carte génétique fine du locus $R_{US}$ dans la famille 54	97
1.2.4	L'allèle candidat $R_{US}$	97
1.3	Caractérisation de l'allèle candidat pour le locus $R_{US}$	98
1.3.1	Alignement des ESTs avec les deux allèles du gène candidat	99
1.3.2	Comparaison entre les deux allèles du locus	99
1.3.3	Comparaison entre les autres allèles des autres gènes	100
1.3.4	Analyse de l'expression du gène candidat pour $R_{US}$	100
<b>2</b>	<b>Vers le clonage positionnel du locus <math>R_1</math></b>	<b>101</b>
2.1	Cartographie génétique fine du locus $R_1$	102
2.1.1	Sélection des individus recombinants proches du locus $R_1$	102
2.1.2	Développement de marqueurs génétiques	102
2.1.3	Cartographie fine du locus $R_1$	103
2.2	Cartographie comparée	103



**Figure 46** : Alignement de la carte génétique consensus des parents de la famille 54 avec les cartes physiques du 101-74.

A : carte physique de l'haplotype  $hR_{US}$ , C : carte physique de l'haplotype  $hr_{US}$ . Les clones BAC envoyés au séquençage sont encadrés en rouge. B : carte génétique consensus des deux parents de la famille 54. Les marqueurs en bleu sont spécifiques du père 101-74, en violet ils sont spécifiques de la mère 73028-62 et en noir ils sont communs aux deux parents.

# 1 Le clonage positionnel du locus $R_{US}$

L'analyse de l'organisation d'une région très complexe de près de 4 Mb représentait une étape indispensable et obligatoire pour le clonage positionnel du locus  $R_{US}$ .

Pour mieux localiser le locus  $R_{US}$  il était important d'analyser les séquences de l'individu 101-74 qui porte ce locus. Je présente donc dans ce chapitre les principaux résultats des premières annotations des clones BAC que j'ai pu effectuées en dernière étape de ma thèse. Le but était de partir de ces séquences pour améliorer la cartographie fine du locus  $R_{US}$  et de proposer la séquence d'un gène candidat pour l'allèle  $R_{US}$ .

## 1.1 Séquençage des clones BAC

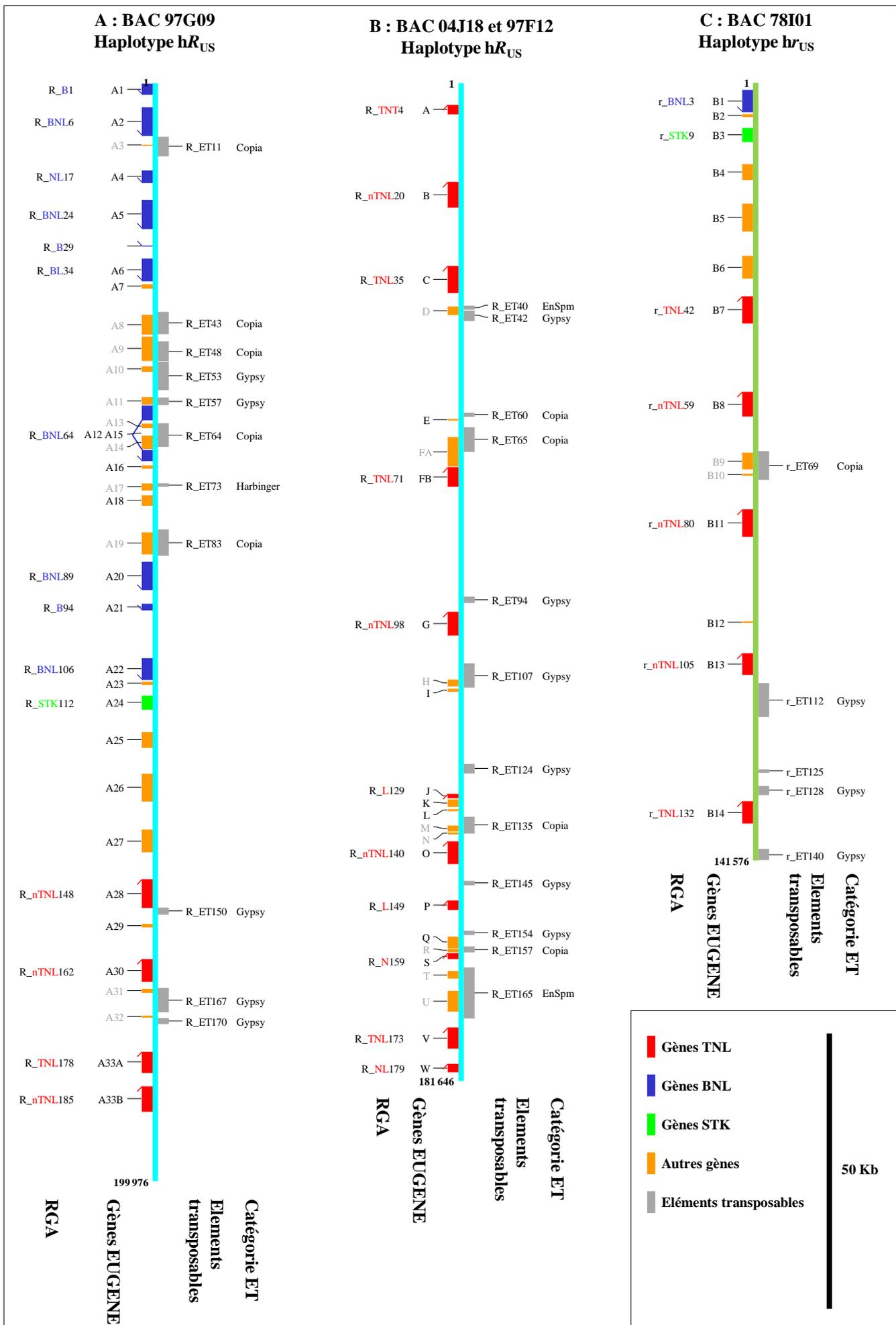
Les quatre clones BAC 04J18, 97F12, 78I01 et 97G09 ont été sélectionnés pour être séquencés (figure 46) :

- Les clones BAC chevauchant 04J18 et 97F12 de l'haplotype  $hR_{US}$  sont porteurs de l'allèle  $R_{US}$ . Le clone 97F12 porte les marqueurs INRA\_s165-1 et RGAs297 qui encadrent le locus  $R_{US}$ . Le clone 04J18 porte uniquement le marqueur INRA\_s165-1, mais il chevauche le clone 97F12 du côté où il n'y a qu'un seul recombinant qui permet de localiser le locus  $R_{US}$  (Cf. § IV-1.2.4.1). De plus, il porte certainement des gènes TNL du clade 1 proches de ceux qui sont dans l'intervalle où se trouve  $R_{US}$ .

- Le clone BAC 78I01, de l'haplotype  $hr_{US}$ , porte le marqueur génétique 14N08-F d'un côté du locus  $R_{US}$ , mais ne porte pas d'autres marqueurs qui permettent de connaître maintenant s'il est porteur de l'allèle  $r_{US}$ .

- Le clone BAC 97G09 de l'haplotype  $hR_{US}$  porte certainement d'autres gènes TNLs du clade 1 dont le locus  $R_{US}$  fait probablement partie. Il correspond aussi à la même zone que le locus  $R_1$  ; il peut donc être une aide pour le clonage positionnel de  $R_1$  par cartographie comparée.

Le fingerprint de ces BAC a été effectué avec l'enzyme *HindIII* pour vérification. Le séquençage des clones BAC a été effectué au Génoscope en lecture de type Sanger. Pour les deux premiers BAC (04J18 et 97F12), des sous-clones de 10 Kb ont été produits pour ce séquençage. L'assemblage de ces deux clones BAC chevauchants produit une séquence de 181 646 pb. La séquence du clone BAC 97G09 fait 199 976 pb et celle de 78I01 fait 141 576 pb.



**Figure 47** : Annotation des clones BAC du 101-74.

A : représentation de l'annotation des clones BAC chevauchant 04J18 et 97F12, B : du clone BAC 97G09, C : du clone BAC 78I01. Les gènes sont représentés à gauche de chaque séquence : les gènes de la catégorie des BNLs sont représentés en bleu, les gènes de la catégorie des TNLs sont représentés en rouge, les gènes STK sont représentés en vert et les autres gènes sont représentés en orange. Les éléments transposables sont représentés à droite des séquences en gris. Pour les gènes NBS-LRR une flèche indique l'orientation des gènes.

## 1.1.1 Annotation des séquences génomiques

### 1.1.1.1 Annotateur EUGENE

L'annotation automatique afin de détecter les CDSs sur la séquence des clones BAC a été faite par le programme EUGENE. Vingt trois gènes ont été prédits sur les clones BAC chevauchants 04J18 et 97F12, notés de A à W sur la figure 47. Neuf d'entre eux possèdent des CDS de moins de 1 000 pb. Trente trois gènes ont été prédits sur le clone BAC 97G09, notés d'A1 à A33 sur la figure 47. Quatorze gènes ont été prédits sur le clone BAC 78I01, notés de B1 à B33 sur la figure 47.

Ces gènes ont été par la suite expertisés par rapport à leur homologie avec des gènes de résistance connus et en parallèle l'annotation des éléments transposables a été effectuée (*voir ci-dessous*).

### 1.1.1.2 Les gènes de résistance

Les motifs MEME des gènes NBS-LRR donnés par Kolher *et al*, (2008) et Meyers *et al*, (2003) (annexe 2) ont été alignés avec le logiciel tBLASTn sur les séquences des clones BAC couvrant les deux haplotypes  $hR_{US}$  ou  $hr_{US}$ . Les différents alignements (des différents domaines) ont été raboutés suivant leur orientation et leur proximité, ce qui a permis de déduire 31 gènes NBS-LRR putatifs (figure 47). Ces gènes ont été nommés dans l'ordre, par R\_ ou r\_ suivant qu'ils appartiennent à l'haplotype  $hR_{US}$  ou  $hr_{US}$ , puis par les domaines qu'ils contiennent, et enfin par un chiffre qui correspond à leur localisation sur la séquence des clones BAC.

Pour les clones BAC chevauchant 04J18 et 97F12, les alignements avec les motifs MEME ont permis de caractériser 11 gènes putatifs du clade 1 des TNLs, qui sont tous dans la même orientation. Ces 11 gènes putatifs co-localisent parfaitement avec les gènes prédits par le logiciel EUGENE, à l'exception d'une différence, pour le gène R\_TNL71 qui est bien plus grand ; la fin du gène ne porte pas de motif MEME sur près de six Kb. Parmi ces 11 gènes putatifs, six seulement semblent complets, avec dans cette ordre, les domaines TIR, NBS et LRR, dont 3 ont aussi en premier le domaine NLS d'adressage au noyau. Pour quatre autres gènes putatifs, ils ne portent qu'un ou deux de ces domaines : deux gènes avec seulement le domaine LRR, un gène avec le domaine NBS et un gène avec les domaines NBS et LRR. Par l'alignement des gènes entre eux, ils ont pu être placés dans le clade 1 des TNL. Le tout dernier gène possède un enchaînement particulier avec, dans cet ordre, les domaines TIR et NBS puis de nouveau un domaine TIR.



Pour le clone BAC 78I01 de l'haplotype  $hr_{US}$ , six gènes putatifs NBS-LRR ont été détectés. Ils semblent tous complets avec les domaines NBS et LRR, précédés pour cinq d'entre eux du domaine TIR, et le dernier du domaine BED. Les gènes TNL sont dans la même orientation, en sens inverse du gène BNL. Ces six gènes putatifs co-localisent parfaitement avec les gènes prédits par le logiciel EUGENE. Il y a aussi un gène STK, qui a été déterminé par comparaison avec l'annotation automatique du génome de Nisqually-1.

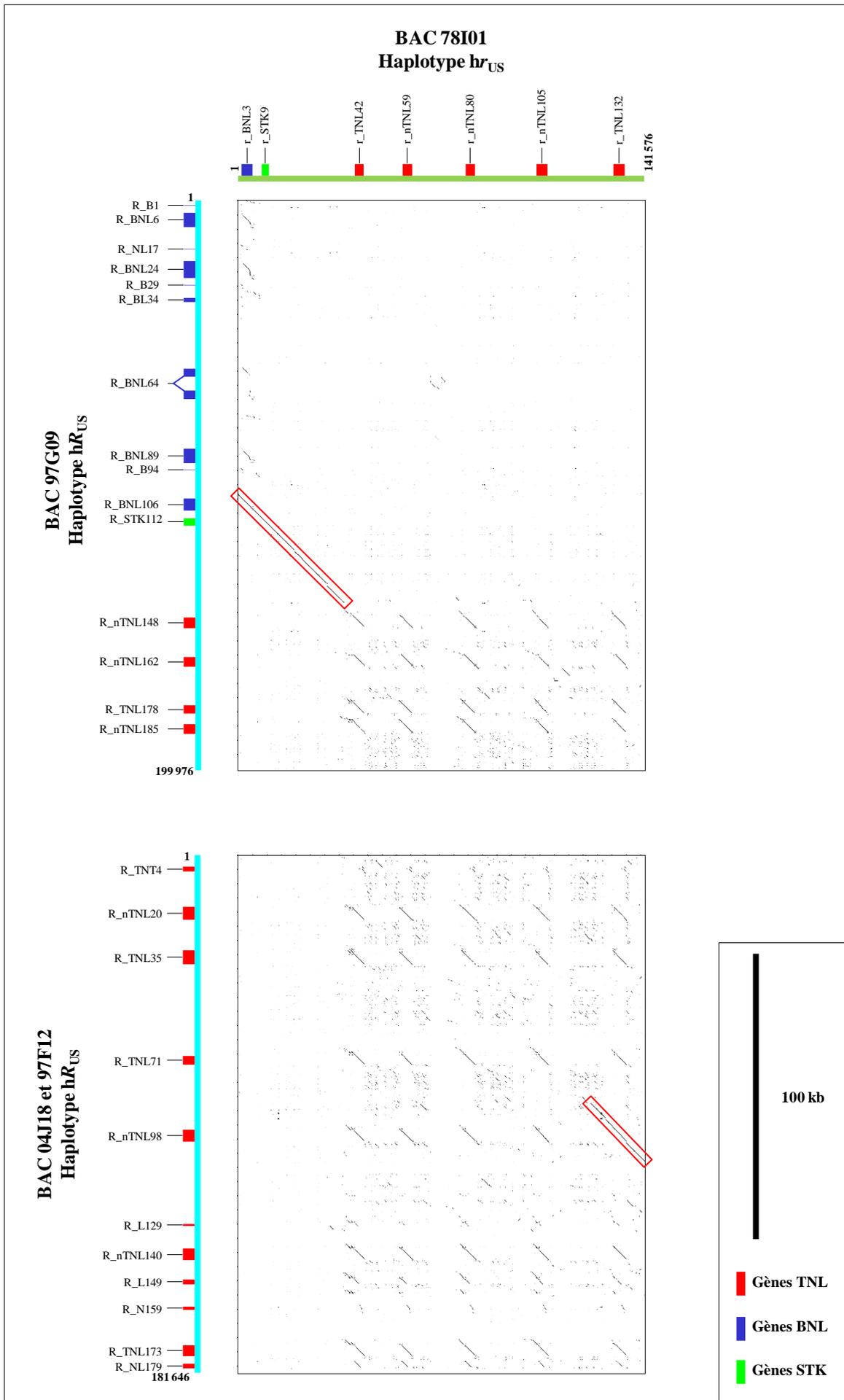
Pour le clone BAC 97G09, 14 gènes putatifs possèdent des domaines des gènes NBS-LRR. Parmi ces 14 gènes putatifs, neuf seulement semblent complets avec les domaines NBS et LRR, précédés pour cinq d'entre eux par le domaine BED et 4 par le domaine TIR. Pour les cinq autres gènes putatifs, trois d'entre eux ne possèdent qu'un domaine BED, un gène porte les domaines BED et LRR et le dernier possède les domaines NBS et LRR : celui-ci a pu être placé par alignement avec les autres gènes dans la catégorie des BNL. Les gènes TNL entre eux sont dans la même orientation, de même pour les gènes BNL entre eux, par contre les deux catégories de gènes ont des orientations opposées. Treize gènes putatifs NBS-LRR co-localisent parfaitement avec les gènes prédits par le logiciel EUGENE, et un gène B ne correspond pas à une annotation du logiciel EUGENE. Il y a aussi un gène STK qui a été déterminé par comparaison avec le génome de Nisqually-1.

### **1.1.1.3 Eléments transposables**

La détection et l'annotation des éléments transposables (ETs) ont été effectuées par alignement des séquences des clones BAC sur le site de GIRI (<http://www.girinst.org>) en utilisant le programme CENSOR et la base de données d'ETs RepBase (release 15.10, octobre 2010). Les alignements représentant un même TE sur la séquence ont ensuite été regroupés lorsqu'ils avaient des coordonnées et une orientation compatibles : la figure 47 présente ici un descriptif non détaillé des éléments transposables présents sur la séquence des clones BAC, mais il ne s'agit pas d'une annotation exhaustive.

Pour les clones BAC 04J18 et 97F12, 12 ETs ont été détectés ; il s'agit des éléments transposables de type EnSpm, Copia et Gypsy. Les analyses montrent qu'une partie du gène F correspond à un ET (nommé FA) et l'autre partie possède les domaines TIR, NBS et LRR (nommé FB). Tous les gènes prédits par EUGENE et qui ne sont pas des gènes NBS-LRR, co-localisent avec les ETs (figure 47). Ils représentent en effet des morceaux de CDS des ETs telle que la transposase.

Pour le clone BAC 97G09, 11 ETs ont été détectés ; il s'agit d'ETs de type Copia, Gypsy et Harbinger. Un ET de type Copia s'est inséré dans le gène R\_BNL64.



**Figure 48** : Alignement des séquences des clones BAC par Dot Plot.

Les clones BAC de l'haplotype  $hr_{US}$  sont en ordonnée et le clone BAC de l'haplotype  $hr_{US}$  est en abscisse. Les segments chromosomiques orthologues sont encadrés en rouge. Sur les séquences des clones BAC sont représentés les gènes de résistance : en rouge les gènes TNL, en bleu les gènes BNL et en vert les gènes STK.

Pour le clone BAC 78I01, quatre ETs ont été détectés ; il s'agit d'ETs de type Gypsy et Copia.

Dans l'ensemble, les ETs sont pour la plupart des éléments à LTR. Dans la plupart des cas, les LTR ne sont pas détectables par l'alignement de la séquence sur elle-même. Les résultats des alignements des clones BAC sur la base d'ETs montrent des alignements partiels ; ce sont probablement des ETs incomplets mais cela peut être aussi des ETs différents de ceux de RepBase.

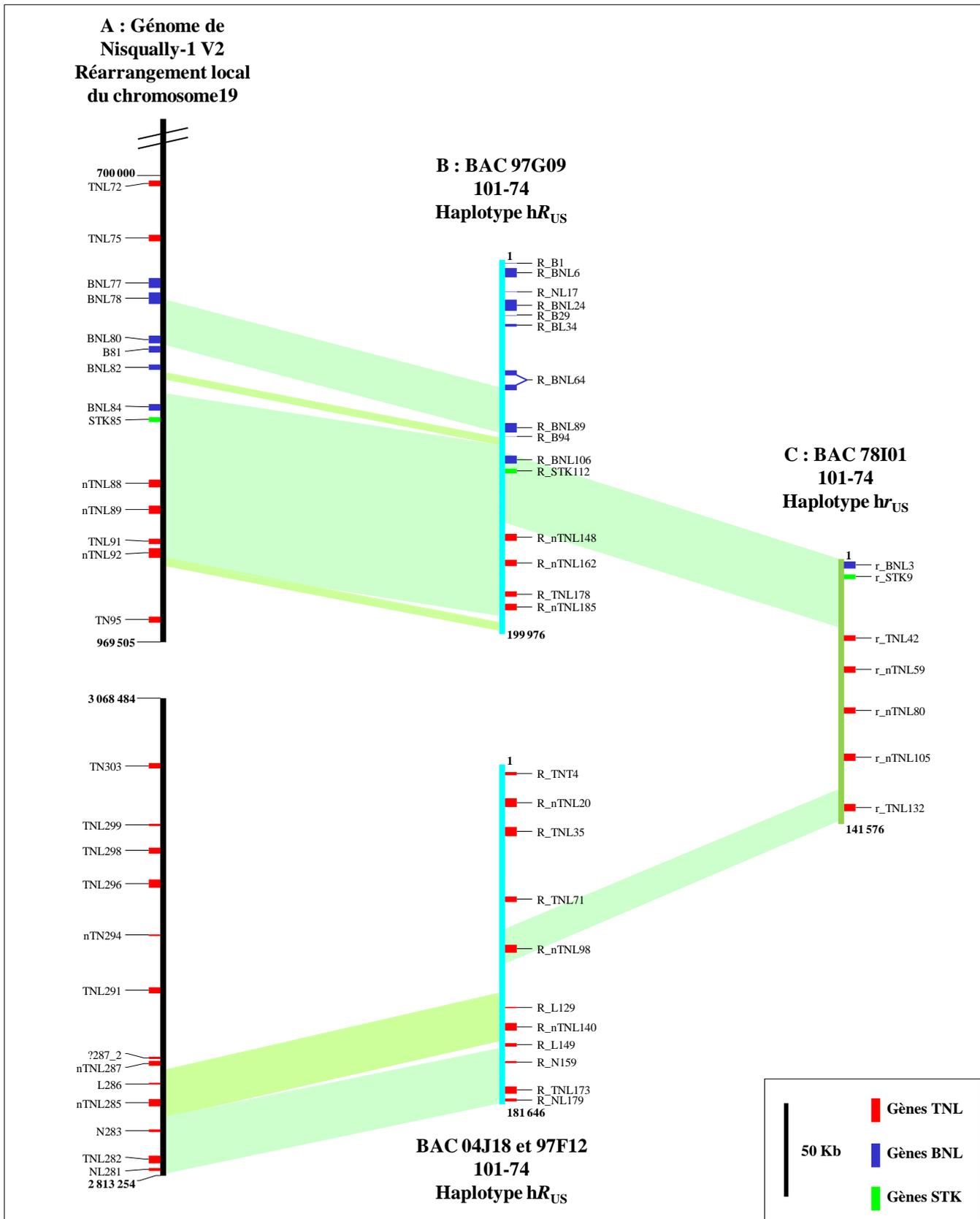
## **1.1.2 Comparaison des séquences des clones BAC et du génome de référence**

### **1.1.2.1 Orthologie entre les deux haplotypes du 101-74**

Afin de déterminer les segments chromosomiques orthologues entre les deux haplotypes du 101-74, des alignements par Dot Blot ont été produits entre les séquences des clones BAC provenant des deux haplotypes différents (figure 48). Entre les clones BAC 97G09 et 78I01, il y a un segment chromosomique orthologue d'environ 37 Kb (encadré en rouge dans la figure 48), et entre les clones 04J18-9F12 et 78I01 il y a un segment chromosomique orthologue d'environ 18 Kb. Entre ces deux segments, il n'est pas possible de déterminer de l'orthologie entre les séquences des deux haplotypes. Ceci rend impossible la détermination d'orthologie pour un gène dans la plupart des cas.

Entre les clones BAC 97G09 et 78I01, il est possible de déterminer clairement les deux formes alléliques pour deux gènes : R\_BNL106 avec r\_BNL3, R\_STK112 avec r\_STK9. Dans le cluster de TNL, il n'est pas possible de déterminer une orthologie stricte entre les gènes des deux haplotypes. Chacun de ces gènes s'aligne avec tous les autres membres du cluster du deuxième haplotype sans que les identités et similarités des séquences puissent privilégier une orthologie. Les régions inter-géniques ne s'alignent quasiment pas, aucune paire d'allèles correspondante à un gène n'ayant pu être définie.

Entre les deux clones BAC chevauchants 04J18 et 97F12 et le clone BAC 78I01, les mêmes observations sont faites pour les gènes TNL à l'exception du couple d'allèles R\_nTNL98 et r\_TNL132. Mais ces deux allèles du même gène ne portent pas les mêmes domaines, l'allèle porté par l'haplotype hR<sub>US</sub> possède en plus le domaine NLS d'adressage au noyau. De plus, ces allèles s'alignent avec tous les autres gènes de l'autre haplotype. Il n'y a donc que le bon alignement des zones inter-géniques entourant ces deux allèles qui permettent de déterminer l'orthologie entre ces gènes.



**Figure 49** : Alignement entre les séquences des BAC du 101-74 et le génome de Nisqually-1 V2.

A : La séquence du génome de Nisqually-1. B : les séquences des clones BAC de l'haplotype  $hR_{US}$  du 101-74. C : la séquence du clone BAC de l'haplotype  $hr_{US}$ . Les alignements entre les différentes séquences représentent les segments chromosomiques définis comme orthologues. Sur les séquences des clones BAC sont représentés les gènes de résistance : en rouge les gènes TNL, en bleu les gènes BNL et en vert les gènes STK.

### **1.1.2.2 Orthologie entre l'haplotype 101-74 et Nisqually-1**

De même, les alignements par Dot Plot ont été effectués entre l'haplotype  $hR_{US}$  du 101-74 et le génome de Nisqually-1 V2 pour caractériser les segments chromosomiques orthologiques entre ces séquences (figure 49). Même s'il y a plus de régions qui s'alignent, il reste beaucoup de régions qui ne sont pas orthologiques. Une grande partie correspond au cœur du cluster de gènes TNL du clade 1 et une seconde partie correspond au cluster de gènes BNL, qui est plus étendue pour l'haplotype  $hR_{US}$ .

### **1.1.2.3 Différence de taille**

Ces alignements révèlent des différences importantes de taille entre les deux haplotypes du 101-74 et le génome de Nisqually-1 V2 (figure 49). Ils confirment l'expansion de taille de l'haplotype  $hR_{US}$  par rapport à l'haplotype  $hr_{US}$  remarquée avec la carte physique (Cf. § III-2.5.5.3). Le segment chromosomique non orthologue entre les deux haplotypes fait 88 Kb pour l'haplotype  $hr_{US}$  avec 4 gènes TNL et, au minimum, 147 Kb pour l'haplotype  $hR_{US}$  avec 8 gènes TNL. Cette différence est au cœur du cluster de gènes TNL du clade 1. Les segments non orthologiques du cluster de TNL semblent plus petits chez l'haplotype  $hR_{US}$  que chez Nisqually-1 ; elle fait au minimum 124 Kb pour l'haplotype  $hR_{US}$  avec 4 gènes TNL et au minimum 239 Kb pour l'individu Nisqually-1 avec 8 gènes TNL. Un autre segment non orthologue entre l'haplotype  $hR_{US}$  et Nisqually-1 se situe au cœur du cluster de BNL, qui semble plus étendu en nombre de gènes mais aussi en ETs dans l'haplotype  $hR_{US}$  que pour l'individu Nisqually-1.

Néanmoins, il est difficile de conclure sur ces différences de taille basées sur les séquences disponibles de Nisqually-1 V2 et des clones BAC de  $hR_{US}$ . En effet, le manque de la banque BAC donne un trou de séquence dans l'haplotype  $hR_{US}$  dont la taille est difficilement estimable. De même, le génome de Nisqually-1 est mal assemblé ; il a donc été proposé un nouvel ordonnancement de différents contigs dans cette zone. La taille des trous entre ces différents contigs réassemblés n'a pas pu être estimée. De plus, comme la séquence du génome n'est pas un haplotype mais le mélange (chimère) des 2 haplotypes de l'individu Nisqually-1, il est donc possible que les différents contigs du génome proviennent des différents haplotypes qui normalement doivent se chevaucher ; en effet, comme nous avons pu le voir entre les deux haplotypes du 101-74, certaines séquences d'haplotypes différents correspondant à une même zone ne s'alignent pas, il n'est donc pas possible de le savoir pour Nisqually-1 sans un travail approfondi sur ses deux haplotypes séparément.



## 1.2 Cartographie génétique fine de $R_{US}$

### 1.2.1 Sélection des recombinants proches du locus $R_{US}$

Le séquençage du segment chromosomique portant  $R_{US}$  étant réalisé, une cartographie génétique plus fine a été engagée en exploitant ces séquences d'une part et l'existence de recombinants entre les marqueurs génétiques les plus proches d'autre part. Seuls les 26 individus recombinants entre les marqueurs 97G09-R-1 et RGAs352 pour l'haplotype provenant du père 101-74, intervalle dans lequel se trouve  $R_{US}$ , ont été retenus pour la suite de l'étude.

#### 1.2.1.1 Re-phénotypage des recombinants

Bien que les tests pathologiques soient faits dans des conditions contrôlées et normalisées, le phénotype de 25 individus, parmi les 26 recombinants, a été vérifié par la réalisation d'une expérience supplémentaire : ceci dans l'objectif de s'affranchir du biais entre les différentes expériences, ce qui pouvait fausser l'évaluation phénotypique et donc le classement des individus. Cette vérification a permis aussi de confirmer la localisation du locus  $R_{US}$ .

### 1.2.2 Développement de nouveaux marqueurs génétiques

Le développement de nouveaux marqueurs génétiques basés sur la PCR, a été réalisé à partir des séquences des clones BAC chevauchant 04J18 et 97F12 de l'haplotype  $hR_{US}$  du parent 101-74. Une première approche a été menée sur les 11 gènes NBS-LRR en utilisant les amorces consensus dessinées dans les domaines TIR et LRR et une amorce spécifique localisée dans les zones inter-géniques. Quinze nouvelles amorces spécifiques ont été développées (ce qui forme autant de couples d'amorces), sept pour l'amorce consensus sur le domaine TIR et huit pour l'amorce consensus sur le domaine LRR. La forte homologie des gènes NBS-LRR entre eux dans notre zone d'intérêt, a fait qu'il a été très difficile de définir des amorces spécifiques ; l'alignement des gènes entre eux a donc été très utile pour définir les zones spécifiques. Pour compléter, quatre SSR (dont un développé sur Nisqually-1) et cinq STS inter-géniques ont aussi été développés sur la séquence des clones BAC (annexe 5).

Ces nouveaux couples d'amorces ont été testés sur les parents et 6 descendants de la famille 54. Pour deux marqueurs qui ne ségrègent pas, le séquençage de produit d'amplification PCR a permis de détecter des SNPs et de les convertir en trois marqueurs génétiques supplémentaires (par rapport aux SNP), dont deux correspondent à un même locus, chacun marquant un des deux allèles. Au total, 20 nouveaux marqueurs génétiques proches du locus  $R_{US}$



ont été développés (annexe 5) dont 15 ségrègent 1:1 pour le 101-74 et cinq ségrègent 1:3 pour les deux parents de la famille 54 (Cf. figure 33, cas 7, page III-73).

### **1.2.3 Carte génétique fine du locus $R_{US}$ dans la famille 54**

#### **1.2.3.1 Génotypage de la famille 54**

Les ADNg de 24 individus ont été ré-extraits parmi les 26 individus recombinants, ce qui a permis de vérifier leur génotype tout en assurant la quantité suffisante pour la suite des expériences.

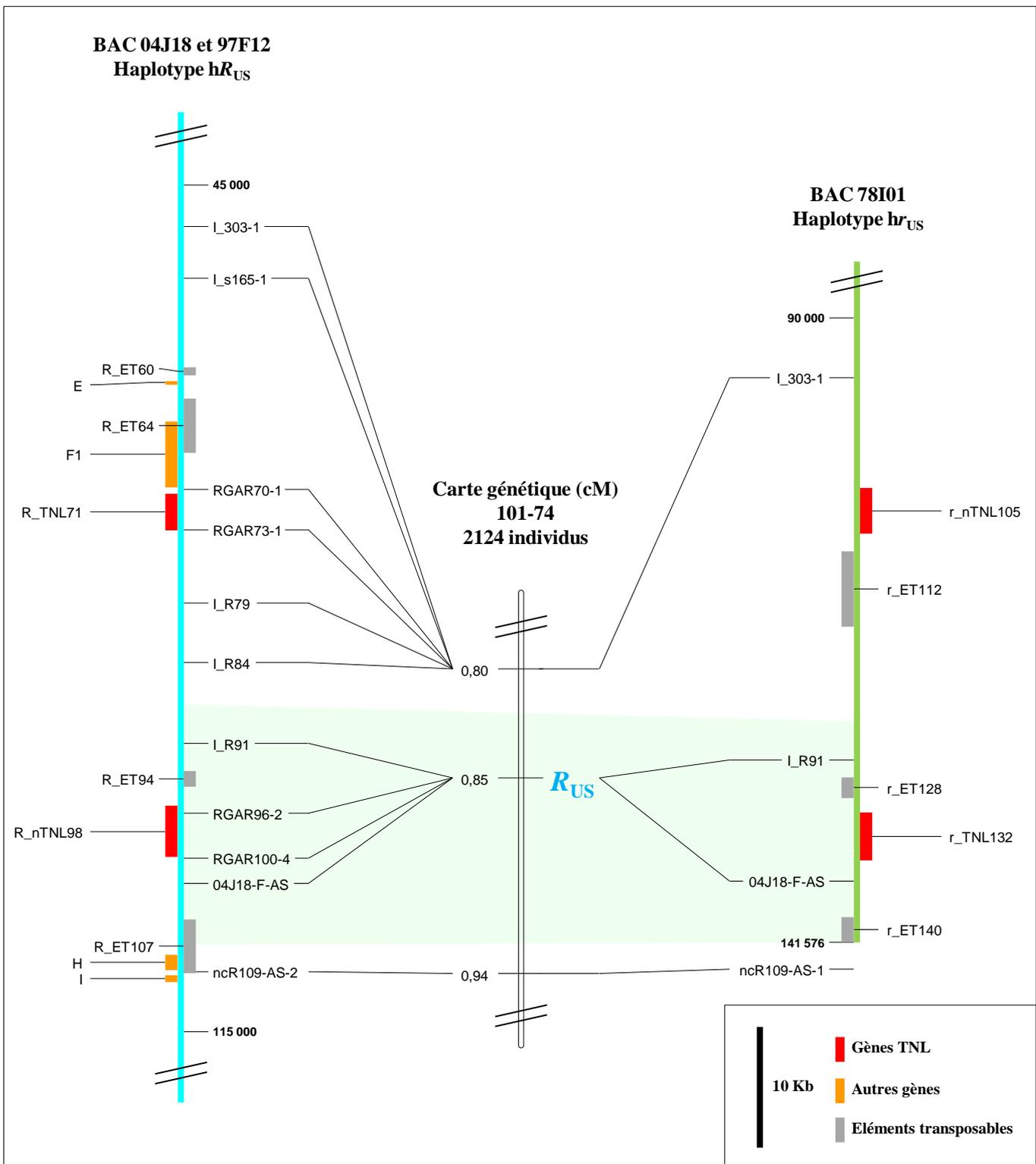
Parmi les 21 nouveaux marqueurs génétiques, 15 ont été utilisés pour génotyper les individus porteurs d'une recombinaison proche de  $R_{US}$ , ce qui forme 13 nouveaux locus cartographiés génétiquement (en gras dans la figure 50) ; deux couples de marqueur proviennent d'un même locus (RGAR96-2 avec RGAR96-4, ncR109-AS-1 avec ncR109-AS-2), ils ont montré comme attendu une co-ségrégation complète (figure 50). Treize marqueurs génétiques préalablement utilisés dans la construction des cartes génétiques locales ont été réutilisés pour réaliser une carte génétique fine et vérifier ces résultats.

Par ailleurs, j'ai également confirmé la localisation physique des marqueurs par e-PCR ou BLASTn sur les séquences des clones BAC de l'haplotype  $hR_{US}$  en regardant la correspondance entre la taille d'amplification attendue et celle réellement obtenue par PCR sur l'ADNg.

#### **1.2.4 L'allèle candidat $R_{US}$**

Après la première étape de cartographie locale, les marqueurs les plus proches délimitaient l'allèle  $R_{US}$  à un intervalle de 0,42 cM et 77 Kb de l'haplotype  $hR_{US}$ , où deux gènes TNL sont prédits. Les nouveaux marqueurs ont permis de mieux délimiter les intervalles génétiques et physiques de l'allèle  $R_{US}$ . Celui-ci est encadré par deux marqueurs génétiques (INRA\_R84, ncR109-AS-2) espacés de 0,14 cM en distance génétique, ce qui correspond à 26 Kb en distance physique. Le locus  $R_{US}$  co-localise avec quatre marqueurs génétiques (INRA\_R91, RGAR96-2, RGAR100-2, 04J18-F-AS). Dans cet intervalle il y a maintenant un seul gène TNL qui est le candidat pour l'allèle  $R_{US}$ . Ce gène identifié,  $R\_nTNL98$  porte aussi en 5' un domaine NLS d'adressage au compartiment nucléaire.

Ce gène est dans la partie chevauchante entre les deux clones BAC 04J18 et 97F12, il a donc été séquencé deux fois indépendamment. Dans la partie chevauchante de ces deux BAC, il y a 59 Kb, ces séquences sont identiques avec deux courtes exceptions de quelques bases dans



**Figure 51** : Alignement de la carte génétique fine de  $R_{US}$  avec les BAC des haplotypes  $hR_{US}$  et  $hr_{US}$  autour du locus.

des régions de faibles complexités inter-géniques. On peut donc considérer que la séquence de ce gène est particulièrement sûre.

Pour le reste de la séquence dans cet intervalle de l'allèle  $R_{US}$ , il y a en amont du gène candidat et en bordure de l'intervalle, un ET de type Gypsy, nommé  $R\_ET107$  avec deux gènes H et I. En aval du gène candidat, il y a un autre ET de type Gypsy, nommé  $R\_ET94$  (figure 51).

#### 1.2.4.1 L'individu recombinant

L'intervalle déterminant le gène  $R\_nTNL98$  comme candidat pour  $R_{US}$ , est basé sur une seule recombinaison observée d'un côté du gène  $R_{US}$  (figure 50), porté par l'individu 661303763. En effet, sans cet individu recombinant, l'intervalle du gène  $R_{US}$  s'élargirait à trois gènes supplémentaires et couvrirait une région génomique non couverte par la banque BAC et donc sans séquence disponible. Pour confirmation, cet individu a été génotypé sur 3 extractions indépendantes d'ADNg. Il a également été phénotypé 2 fois pour chaque souche 98AG69 et 93ID6. Comme il est de phénotype  $r1$ , le phénotypage des deux souches a permis de déterminer son phénotype  $R_{US}$ .

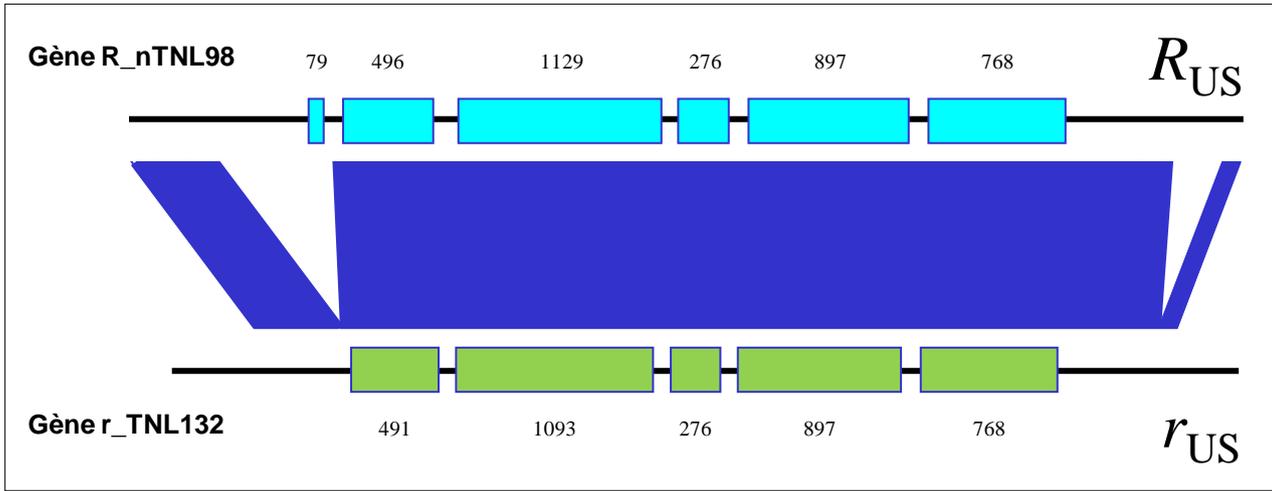
#### 1.2.4.2 L'allèle $r_{US}$

Parmi les 4 marqueurs génétiques co-dominants proches du locus  $R_{US}$ , deux (INRA\_303-1 et ncR109) délimitent un intervalle de 50 Kb environ sur l'haplotype  $hr_{US}$  où deux gènes TNL ( $r\_nTNL105$  et  $r\_TNL132$ ) ont été identifiés. Cet intervalle reste imprécis car l'un des marqueurs flanquants (ncR109-AS-1) n'est pas présent sur le clone BAC séquencé (78I01).

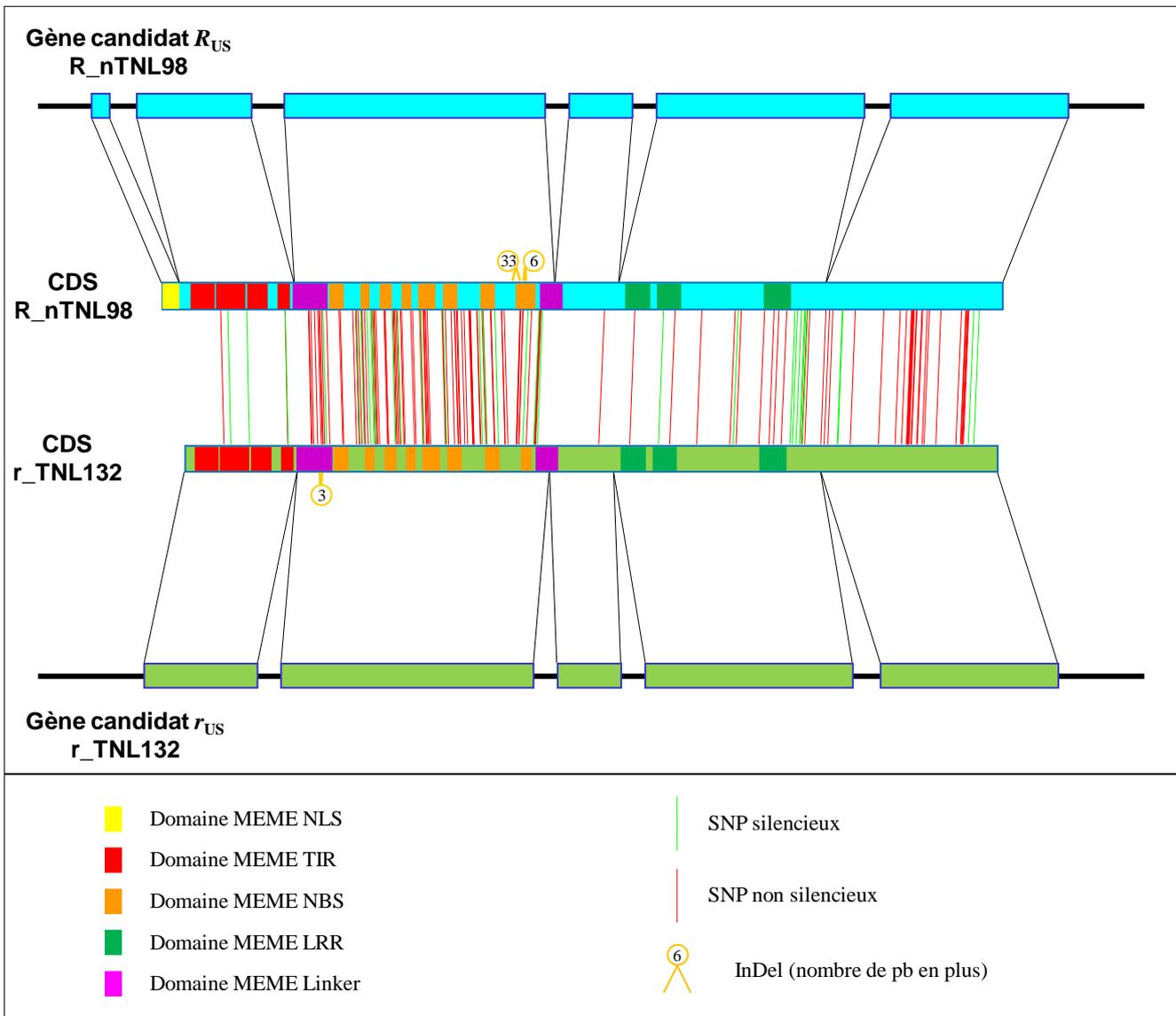
Néanmoins, le gène  $r\_TNL132$  est un bon candidat pour l'allèle  $r_{US}$ . En effet, on peut considérer qu'il est l'orthologue au gène  $R\_nTNL98$  grâce à la colinéarité des séquences des deux haplotypes autour des deux allèles (Cf. § IV-1.1.2.1, figure 51). Autour du locus  $R_{US}$ , les ETs sont conservés entre les deux haplotypes.

### 1.3 Caractérisation de l'allèle candidat pour le locus $R_{US}$

Dans la zone génomique du locus  $R_{US}$ , une orthologie est faite entre l'allèle  $R\_nTNL98$  et l'allèle  $r\_TNL132$  (figure 51). L'alignement représente 18 Kb (Cf. figure 48, page IV-94), mais il n'y a que 82 % de bases identiques et 55 trous dans l'alignement. Il reste donc encore beaucoup d'hétérozygotie, que ce soit dans le gène ou à l'extérieur, y compris dans les régions promotrices du gène.



**Figure 52 :** Structure des gènes candidats pour  $R_{US}$  et  $r_{US}$ .  
 Les cadres sur les séquences montrent les exons dont la taille est indiquée. Les alignements entre les deux séquences montrent les parties homologues.



**Figure 53 :** Comparaison des CDSs des gènes candidats pour  $R_{US}$  et  $r_{US}$ .

### 1.3.1 Alignement des ESTs avec les deux allèles du gène candidat

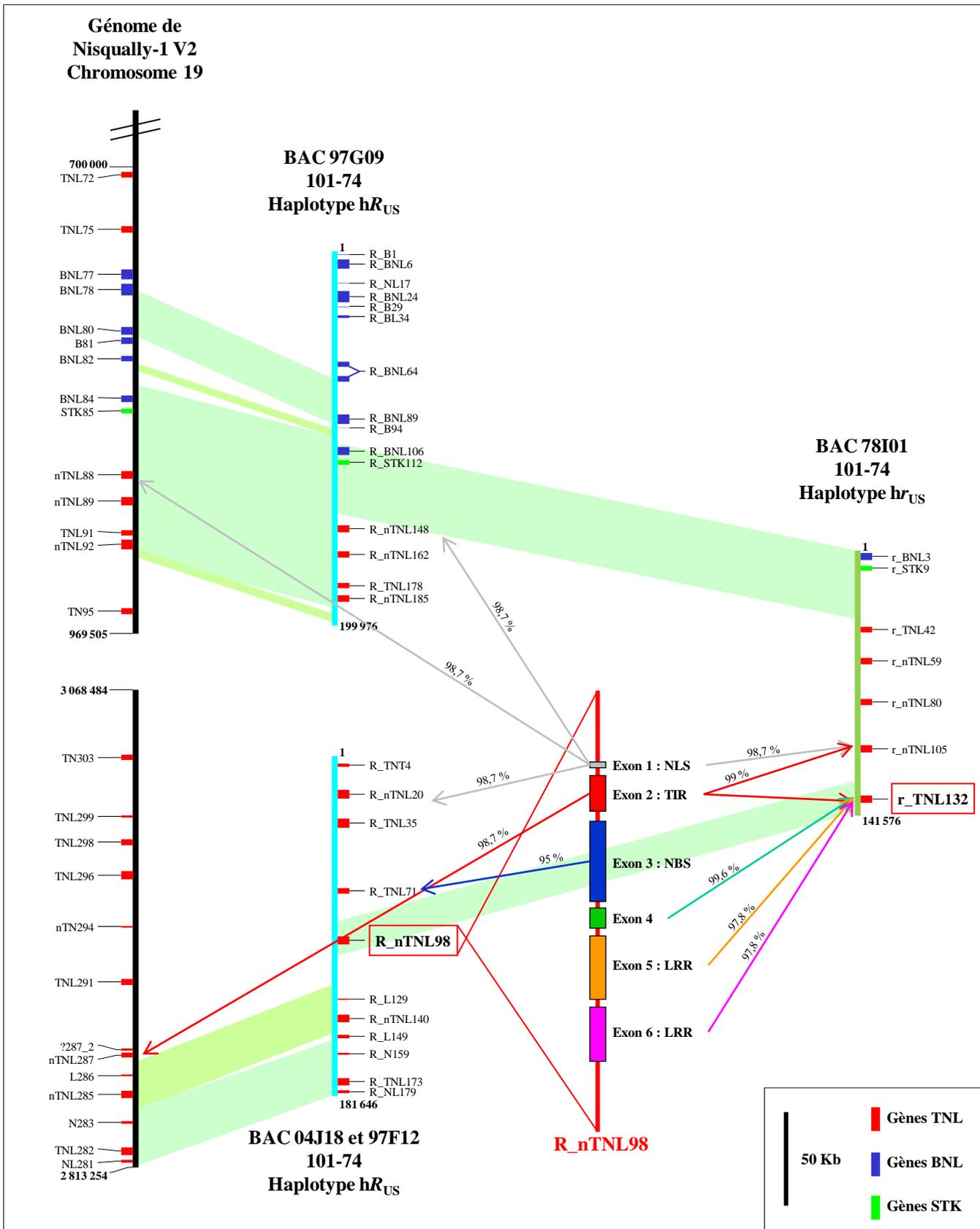
Plus de 400 000 EST sont maintenant disponibles chez le peuplier. La séquence du gène R\_nTNL98 avec ses introns et 1 000 pb situées avant et après le gène, a été utilisée pour rechercher les ESTs. Trois cent soixante quatre EST ont été identifiés en utilisant le logiciel BLASTn sur la base de données EST peuplier du NCBI (e-value inférieure à 1). Ces ESTs ont ensuite été alignés avec la séquence du génome pour vérifier qu'ils provenaient bien du clade 1 des gènes TNL, en regardant la localisation des meilleurs résultats d'alignement. En effet, avec la redondance importante des gènes NBS-LRR, beaucoup d'ESTs provenaient d'autres clades. Au total, 87 ESTs ont été identifiés dans ce clade de gènes et ont été alignés avec les gènes candidat  $R_{US}$  et  $r_{US}$  à l'aide du logiciel MEGA4. Cela a permis de confirmer l'annotation faite par le logiciel EUGENE pour les CDS du gène R\_nTNL98 qui rassemblent 3 642 pb en 6 exons. Les CDS du gène r\_TNL132 font 3 522 pb en 5 exons (figure 52).

### 1.3.2 Comparaison entre les deux allèles du locus

La comparaison globale entre les allèles R\_nTNL98 et r\_TNL132 (figure 52) montre une différence majeure entre les deux allèles : la présence d'un exon supplémentaire au début du gène dans l'allèle  $R_{US}$ . Cet exon de 79 pb code pour le domaine NLS d'adressage au noyau.

Dans la partie commune aux deux CDSs, il y a 3 InDels : un de 3 pb supplémentaires dans le gène r\_TNL132, un de 33 pb et un de 6 pb supplémentaires dans le gène R\_nTNL98. Pour le reste, on trouve aussi 153 SNPs (96,6 % d'identité), dont 122 changent la séquence protéique. La protéine du gène R\_nTNL98 est constituée de 1 214 acides aminés et la protéine du gène r\_TNL132 est constituée de 1 174 acides aminés. Dans la partie commune aux deux protéines, il y a 122 acides aminés différents sur 1173 acides aminés. Il y a donc que 87 % d'identité entre les deux protéines, sans compter les InDels.

Les motifs MEME issus de l'analyse faite par Kolher *et al.*, (2008) et l'étude de Meyers *et al.*, (2003) regroupés dans l'annexe 2, ont été représentés dans la figure 53. Les exons du domaine TIR sont très proches avec seulement 5 SNPs dont 3 silencieux sur 496 pb (98,9 % d'identité). L'exon codant le domaine NBS est le plus hétérozygote, avec les 3 InDels et 96 SNPs dont 16 silencieux pour 1 129 pb (91,2 % d'identité). Les trois derniers exons avec les domaines LRR sont assez bien conservés, avec 54 SNPs dont 12 silencieux pour 1 941 pb (97,2 % d'homologie).



**Figure 54** : Alignement des exons du gène candidat avec les autres gènes du cluster de TNL.

### 1.3.3 Comparaison entre les autres allèles des autres gènes

#### 1.3.3.1 Existe-t-il une autre forme allélique à $R_{US}$ chez Nisqually-1 ?

Dans la séquence du génome, il n'est pas possible de définir un allèle de ce locus car aucun segment chromosomique orthologue n'est définissable pour la région du gène  $R\_nTNL98$ , quelle que soit la version de l'assemblage (figure 49, page IV-95). Pourtant des amplifications PCR ont été produites avec des amorces dessinées sur le gène  $R\_nTNL98$  sur l'ADNg de Nisqually-1 : il est possible qu'un allèle  $r_{US}$  soit présent chez cet individu mais qu'il n'ait probablement pas été couvert dans la séquence du génome.

#### 1.3.3.2 $R_{US}$ parmi le clade de gènes de TNL

Le premier exon du gène  $R\_TNL98$ , codant le domaine NLS, a une identité de 98,7% avec 4 autres gènes (figure 54). Cet exon a une très forte redondance puisqu'il y a aussi 11 gènes qui ont plus de 95% d'identité avec ce premier exon, dont trois ne sont pas localisés dans ce cluster.

Le second exon du gène, codant le domaine TIR, a une identité de 99 % avec deux gènes d'autres haplotypes : le gène candidat pour  $r_{US}$  et son voisin. Il a aussi deux autres gènes qui ont un exon très proche, 98,8 et 98,6 % d'identité. C'est l'exon le plus redondant ; il a 28 gènes qui ont plus de 95 % d'identité avec cet exon dont 3 gènes qui ne font pas partie de la région génomique.

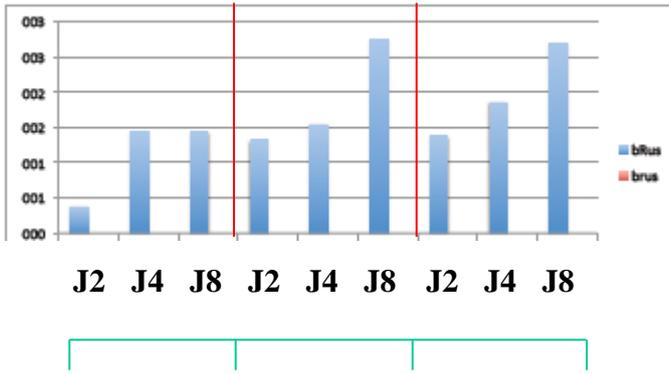
Le troisième exon, codant le domaine NBS, est très différent des autres gènes du clade, notamment de l'allèle candidat  $r_{US}$ . Le seul gène qui possède une homologie de plus de 95 % est un gène porté par l'haplotype  $hR_{US}$ , le plus proche voisin du gène  $R\_nTNL98$  en direction du télomère. Il y a 9 gènes qui possèdent une identité supérieure au gène  $r\_TNL132$  pour cet exon, et qui proviennent du génome de Nisqually-1 et des deux haplotypes du 101-74.

Les trois derniers exons ont comme meilleur alignement le gène  $r\_TNL133$ .

### 1.3.4 Analyse de l'expression du gène candidat pour $R_{US}$

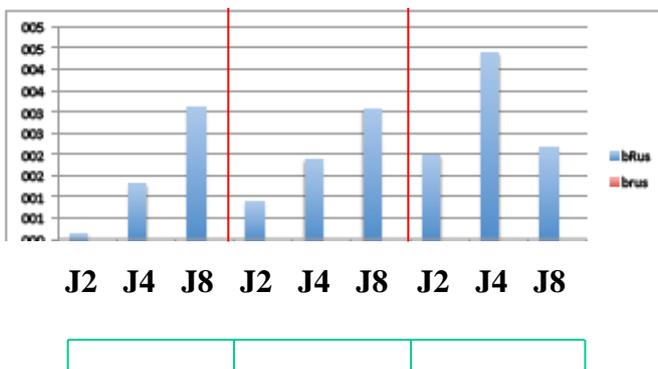
Pour étudier l'expression du gène candidat, il était nécessaire de développer des couples d'amorces dans le gène. La redondance très importante du clade 1 des TNL, dont le gène  $R\_nTNL98$  fait partie, représentait une difficulté majeure pour l'obtention de couples spécifiques de ce gène. Pour y arriver, Frédérique Bitton, ingénieur dans l'équipe, a utilisé l'alignement entre les gènes de ce clade. Il était intéressant de regarder tous les gènes, même ceux qui provenaient d'autres individus : ceux de Nisqually-1 et ceux portés par le clone BAC à proximité du locus

### Premier marqueur



3 répétitions

### Second marqueur



3 répétitions

INRA Nancy IaM

Figure 55 : Expression de l'allèle candidat par RTq-PCR.

*MER* (Lescot *et al.*, 2004). En effet, les séquences disponibles du 101-74 ne contiennent pas tous les gènes de ce clade, différents trous de la banque et certains gènes de ce clade ne sont pas présents dans ce cluster. De plus, il est nécessaire que les amorces ne fonctionnent pas chez le 73028-62 afin de pouvoir voir spécifiquement l'expression de l'allèle  $R_{US}$  dans la famille 54. Les amorces ont été testées sur l'ADNg des deux parents, les individus recombinants proches de  $R_{US}$  de la famille 54, des clones BAC, et des sous clones BAC (Cf. § IV-1.1) pour sélectionner des marqueurs robustes et spécifiques de l'allèle.

Quatre bulks de 10 individus de phénotype  $R_{US}$  et quatre bulks de 10 individus de phénotype  $r_{US}$  ont été constitués à partir d'ADNc. Ils ont été créés à 2, 4 et 8 jours après l'inoculation de *Mlp* et sans inoculation. Deux marqueurs robustes (142 pb et 158 pb sur le dernier exon du gène) ont été utilisés sur des bulks d'ADNc. Une première RT-PCR a été réalisée au laboratoire, qui montrait une amplification de taille attendue chez le bulk  $R_{US}$  mais absent du bulk  $r_{US}$  au temps de 4 jours après inoculation.

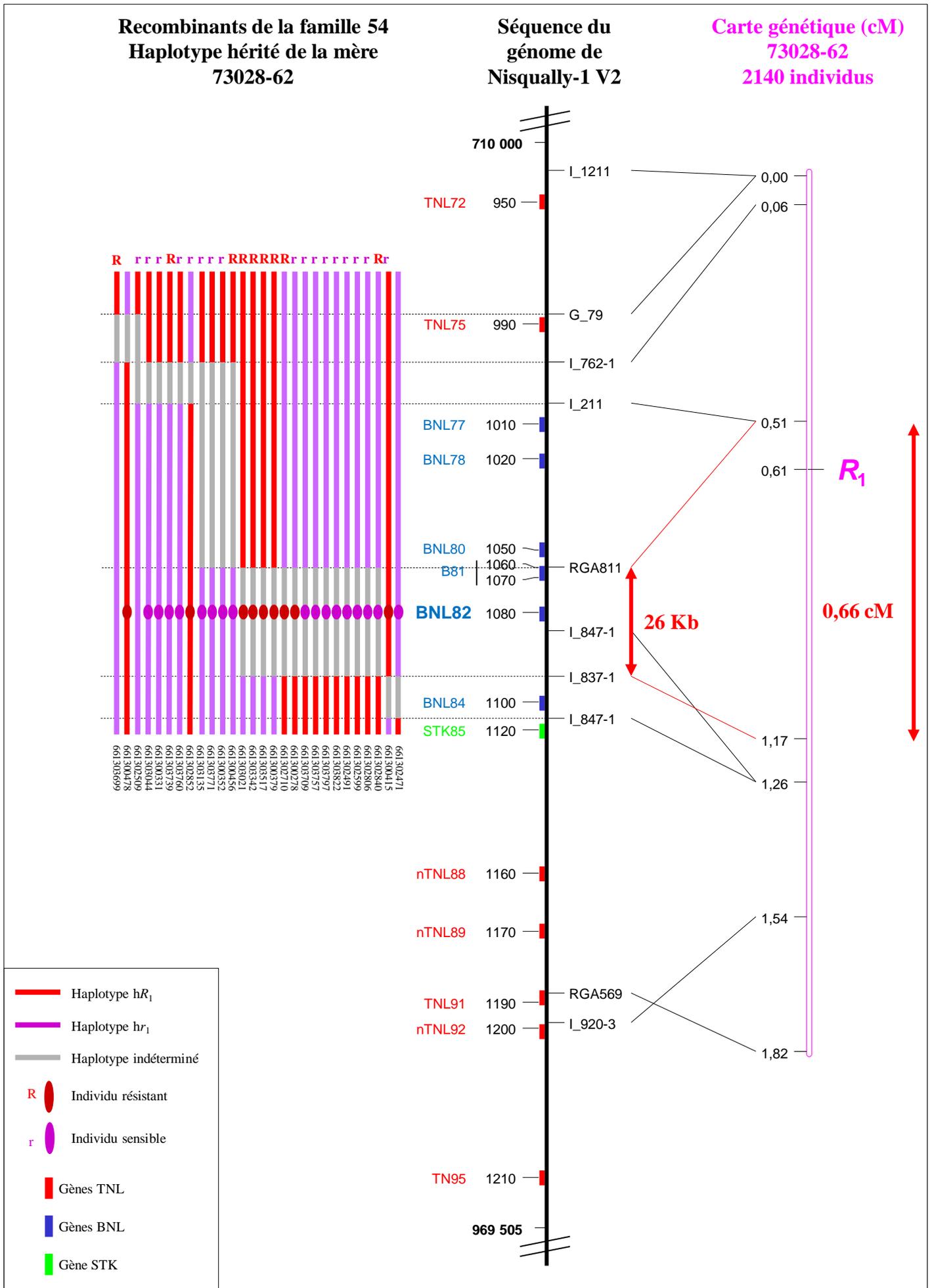
A l'INRA de Nancy, une RTq-PCR a été réalisée sur la série de bulks aux différents temps  $R_{US}$  et  $r_{US}$ , avec 2, 4 et 8 jours après l'inoculation de *Mlp*. Le gène s'exprime de plus en plus avec le temps dans les bulks  $R_{US}$ , et ne s'exprime pas dans les bulks  $r_{US}$  (figure 55).

## 2 Vers le clonage positionnel du locus $R_1$

Il était intéressant aussi de cloner le locus  $R_1$ , contrôlant une résistance qualitative à *Mlp*. Cela permettra d'étudier plus en profondeur les relations entre les deux loci  $R_1$  et  $R_{US}$ , notamment pour savoir si les deux loci, d'une résistance qualitative et d'une résistance quantitative, sont déterminés par des gènes de la même catégorie, voire du même clade de gènes.

Certains résultats obtenus lors de la recherche du locus  $R_{US}$  ont permis d'avancer dans la localisation génomique du locus  $R_1$  proche de  $R_{US}$  ; certains marqueurs développés permettaient de voir la ségrégation pour les deux parents en même temps et d'autres marqueurs développés pour  $R_{US}$ , ne ségrégaient que chez la mère 73028-62 et ont donc servi à cartographier le locus  $R_1$ . Pour la cartographie fine de  $R_1$ , il faut maintenant travailler près de ce locus sur une zone génomique restreinte.

Les ressources génomiques pour l'espèce *P. deltoides*, notamment pour l'individu 73028-62 porteur de  $R_1$ , ne sont pas autant développées que pour l'espèce *P. trichocarpa* ; la base de ce travail est donc la synténie entre les deux espèces. L'alignement des deux cartes génétiques *P. deltoides* et *P. trichocarpa* a montré une bonne conservation de l'ordre des séquences entre les deux espèces pour les 4 premier Mb de l'extrémité du chromosome 19 (Cf. § III-2.4.5 et § III-



**Figure 56 :** Alignement de la carte génétique fine de  $R_1$  et de la séquence du génome V2.  
Le numéro des gènes correspond à la protein ID de l'annotation du JGI.

2.4.6). Il s'agit maintenant de voir si la micro-synténie près du locus  $R_1$  est utilisable pour faire aboutir le clonage du locus avec une stratégie identique à celle développée pour la cartographie fine de  $R_{US}$ . En effet, les marqueurs étant dessinés à partir de séquence *P. trichocarpa*, le nombre de marqueurs transférables est réduit.

## **2.1 Cartographie génétique fine du locus $R_1$**

### **2.1.1 Sélection des individus recombinants proches du locus $R_1$**

Il a été sélectionné 41 individus qui portent une recombinaison entre les marqueurs génétiques INRA\_1211 et RGAs2701-1 sur l'haplotype hérité de la mère 73028-62, parmi les 2140 individus qui sont utilisables pour la cartographie de  $R_1$ . Ces deux marqueurs ne sont pas les marqueurs les plus flanquants du locus  $R_1$ , ceci permet d'avoir une marge de sécurité afin d'être sûr que ce locus  $R_1$  se trouve bien dans l'intervalle défini ; en effet, certains marqueurs proches du locus permettent d'avoir seulement la moitié des génotypes pour la mère 73028-62 dans la famille 54. Il peut aussi y avoir des erreurs de génotypage qui pourraient fausser la localisation des recombinaisons et ainsi restreindre le nombre d'individus utilisable pour la cartographie fine du locus.

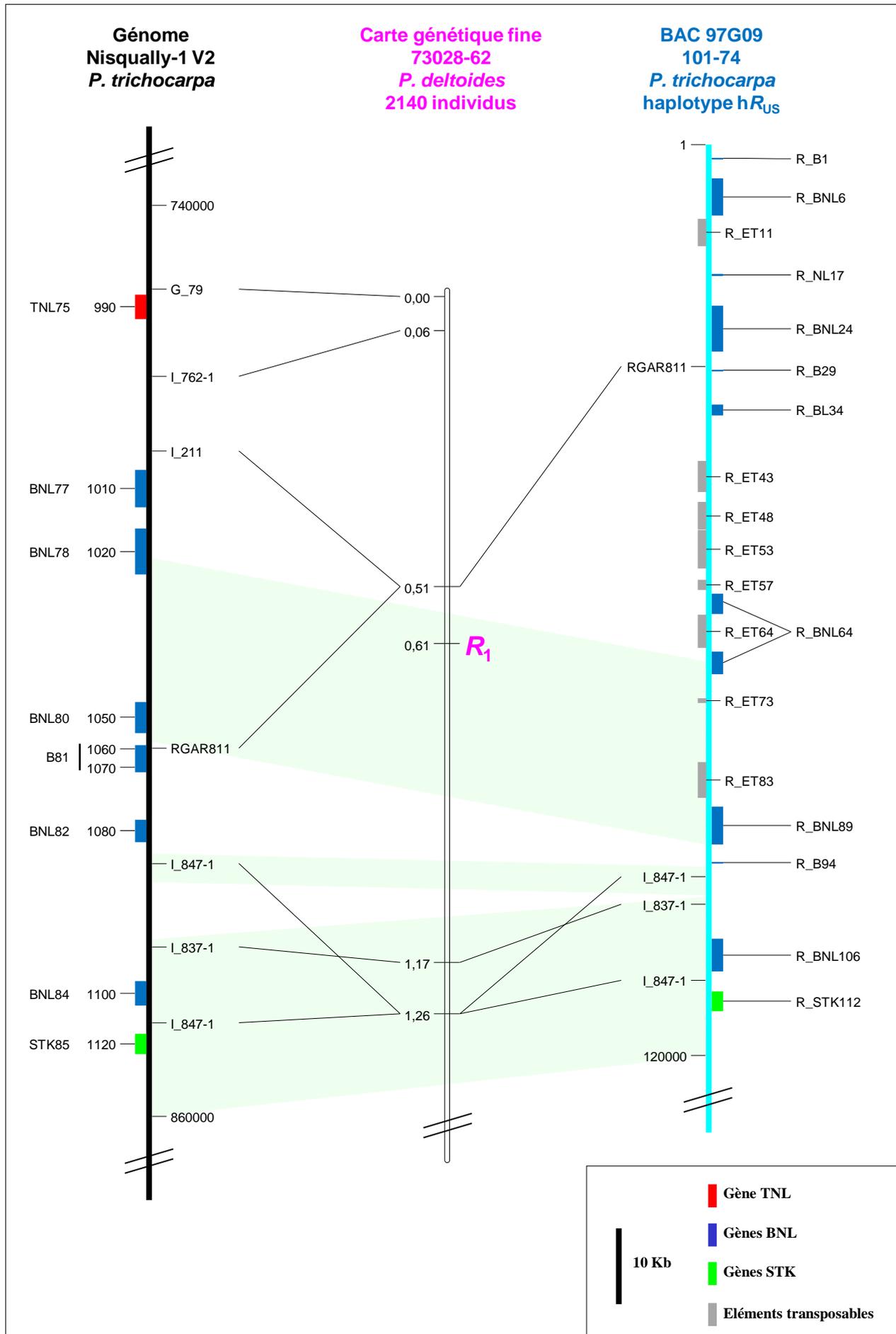
Parmi ces 41 individus recombinants autour de  $R_1$ , l'ADNg de 37 d'entre eux a pu être ré-extrait : ceci a permis de vérifier leur génotype tout en assurant la quantité suffisante pour la suite des expériences.

### **2.1.2 Développement de marqueurs génétiques**

Entre les marqueurs INRA\_1211 et RGAs2701-1, une zone d'environ 270 Kb du génome de Nisqually-1 V2 est utilisée pour développer de nouveaux marqueurs génétiques qui permettront de préciser la localisation du locus  $R_1$ . Plusieurs approches ont été menées successivement : la première a été de dessiner 14 couples d'amorces autour d'un microsatellite, la seconde a été de dessiner neuf couples d'amorces sur les gènes NBS-LRR en utilisant les alignements des gènes entre eux (MEGA4) et enfin la dernière a été de dessiner quatre couples d'amorces STS inter-génique (annexe 6).

Les tests d'amplification et de ségrégation de ces nouveaux couples d'amorces sur les parents et sur 6 descendants de la famille 54, ont permis de révéler cinq nouveaux marqueurs génétiques pour la mère 73028-62 ; quatre marqueurs SSR et un marqueur STS génique.

Au total, avec les 9 marqueurs préalables, 36 couples d'amorces ont été dessinés dans cette zone durant cette thèse, 27 couples fonctionnent chez l'individu 73028-62, parmi lesquels 7



**Figure 57** : Carte génétique fine du locus  $R_1$  alignée avec le génome de Nisqually-1 V2 et le BAC 97G09 du 101-74.

fonctionnent seulement chez l'haplotype  $hr_1$  et un seul fonctionne seulement chez l'haplotype  $hR_1$ . La séquence de Nisqually-1 semble avoir plus d'homologie avec l'haplotype  $hr_1$  qu'avec l'haplotype  $hR_1$ .

### 2.1.3 Cartographie fine du locus $R_1$

Les trois marqueurs génétiques SSR ont servi à génotyper les 37 individus recombinants. Ces nouveaux marqueurs ont permis de restreindre l'intervalle dans lequel se trouve  $R_1$  dans le cluster de gènes des BNL. Après, les couples d'amorces dans les gènes ont été dessinés dans les BNL. Le marqueur STS génique (RGAR811) a permis de restreindre l'intervalle où se trouve  $R_1$  à 0,66 cM. Dans cet intervalle, il n'y a que le seul gène BNL82 sur la séquence du génome de Nisqually-1 V2 pour une distance de 26 Kb (figure 56). Un autre gène se trouve dans cet intervalle : un gène de 759 pb et de fonction inconnue. Les marqueurs inter-géniques n'ont pas permis de confirmer que le gène BNL82 était le locus  $R_1$ .

## 2.2 Cartographie comparée

L'alignement entre le génome de Nisqually-1 V2 et le clone BAC 97G09 par Dot Plot a permis de déterminer les zones complémentaires entre l'haplotype  $hR_{US}$  et le génome de Nisqually-1 (figure 57). Il montre des différences importantes pour la région génomique au niveau du locus  $R_1$ .

L'intervalle où se trouve le locus  $R_1$  représente, sur le génome de Nisqually-1, 26 Kb avec un seul gène BNL candidat, mais sur la séquence du clone BAC 97G09, cet intervalle représente 70 Kb avec 4 gènes de la catégorie des BNL (figure 57). Le gène  $R\_BNL64$  est un gène candidat car il se peut que l'ET qui s'y est inséré, ne soit pas présent chez le 73028-62. Il permet d'exclure les gènes TNL et le gène STK pour le locus  $R_1$ . Il est donc probable que  $R_1$  soit un BNL et donc que les deux loci,  $R_1$  et  $R_{US}$ , proviennent de catégories différentes de gènes NBS-LRR.

Les différences importantes observées entre ces deux séquences proviennent d'individus de l'espèce *P. trichocarpa*, et il est possible que les différences entre les espèces *P. trichocarpa* et *P. deltoides* soient encore plus importantes.



**Partie V**  
**DISCUSSION GENERALE**



# Partie V

## DISCUSSION GENERALE

<b>1</b>	<b>Clonage positionnel des loci de résistance.....</b>	<b>105</b>
1.1	Développement de marqueurs .....	105
1.1.1	La redondance des séquences .....	105
1.1.2	Le polymorphisme.....	106
1.1.3	L'irrégularité.....	106
1.1.4	La qualité des marqueurs .....	106
1.2	Utilité des cartes génétiques.....	106
1.2.1	Ré ordonnancement des séquences du génome .....	106
1.2.2	Phénotypage de qualité.....	107
1.2.3	Construction des cartes génétiques .....	107
1.2.4	L'existence d'individus atypiques.....	109
1.3	Utilité de la banque BAC et des cartes physiques .....	110
1.4	Variation du taux de recombinaison .....	111
1.5	Distorsions de ségrégation.....	112
<b>2</b>	<b>Le séquençage de génome .....</b>	<b>112</b>
2.1	Les erreurs d'assemblage de la séquence du génome .....	112
2.1.1	Hétérozygotie.....	113
2.1.2	Redondance.....	114
2.1.3	Le 101-74 : <i>P. trichocarpa</i> ?.....	114
2.2	Les zones complexes dans le séquençage de génome.....	114
<b>3</b>	<b>Les loci de résistance .....</b>	<b>115</b>
3.1	Le locus $R_{US}$ .....	115
3.1.1	Le polymorphisme du locus $R_{US}$ .....	115
3.1.2	$R_{US}$ est-il le seul gène qui confère la résistance à <i>Mlp</i> dans le cluster ?.....	117
3.1.3	L'expression de l'allèle candidat .....	117
3.1.4	miARN.....	118
3.2	Le locus $R_1$ .....	118
3.2.1	L'exploitation de la synténie entre <i>P. deltoïdes</i> et <i>P. trichocarpa</i> .....	118
3.2.2	Les autres gènes proches de $R_1$ .....	119
3.3	$R_{US}$ et $R_1$ : deux loci différents .....	119
3.4	Le locus <i>MER</i> .....	120
<b>4</b>	<b>Cluster de gènes R .....</b>	<b>121</b>
4.1	Un super cluster de gène R .....	121
4.2	L'évolution des gènes NBS-LRR .....	121
4.2.1	Les différents clusters.....	121
4.2.2	Clade 1 des TNL .....	122
4.2.3	Le cluster de BNL où se trouve $R_1$ .....	123
4.3	Les éléments transposables .....	123



# 1 Clonage positionnel des loci de résistance

Dès le début de cette thèse, des difficultés ont été rencontrées dont la cause principale est la structure complexe de l'extrémité du chromosome 19. Ces difficultés ont été contournées, ce qui a permis de faire aboutir le clonage du locus  $R_{US}$ , qui est le premier clonage positionnel d'un gène chez les arbres forestiers. Le clonage du gène  $R_1$  n'a pas pu aboutir mais ce travail de thèse a permis de le faire considérablement avancer.

Malgré la disponibilité de la séquence du génome *P. trichocarpa* clone Nisqually-1, il a été indispensable de construire une carte génétique fine et une carte physique locale pour aboutir au clonage du locus  $R_{US}$ . En effet, la séquence du génome qui a été publiée est de qualité médiocre : il y a beaucoup de trous et de scaffolds non assignés aux chromosomes. De plus, la séquence du génome a été réalisée à partir d'un individu différent de ceux utilisés dans cette étude. Les différences entre ces individus étaient donc une difficulté supplémentaire. Ce travail de clonage positionnel a donc nécessité un constant aller et retour entre les cartes génétiques, les cartes physiques et la séquence du génome. Ce premier chapitre explique les difficultés rencontrées et les techniques pour les contourner afin d'aboutir à ce clonage positionnel.

## 1.1 Développement de marqueurs

Plusieurs challenges se sont présentés rendant difficile le développement des marqueurs utiles à l'élaboration des cartes génétiques et physiques : le polymorphisme et la redondance de l'extrémité du chromosome 19. Durant cette étude, plus de 430 couples d'amorces ont été dessinés, mais seulement 96 d'entre eux ont été utilisés en cartographie génétique, et 95 en cartographie physique.

### 1.1.1 La redondance des séquences

Certains marqueurs ont été développés sur plusieurs locus du génome de Nisqually-1 (Cf. figure 32, page III-72), bien qu'il ait eu l'utilisation conjointe des logiciels ePCR et BLAST pour en vérifier la spécificité. Parmi ces couples d'amorces, les expériences ont permis de voir qu'ils n'étaient pas tous multi-locus et qu'il s'agissait d'erreurs dans l'assemblage du génome. A l'inverse, de nombreux couples d'amorces spécifiques d'un seul locus se sont révélés expérimentalement multi-locus. Et il en reste beaucoup qui doivent l'être aussi, mais que les expériences faites, n'ont pas permis de révéler.



### 1.1.2 Le polymorphisme

Le polymorphisme entre les haplotypes du génotype 101-74 et le génome de Nisqually-1 (SNP, InDel de tailles variables) n'a pas permis d'utiliser tous les marqueurs dans la famille 54. Les marqueurs développés sur certaines parties de l'extrémité du chromosome 19 de Nisqually-1 ne fonctionnaient que chez la mère *P. deltoides* et pas chez le 101-74 *P. trichocarpa*. La difficulté due au polymorphisme l'était particulièrement pour les régions des loci  $R_1$  et  $R_{US}$  ; dans la figure 32, page III-72,  $R_{US}$  est placé au début du scaffold 117, où de nombreux couples d'amorces ne fonctionnent que chez le 73028-62.

### 1.1.3 L'irrégularité

La région génomique cible est constituée de séquences avec un fort niveau de polymorphisme et une forte redondance, mais la principale difficulté a été que ces deux paramètres n'étaient pas réguliers. Il a donc fallu dans ce travail constamment adapter le type de marqueurs à chaque région à l'intérieur de l'extrémité du chromosome 19.

### 1.1.4 La qualité des marqueurs

A de nombreuses reprises, plusieurs couples d'amorces ont été dessinés sur un même locus. Généralement, ils ont donné les mêmes résultats à la différence qu'ils n'avaient pas la même qualité d'amplification. La sélection du meilleur marqueur parmi ceux testés a permis d'obtenir les marqueurs utiles pour l'élaboration des cartes génétiques et physiques fiables, indispensables à un travail de qualité.

## 1.2 Utilité des cartes génétiques

La carte génétique devait être la plus aboutie possible, en effet, il n'était pas possible de faire une approche "gène candidat" : le nombre de gènes NBS-LRR et leur similarité nucléotidique aurait rendu impossible une telle approche. Il fallait donc l'élaboration d'une carte génétique de qualité, où la localisation du locus  $R_{US}$  devait être parfaite.

### 1.2.1 Ré ordonnancement des séquences du génome

Les marqueurs génétiques développés sur la séquence et positionnés par cartographie génétique conduisaient aussi à un ordre différent de celui indiqué dans la séquence du génome, ce qui obligeait à repartir du génome pour redessiner de nouveaux marqueurs. La reconstruction d'un ordre plus cohérent entre carte génétique et séquences assemblées apparaissait donc



incontournable pour redessiner de nouveaux marqueurs efficaces ; cette étape a amené logiquement plus de robustesse dans les ressources disponibles.

### **1.2.2 Phénotypage de qualité**

Le clonage positionnel d'un gène demande un phénotypage de qualité afin de localiser correctement l'allèle candidat pour le locus  $R_{US}$ .

La meilleure statistique utilisée pour la qualité du phénotypage de la famille 54 a abouti à quasiment aucune incohérence entre le phénotype et le génotype pour le locus  $R_1$  mais il y a plus d'incohérences pour le locus  $R_{US}$  (Cf. figure 35, page III-77). Ces quelques incohérences sont sûrement dues à quelques erreurs humaines et ont été vérifiées pour le phénotype comme pour le génotype. Comme l'étude de Jorge *et al*, (2005) l'a montré, il existe plusieurs autres QTL dans cette population en ségrégation. Si tous les autres QTL ont un effet contraire au locus  $R_{US}$ , il se peut que la lecture du phénotype soit faussée pour le caractère « taille des sores ». Mais il s'agit là d'une partie mineure, en effet plus de 70% de la différence phénotypique est expliqué par le locus  $R_{US}$ . Dans le cas d'incohérence génotype [ $R_{US}$ ] et phénotype  $r_{US}$ , il se peut aussi qu'il y ait une mutation qui altère le gène  $R_{US}$ . Le cas le plus probable est une insertion d'un élément transposable.

### **1.2.3 Construction des cartes génétiques**

Le clonage positionnel d'un gène demande une carte génétique fine de qualité afin de déterminer correctement les marqueurs qui encadrent et localisent correctement le gène candidat pour le locus  $R_{US}$

#### **1.2.3.1 Marqueurs génétiques dominants et co-dominants**

L'obtention de marqueurs co-dominants est plus intéressante, ils permettent d'introduire moins de faux négatif, ce qui arrive avec les marqueurs dominants. Dans ce cas, il est impossible de différencier un individu homozygote "absent" d'un résultat négatif issu de l'absence d'amplification d'un fragment PCR. Mais il n'est pas toujours possible d'avoir des marqueurs co-dominants ; pour certaines zones, il était impossible de trouver ces marqueurs dans le cas de grands InDels.

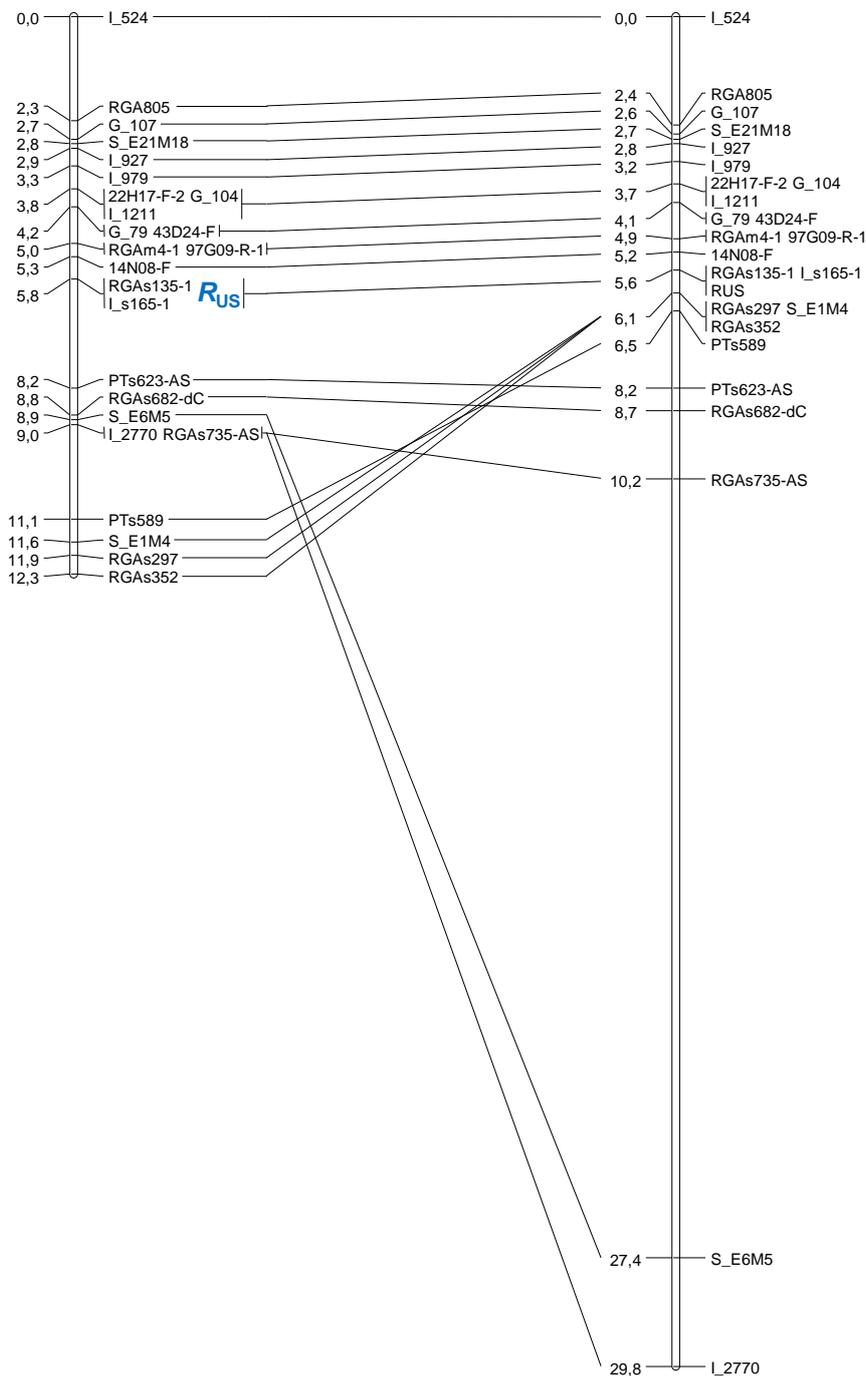
#### **1.2.3.2 Les données des cartes génétiques locales**

La carte génétique de l'individu *P. deltoides* est effectuée avec une matrice moins complète, ne comportant que 82,6 % de données présentes. Ceci est dû à plusieurs marqueurs

**Carte génétique locale**  
**101-74 *P. trichocarpa***  
**LG XIX**  
**1393 individus**

**A : MAPMAKER**

**B : Tmap**



**Figure 58** : Comparaison entre les logiciels de cartographie génétique

MAPMAKER et Tmap.

A : Carte génétique élaborée avec le logiciel MAPMAKER. B : Carte génétique élaborée avec le logiciel Tmap. les deux cartes génétiques sont élaborées avec les mêmes données correspondantes aux 336 premiers individus et la première extension de la famille 54.

dont l'amplification qui ségrège pour la femelle *P. deltoides*, est masquée par l'amplification du père *P. trichocarpa* pour la moitié des génotypes de la famille 54 (cas n°2 et cas n°6 dans la figure 33, page III-73).

Les couples d'amorces étant dessinés à partir de séquences *P. trichocarpa*, le nombre d'amorces utilisables pour la cartographie de l'individu maternel appartenant à l'espèce *P. deltoides*, est nettement réduit. De plus, lorsque les amorces amplifient chez *P. deltoides*, elles amplifient aussi chez *P. trichocarpa*. La ségrégation des allèles *P. deltoides* est alors masquée par la présence des allèles *P. trichocarpa*.

### **1.2.3.3 Les logiciels de cartographie génétique**

Le logiciel de cartographie génétique a une grande importance dans l'obtention d'une carte génétique de qualité. Plusieurs logiciels ont été testés :

#### **1.2.3.3.1 MAPMAKER**

Le logiciel utilisé dans ce travail de thèse a plusieurs défauts :

- La fonction « détection des erreurs » du logiciel MAPMAKER ne semble pas fonctionner correctement. Une erreur est systématiquement déclarée quand il y a une donnée manquante à proximité d'une recombinaison.

- Lorsque les marqueurs sont trop proches, le logiciel ne permet pas de les ordonner de façon fiable.

- MAPMAKER ne fonctionne pas avec un grand nombre d'individus : il lui est impossible de prendre plus de 2000 individus en une seule analyse.

#### **1.2.3.3.2 Tmap**

Un autre logiciel, Tmap (Cartwright *et al*, 2007) a été testé avec les mêmes données de cartographie génétique. Cette comparaison a été faite avec la carte génétique du 101-74 pour 1393 individus et 26 marqueurs. Ce sont les données que j'avais à la date du 20 novembre 2008, elles comportaient 90,9 % de données présentes et un taux d'erreurs global de 0,82 % (figure 58). L'ordre des marqueurs génétiques donné par ce logiciel n'était pas en cohérence avec mon analyse manuelle, qui était confirmée par la cartographie physique. Le logiciel Tmap produit une carte génétique la plus petite possible en taille, même s'il y a de nombreuses erreurs sur certains marqueurs. Ce problème ne peut être contourné, car l'ordre des marqueurs donné par le logiciel Tmap ne peut être modifié.

Par conséquent, il était donc important de faire une analyse manuelle approfondie qui permettait de déterminer chaque erreur et chaque recombinaison, et par conséquent de localiser



correctement les loci de résistance. Le logiciel MAPMAKER a été retenu car il permettait de forcer l'ordre des marqueurs déterminés manuellement.

#### **1.2.3.4 Carte génétique consensus entre les parents**

La comparaison entre la carte 101-74 et la carte génétique 73028-62 montre une bonne conservation de l'ordre des marqueurs, mais révèle des différences du taux de recombinaison (Cf. § III-2.4.5). Une zone peut avoir un taux de recombinaison différent entre les deux parents. Par conséquent, une trop grande différence dans les distances génétiques peut aboutir à une carte consensus dont l'ordre des marqueurs est erroné. Comme le montre la figure 45, page III-88, l'ordre des marqueurs est incohérent par rapport au génome, bien que ces mêmes marqueurs soient en cohérence quand il s'agit des cartes génétiques parentales individuelles. Il était donc important de signaler la provenance de chaque marqueur. La construction de cartes génétiques consensus est donc dangereuse.

### **1.2.4 L'existence d'individus atypiques**

#### **1.2.4.1 Chez la famille 54**

L'existence d'individus atypiques (Cf. § III-2.4.2) a rendu indispensable de regarder précisément les données de génotypage pour chaque individu de la famille 54. En effet, avec le jeu de marqueurs en majorité dominants, on ne pouvait pas les identifier facilement et les rares marqueurs génétiques co-dominants qui montraient la présence des deux allèles d'un des parents étaient interprétés comme des erreurs et éliminés des résultats. Et avec ce jeu de marqueurs dominants qui est dans les deux phases (certains en association avec  $R_{US}$  et d'autres en répulsion), les événements de recombinaison étaient surestimés et par conséquent, la cartographie génétique aboutissait à des distances génétiques surévaluées.

Le regénotypage de ces individus était important pour vérifier qu'il ne s'agissait pas de contaminations entre ADN<sub>g</sub> de plusieurs individus.

#### **1.2.4.2 Dans d'autres études**

L'existence d'individus atypiques tels que des triploïdes ou des trisomiques dans des hybrides inter-espèces a déjà été remarquée. Bradshaw & Stettler, (1993) furent les premiers à caractériser des individus triploïdes ou trisomiques dans des descendances *P. trichocarpa* x *P. deltoides*. Cervera *et al*, (2001) ont éliminé 7 individus triploïdes ou trisomiques sur 232 des descendances de la mère S9-2 *P. deltoides*. Stirling *et al*, (2001), dans une tentative de clonage



positionnel du locus *MXC3*, ont quant à eux déterminé 126 individus porteurs des deux haplotypes maternels *P. deltoides* sur une population  $F_1$  de 2028 individus.

### 1.2.4.3 Différences phénotypiques

Au niveau phénotypique, ces individus atypiques de la famille 54 n'ont rien de particulier pour les différents paramètres quantitatifs et qualitatifs de la résistance (Cf. figure 35, page III-77). Ils sont en cohérence avec le caractère dominant des deux loci  $R_1$  et  $R_{US}$  : lorsqu'il y a présence de l'allèle  $R_{US}$ , le phénotype de ces individus est  $R_{US}$  et lorsqu'il y a absence de l'allèle  $R_{US}$ , le phénotype est  $r_{US}$ . Il en est de même pour le locus  $R_1$ .

Bradshaw & Stettler, (1993) ont montré que ces hybrides triploïdes avaient des caractéristiques phénotypiques différentes des descendants diploïdes issus d'un même croisement. Les caractères étudiés étaient la densité stomatique, le diamètre des cellules épidermiques et la longueur des pétioles.

Bradshaw & Stettler, (1993) ont montré que ces hybrides triploïdes étaient stériles. Ce caractère est intéressant pour les peupleraies cultivées : l'absence de pollen viable et / ou de graines pleines empêcherait le flux de gènes hybrides dans les populations voisines naturelles et de réduire ou d'éliminer la nuisance des allergènes du pollen et du "coton". Ces hybrides stériles seraient des candidats logiques pour la manipulation par génie génétique par transgénèse qui ne seraient pas libérés dans l'environnement.

## 1.3 Utilité de la banque BAC et des cartes physiques

Le clonage positionnel des gènes de résistance reste un challenge sans une banque BAC d'un génotype porteur de la résistance, même s'il y a la séquence complète d'un génome de la même espèce.

La qualité de cette séquence n'était pas suffisante pour développer tous les marqueurs nécessaires à l'aboutissement de ce travail, car il y avait beaucoup de trous. La carte physique *via* les BES a été utilisée pour combler ces trous entre les différents contigs de la séquence. Néanmoins, la disponibilité de la séquence du génome a été un avantage non négligeable dans ce travail, il a permis d'accélérer les travaux de cartographie génétique et physique avec l'élaboration rapide de marqueurs.

Pour caractériser correctement les zones très polymorphiques entre l'individu qui a servi à la séquence du génome et les deux haplotypes de l'individu de la banque BAC, il était nécessaire d'utiliser les BES pour solidifier et compléter l'ordre des contigs. En effet, rien ne pouvait réellement nous permettre, à partir du génome de Nisqually-1, de trouver l'un ou l'autre



allèle du locus  $R_{US}$  (Cf. figure 49, page IV-95). Rien qu'au sein de la même espèce, les différences de l'ordre des gènes NBS-LRR dans un cluster n'est pas exactement la même. Ceci provient du fait il n'y a pas de relation d'orthologie dans ces gènes provenant de plusieurs haplotypes de la même espèce et même entre les deux haplotypes d'un individu vivant, comme c'est le cas pour les gènes NBS-LRR sur l'extrémité du chromosome 19 chez le 101-74.

Il a fallu donc entreprendre une démarche de marche sur le chromosome pour combler certaines zones. De plus, les marqueurs BES ont permis d'orienter les clones BAC dans le contig. Les autres marqueurs permettaient seulement de localiser un clone BAC dans le contig mais ne permettaient pas de l'orienter.

Les tentatives de clonage positionnel des gènes de résistance, sans l'utilisation d'une banque BAC du génotype portant l'allèle conférant la résistance, n'ont que peu de chance d'aboutir. Par exemple, Coleman *et al*, (2009) chez la vigne, ont construit une carte génétique du locus *REN1* qu'ils comparent à la séquence du génome référence disponible. Cet investissement permet de comprendre l'organisation et les gènes candidats qui contrôlent cette résistance, mais il sera très difficile d'identifier le locus précisément. Le clonage du locus  $R_1$  est certainement dans le même cas, et il serait très utile d'investir dans une banque BAC de l'individu 73028-62 pour faire aboutir ce clonage.

Un manque de représentativité de la banque BAC a failli être très problématique à proximité du locus  $R_{US}$ . En effet, il y a seulement quatre clones BAC pour l'allèle  $R_{US}$  alors que l'on en attendait sept ou huit. De plus, le gène est à proximité d'un trou de la carte physique à 90 Kb. C'est encore plus juste pour l'allèle  $r_{US}$  avec seulement un clone BAC avec un trou dans la carte physique à 10 Kb. Bien qu'il y ait eu cette démarche de marche sur le chromosome, les cartes physiques ne sont pas complètes, il y a bien eu un investissement très important sur les marqueurs proches des extrémités des contig pour cribler la banque, mais sans réussite. Il reste donc des régions non caractérisées dues au manque de clones dans la banque BAC. La couverture en clones BAC par haplotype est de 7,75 X ce qui n'est pas suffisant pour espérer avoir une couverture complète de chacun des haplotypes. Pour limiter le nombre de régions génomiques absentes des banques, il faudrait utiliser plusieurs enzymes pour leurs constructions.

## 1.4 Variation du taux de recombinaison

La variation du taux de recombinaison le long du chromosome 19 du peuplier n'est pas une particularité du génome de cette espèce. Par exemple, la tentative de clonage positionnel du



locus *MXC3* chez le peuplier n'a pas eu cette chance car le taux de recombinaison est nul autour du locus sur 300 kb (Stirling *et al*, 2001).

La possibilité de tomber dans une zone non recombinante aurait pu rendre impossible cette démarche de clonage positionnel. En effet, avec plus de 2 000 individus, nous pouvions espérer avoir un intervalle de moins de 0,1 cM ce qui représente 20 kb (200 Kb/cM pour le peuplier ; Tuskan *et al*, 2006).  $R_{US}$  est localisé dans un intervalle de 0.14 cM qui représente 25kb, ce qui correspond quasiment à la valeur attendue. Mais cet intervalle est délimité par une seule recombinaison pour un côté de l'intervalle (Cf. § IV-1.2.4.1), sans laquelle  $R_{US}$  serait localisé dans une région sans recombinaison pour au moins 91 Kb supplémentaires.

Pour le locus  $R_1$ , il semble qu'il y ait un taux de recombinaison plus propice au clonage positionnel du locus, mais il manque une mesure physique fiable de l'intervalle dans lequel il se trouve. Il reste donc des recombinants avec lesquels il est possible de d'améliorer encore la cartographie génétique de  $R_1$ .

## 1.5 Distorsions de ségrégation

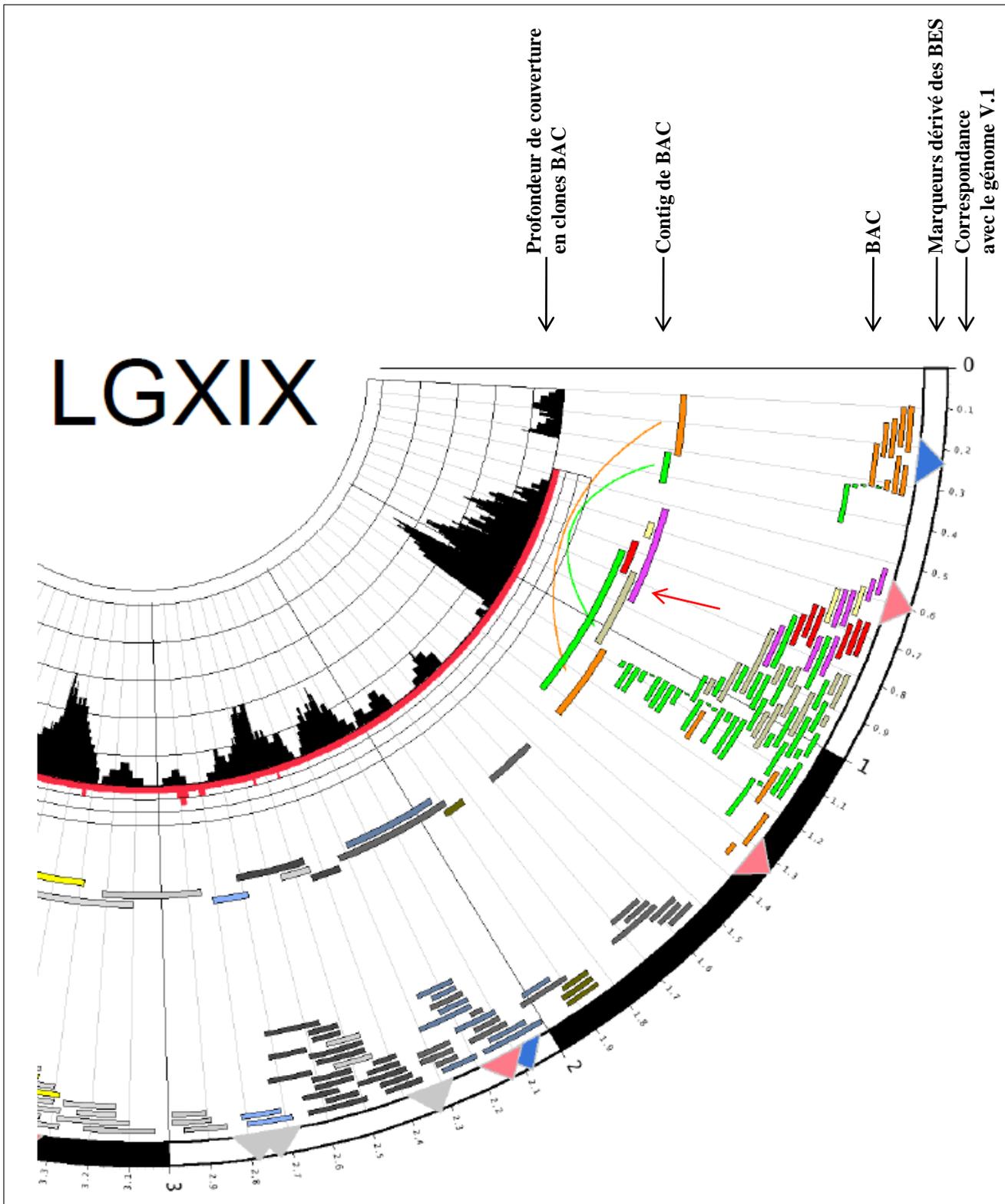
Il y a peu de distorsions de ségrégation dans la famille 54, seuls les marqueurs co-localisants avec  $R_{US}$  présentent une distorsion lors de l'analyse des 336 premiers individus. Des individus en majorité de phénotypes  $r_1 r_{US}$  trop sensibles sont morts, ce qui a diminué le nombre des individus de cette catégorie. Sur l'ensemble de la famille, il y a donc eu une bonne conservation des descendants, quel que soit le niveau de sensibilité.

A vrai dire, pour un clonage positionnel, ce n'est pas forcément un problème, car ce qui est intéressant est surtout de posséder des individus porteurs d'une recombinaison proche du locus d'intérêt pour le localiser finement. Par contre, un faible taux de distorsion permet de faire une analyse correcte de QTL ce qui est indispensable pour le clonage d'un locus quantitatif tel que  $R_{US}$ .

## 2 Le séquençage de génome

### 2.1 Les erreurs d'assemblage de la séquence du génome

Des erreurs de l'assemblage de la séquence du génome ont été révélées lors de ce travail de thèse. Elles sont principalement dues à la forte hétérozygotie, mais aussi aux diverses séquences répétées.



D'après Kelleher et al, 2007

**Figure 59** : La carte physique de Nisqually-1 par fingerprint pour l'extrémité du chromosome 19. De l'extérieur du cercle vers l'intérieur : Les chiffres indiquent la correspondance avec la séquence du génome V1. Les triangles indiquent la présence de marqueur dérivés de BES, en bleu les marqueurs intégrés dans une carte génétique *P. trichocarpa*, en rose les marqueurs qui n'ont pas pu être intégrés à une carte génétique, en gris les marqueurs qui n'ont pas été testés. Les barres représentent les clones BAC dont la couleur correspond au contig auxquels ils appartiennent. Le dernier histogramme correspond à la profondeur en couverture en clone BAC avec une échelle de 5X. Les lignes rassemblant plusieurs contigs montre un même contig placé dans deux zones différentes du génome. La flèche rouge indique la zone où il y a superposition de plus de deux contigs de clones BAC.

### 2.1.1 Hétérozygotie

L'assemblage des séquences produites pour le séquençage en WGS est très compliqué chez un individu diploïde et hétérozygote. En effet, l'assembleur ne peut pas détecter la complémentarité entre les deux haplotypes quand la zone est très hétérozygote, il les place alors dans deux régions différentes.

Quand la zone est fortement hétérozygote, il se produit un effet dit « effet ciseaux » qui conduit à une inversion des séquences. Ce phénomène a été remarqué pour la première fois lors de la construction de cartes physiques par fingerprint chez la vigne (Scalabrin *et al*, 2010). Le logiciel d'assemblage ne reconnaît pas les deux allèles d'une même zone génomique ; par conséquent, il les place l'une après l'autre. La première est correctement placée et la seconde est placée à la suite mais dans le sens opposé. Ces contigs de clones BAC correspondants à la zone cible sont localisés dans 2 zones différentes (figure 59). Le contig de clones BAC de la carte physique de Nisqually-1 le plus à l'extrémité du chromosome 19, est mal localisé (figure 59). On peut émettre l'hypothèse que ce phénomène est à l'origine de certaines erreurs d'assemblage repérées dans les deux versions de la séquence du génome. Cette même zone est montrée comme fortement hétérozygote chez Nisqually-1 (Yin *et al*, 2008). Ces erreurs d'assemblage pour une part partent de la zone qui est très fortement hétérozygote chez le 101-74 et dans laquelle sont les loci  $R_1$  et  $R_{US}$  (zone 3 dans la figure 47, page III-92).

Mais le plus étonnant dans le cas du peuplier est que ces zones mal assemblées en inversion se trouvent à des distances très éloignées ; 800 kb dans la version V1 et 1 800 kb dans la version V2.

L'élaboration d'une carte génétique dense aurait permis de contourner ce genre d'erreurs d'assemblage. Il n'y a eu que 356 marqueurs microsatellites qui ont été utilisés dans la première version de l'assemblage de Nisqually-1 et qui ont permis d'assigner 155 scaffolds aux chromosomes (Tuskan *et al*, 2006) : cela fait en moyenne moins d'un marqueur génétique tous les Mb. Dans la seconde version, l'augmentation des marqueurs dans les cartes génétiques a permis d'assigner plus de séquences aux chromosomes et de corriger certaines erreurs (Drost *et al*, 2009).

L'erreur a été de prendre un individu porteur de deux haplotypes et d'en faire une seule séquence pour les deux haplotypes, ce qui pose des problèmes pour des régions très hétérozygotes. Les cartes physiques et génétiques permettraient de mieux prendre en compte les deux haplotypes et ainsi éviter ce type d'erreur. C'est une nécessité lorsque la région génomique est complexe comme celle d'un cluster de gène NBS-LRR.



### 2.1.2 Redondance

L'assemblage par WGS peut aussi être problématique lorsque la zone est redondante. En effet, l'assembleur ne peut pas détecter si des zones différentes et très redondantes entre elles, ne proviennent pas d'une même zone, et il les incorpore dans une seule zone.

C'est le cas pour une autre zone, très redondante, qui semble mal assemblée dans le génome de Nisqually-1 (zone 2 dans la figure 44, page III-86). Dans cette région il existe une forte redondance due à des duplications récentes, où même les zones inter-géniques sont redondantes. Dans la carte physique de Nisqually-1, produite par fingerprint (Kelleher *et al*, 2007), il y a superposition de plus de 2 contigs de clones BAC (flèche rouge dans la figure 59). Le séquençage de ces clones BAC aurait permis de caractériser correctement la zone.

### 2.1.3 Le 101-74 : *P. trichocarpa* ?

Est-ce que les différences observées entre le 101-74 et Nisqually-1 révèlent des différences qui existent au sein de l'espèce *P. trichocarpa* ? Cervera *et al*, (2005) ont produit une phylogénie du genre *Populus* avec des marqueurs AFLP sur des populations naturelles dont l'individu 101-74 faisait partie. Cent cinquante et une bandes polymorphiques ont été analysées, ce qui a permis de regrouper l'individu 101-74 parmi les autres individus *P. trichocarpa*. L'individu 101-74 n'est donc pas un individu d'une autre espèce ou même un individu atypique de l'espèce *P. trichocarpa*. L'individu 101-74 et Nisqually-1 sont tous les deux originaires de l'Oregon, récoltés le long de la rivière Nisqually. Le polymorphisme et la redondance observés chez le 101-74 dans cette région génomique et entre le 101-74 et Nisqually-1 ne sont donc pas dus à une différence d'espèce.

Les différences dans l'ordre des séquences entre le 101-74 et Nisqually-1 ne sont donc certainement pas dues à une différence d'espèce, mais il s'agirait bien d'erreurs d'assemblage. Ces erreurs de l'assemblage du génome de Nisqually-1 sont confirmées par la carte génétique de la mère 73028-62 colinéaire à celle de l'individu 101-74. Il renforce aussi l'idée que le polymorphisme observé dans les régions des loci  $R_1$  et  $R_{US}$  est une particularité de ces zones.

## 2.2 Les zones complexes dans le séquençage de génome

Beaucoup de génomes de diverses espèces diploïdes, animales ou végétales, ont été séquencés avec succès. Cependant, chez les espèces les plus hétérozygotes, la gestion des haplotypes des zones complexes n'est pas assez prise en compte pour assurer un assemblage de



bonne qualité. Avec les avancées des techniques de séquençage, de plus en plus de génomes d'espèces hétérozygotes, seront séquencés.

Il est généralement recherché de placer seulement les scaffolds non assignés aux chromosomes par l'établissement de carte génétique à postériori. Il devrait aussi être recherché ces zones complexes pour pouvoir mieux les assembler.

Les résultats de ce travail de thèse montrent que pour les zones redondantes et polymorphes, l'établissement de cartes génétiques et physiques devrait probablement être envisagé. L'utilisation de clones BAC permet de séquencer indépendamment les deux haplotypes et donc de contourner les deux problèmes majeurs propres au cluster de gènes de résistance : redondance et hétérozygotie (duplication, SNP, InDel). Le développement de cartes génétiques permet de différencier les haplotypes. Elle permet donc de pouvoir mieux assembler les zones les plus hétérozygotes d'un génome.

Avec les nouvelles technologies de séquençage, les données sont plus nombreuses pour un moindre coût de production, mais les assemblages sont de moindre qualité. Pour un génome complexe comme celui du peuplier, il est maintenant possible de faire l'assemblage par alignement sur la référence du génome déjà produit. Dans des zones comme celle de cluster de gènes de résistance, on peut s'attendre à des résultats partiels, et même erronés.

## 3 Les loci de résistance

### 3.1 Le locus $R_{US}$

Avec l'alignement de la carte génétique et de la séquence des clones BAC chevauchants 04J18 et 97F12, nous avons pu déterminer que le locus  $R_{US}$  correspond au gène R\_nTNL98. Mais ce résultat ne tient qu'à une seule recombinaison de la famille 54, porté par l'individu 661303763 (Cf. § IV-1.2.4.1). Le gène candidat pour  $r_{US}$  n'est pas déterminé avec uniquement les marqueurs génétiques, mais surtout pas l'orthologie qui existe entre la région du gène candidat pour  $R_{US}$  et celle du gène candidat pour  $r_{US}$ .

#### 3.1.1 Le polymorphisme du locus $R_{US}$

De nombreuses différences ont été remarquées entre les allèles de  $R_{US}$  et  $r_{US}$ . Mais nous ne pouvons pas conclure sur la responsabilité de l'une d'entre elles pour expliquer la différence phénotypique.



Comme l'allèle  $r_{US}$  n'est pas le plus proche en terme de séquence pour certaines parties du gène R\_nTNL98, candidat pour  $R_{US}$ , il est donc plus intéressant de comparer l'allèle  $R_{US}$  avec tous les gènes du clade 1 de TNL. En effet, les autres gènes de ce clade chez le 101-74 ne confèrent pas de résistance, pour le paramètre quantitatif taille des sores. Cette comparaison peut être étendue à l'ensemble des gènes de ce clade provenant de la séquence du génome de Nisqually-1. Cet individu est de phénotype gros sores donc  $r_{US}$ , il ne possède donc pas de gène responsable de résistance quantitative agissant sur la taille des sores. La comparaison de la séquence du gène candidat avec les autres gènes du même clade peut fournir quelques indices sur les raisons de la résistance de l'allèle  $R_{US}$ . Y a-t-il quelque chose de spécifique dans le gène candidat qui explique cette résistance ?

La résistance conférée par  $R_{US}$  peut aussi être due au bon enchaînement de ces domaines : avoir le bon promoteur, la présence du domaine NLS, un bon domaine TIR et NBS, ainsi qu'un domaine LRR qui permet l'interaction directe ou indirecte avec l'effecteur de *Mlp*.

#### **3.1.1.1 Domaine NLS**

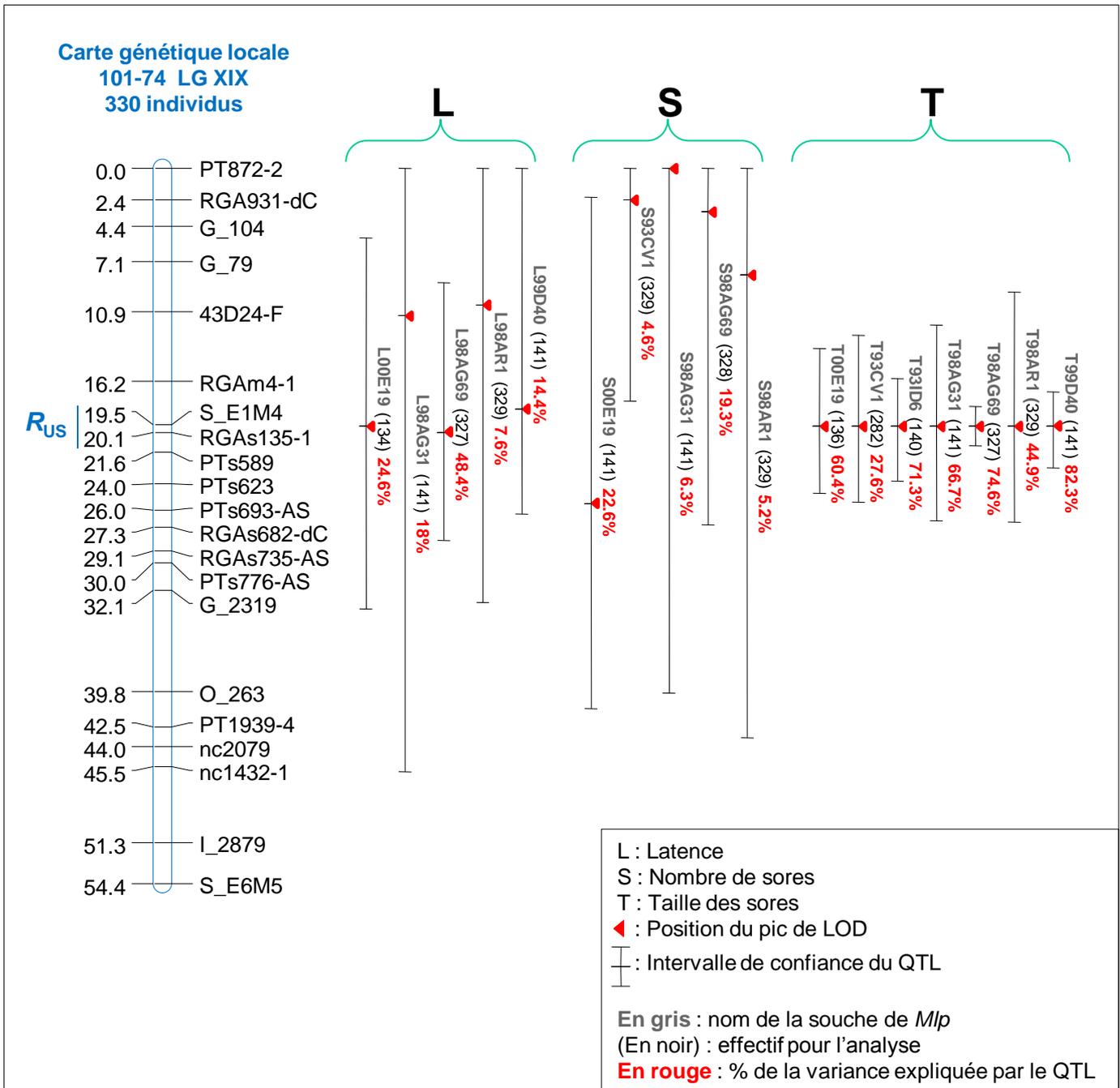
La différence la plus marquante est la présence d'un domaine NLS, d'adressage au noyau, chez l'allèle  $R_{US}$  comparé à l'allèle  $r_{US}$  qui n'en possède pas (Cf. § IV-1.3.2). Parmi les 15 gènes TNL complets (possédant l'enchaînement de ces trois domaines) du clade 1 caractérisés dans la séquence des clones BAC du 101-74, 9 d'entre eux portent le domaine NLS.

Le domaine NLS du gène candidat pour  $R_{US}$  est différente de deux acides aminés par rapport au domaine utilisé par Kolher *et al*, (2008) (Cf. § I-5.2.1.4) pour la validation de sa fonctionnalité (figure 60). Il reste donc à prouver la fonctionnalité du domaine NLS pour le gène candidat R\_nTNL98.

Est-ce que la localisation cellulaire de la protéine peut avoir de l'importance dans l'expression de la résistance ? On peut très bien imaginer, dans le modèle de garde, un effecteur qui cible une protéine du compartiment nucléaire et que la protéine  $R_{US}$  détecte la modification due à l'interaction effecteur-protéine. Dans cette hypothèse, l'effecteur n'est pas détectable dans le cytoplasme, il est donc indispensable que la protéine R soit localisé dans le noyau.

#### **3.1.1.2 Domaine LRR**

Le domaine LRR est peut-être à regarder en priorité car il est considéré comme responsable de l'interaction plante-pathogène et donc de la spécificité de l'interaction. Les études publiées montrent que la partie N-terminale des gènes de type NBS-LRR sont moins polymorphes que leur partie C-terminale et donc que les domaines LRR sont les plus



**Figure 61** : QTLs de résistance quantitative évaluée en laboratoire chez le 101-74 vis-à-vis de sept souches de *Mlp*.

Cartographie génétique disponible au 26/07/2007 uniquement avec les marqueurs à position robuste. Les QTLs des différents paramètres de la résistance qualitative sont alignés par rapport à la carte génétique.

polymorphes. Mais ce n'est pas le cas si l'on compare les gènes candidats pour  $R_{US}$  et  $r_{US}$ , où le domaine NBS est plus polymorphe que le domaine LRR. Mais en comparant l'ensemble de gènes du clade 1 de TNL, les domaines LRR sont bien plus polymorphes.

### 3.1.1.3 Domaine NBS

Il existe plus de polymorphismes entre les domaines NBS qu'entre les domaines LRR entre les deux allèles du locus  $R_{US}$ .

### 3.1.2 $R_{US}$ est-il le seul gène qui confère la résistance à *Mlp* dans le cluster ?

Une nouvelle détection de QTL a été entreprise par Véronique Jorge avec une carte génétique dont les données datent du 26 juillet 2007 et avec les données phénotypiques de Jorge *et al.*, (2005) (figure 61). Des QTL préalablement localisés dans le même intervalle sont localisés avec une meilleure précision. Les deux autres paramètres de la résistance qualitative (la latence et le nombre de sores) montrent une ségrégation dans la même zone génomique. La question qui en découle est de savoir si d'autres gènes, présents dans la région génomique de  $R_{US}$ , pourraient être responsables de la variabilité phénotypique observée pour ces deux autres paramètres. En effet il y a un nombre considérable de gènes NBS-LRR qui sont susceptibles de contrôler un autre paramètre de la résistance à *Mlp*. De plus, un fort QTL tel que  $R_{US}$ , peut masquer un autre QTL dont l'effet est moins fort.

La latence est un paramètre qui est lié à la taille des sores ; plus l'apparition des premiers sores est tardif, plus les sores seront petits au bout de 14 jours.

Par contre, le nombre de sore ne paraît pas être lié à la taille des sores. Il se peut que ces paramètres soient déterminés par d'autres loci que  $R_{US}$ . Pour ce paramètre, les pics de LOD ne sont pas localisés au même endroit que  $R_{US}$ , mais ce sont des % qui sont beaucoup plus faibles.

### 3.1.3 L'expression de l'allèle candidat

L'induction d'un gène de résistance par l'infection de l'agent pathogène, comme c'est le cas pour  $R_{US}$ , n'a été reportée que dans un petit nombre de cas. D'autres cas ont montré que certains gènes NBS-LRR étaient affectés par le stade de développement ou par les conditions environnementales (Tan *et al.*, 2007), et qu'ils étaient exprimés spécifiquement dans certains tissus.

Le nombre réduit d'EST détecté pour des gènes NBS-LRR et les gènes apparentés d'*A. thaliana*, révèle un niveau d'expression relativement bas (Tan *et al.*, 2007). C'est aussi le cas chez le peuplier.



Quels sont les mécanismes qui existent entre l'attaque du pathogène et l'induction du gène de résistance qui le reconnaît spécifiquement et permet le déclenchement des mécanismes de résistance ? Il y a sûrement beaucoup de cas différents, mais qu'en est-il pour le gène candidat *R\_nTNL98* dont l'expression est augmentée par *Mlp* ? Deux hypothèses :

- Sans *Mlp*, le gène est exprimé mais à un taux très faible ce qui le rend indétectable, et la reconnaissance de l'agent pathogène par le produit du gène provoque une autorégulation positive de lui-même et donc de son expression qui va augmenter.

- Il peut s'agir aussi d'une perception universelle d'une attaque d'un agent pathogène qui provoque l'expression de ce gène, et dont le produit reconnaît spécifiquement *Mlp*.

Tan *et al*, (2007) ont montré aussi qu'il n'y avait pas de corrélation entre le placement des gènes dans les groupes et sous groupes des gènes NBS-LRR ou leur localisation et des motifs d'expression.

Même si les résultats de Q-PCR obtenus avec des amorces spécifiques de *R<sub>US</sub>* montrent que l'expression de l'allèle est stimulée lors de l'interaction, cela ne constitue pas une preuve suffisante.

### **3.1.4 miARN**

Klevebring *et al*, (2009) ont montré la présence de nombreux miARN sur l'extrémité du chromosome 19, notamment des miARN qui ont une action sur les gènes NBS-LRR.

## **3.2 Le locus *R<sub>1</sub>***

### **3.2.1 L'exploitation de la synténie entre *P. deltoides* et *P. trichocarpa***

L'hétérozygotie dans la zone de *R<sub>1</sub>* pour l'individu 101-74 de *P. trichocarpa* est importante, mais aussi entre le 101-74 et Nisqually-1. Mais quelle est la situation pour l'hétérozygotie entre les deux haplotypes chez l'individu 73028-62 *P. deltoides* pour ce cluster de BNL ? Le clonage positionnel de *R<sub>1</sub>* se trouve sûrement dans une impasse due au fort polymorphisme de la zone, mais en plus, il y a la barrière de l'espèce.

Même si on obtient des marqueurs sur des gènes qui co-localisent avec la résistance *R<sub>1</sub>*, on ne pourra pas affirmer qu'il s'agit du gène *R<sub>1</sub>*. En effet, il se peut qu'il y ait d'autres gènes qui co-localisent avec le locus *R<sub>1</sub>* dont un soit le gène responsable de la résistance.



### 3.2.2 Les autres gènes proches de $R_1$

Une hypothèse a été donnée sur la présence d'une résistance contournée par la virulence 7 qui est contrôlée par deux gènes en répulsion dans l'individu 73028-62 (Cf. § I-6.2.2.1). Le fait que deux gènes NBS-LRR sont nécessaires pour une résistance a déjà été rapportée (Ashikawa *et al*, 2008).

Avec les cinq individus qui sont résistants à la souche 93CV1 (Cf. § I-6.2.2.1), une hypothèse a été qu'il y a deux gènes en répulsion et que ces cinq individus portent une recombinaison entre ces deux gènes et donc qu'ils sont les seuls à posséder la résistance héritée du 73028-62. Mais ces cinq individus sont de génotype [ $R_1$ ] ne portant pas de recombinaison génomique à proximité du locus  $R_1$ . L'hypothèse de recombinaison entre les deux gènes n'est donc pas vérifiée.

Ceci ne remet pas en cause l'hypothèse de deux gènes en répulsion, qui explique que la plus grande partie de la descendance de la famille 54 n'a pas hérité de la résistance à la souche 93CV1. Même si cela paraît peu probable, il faudrait donc une réactivation d'un allèle en répulsion par rapport à  $R_1$ .

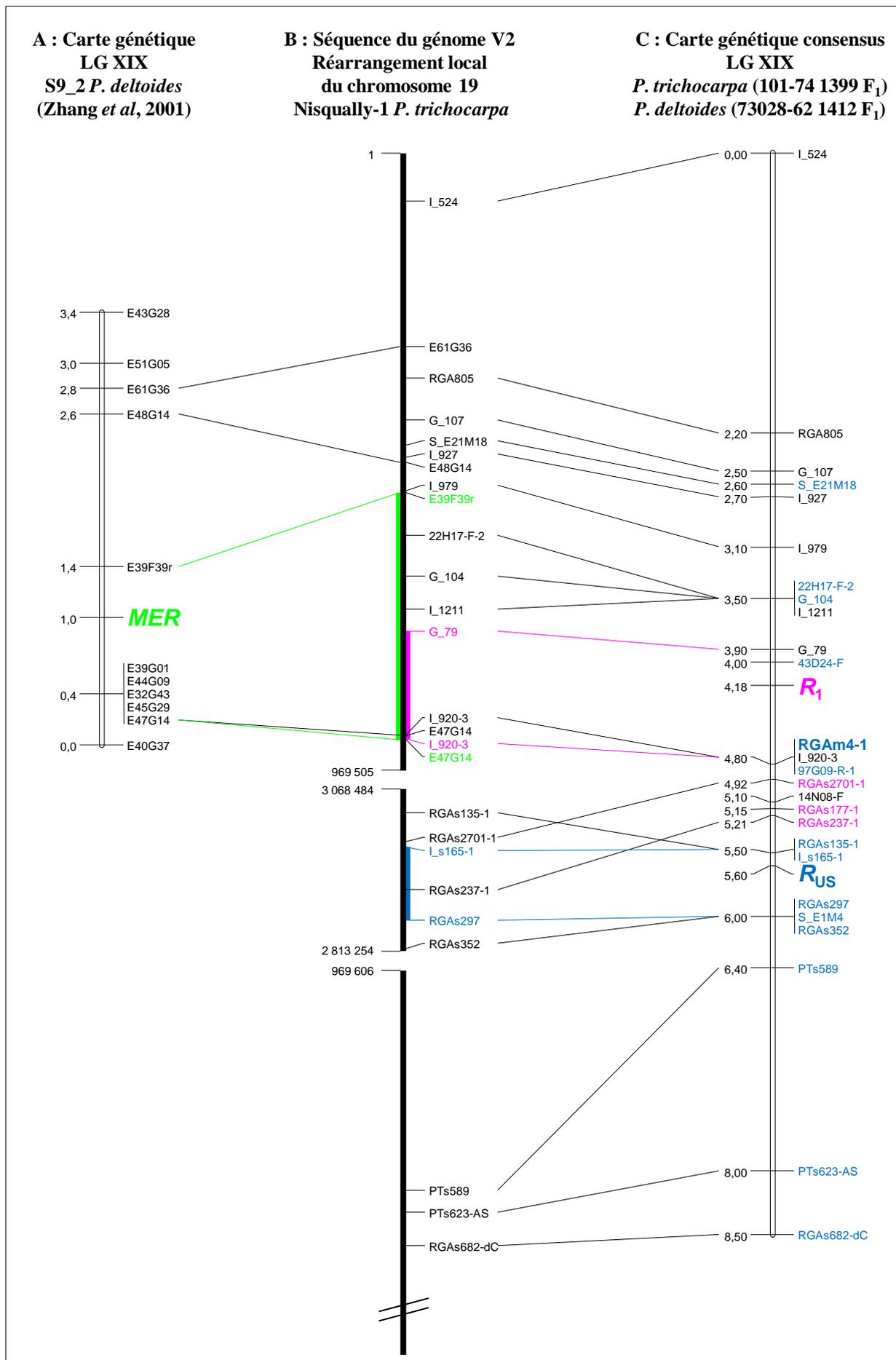
### 3.3 $R_{US}$ et $R_1$ : deux loci différents

La région du locus  $R_1$  est différente de celle du locus  $R_{US}$ . Le premier est localisé dans un cluster de gènes BNLs, le second fait parti d'un cluster de gènes TNLs. Mais ils sont tous les deux localisés à une extrémité du chromosome 19. La co-localisation de loci de résistances qualitatives et quantitatives dans un même cluster de gènes NBS-LRR a déjà été remarquée comme par exemple chez le haricot, et la vigne (Moroldo *et al*, 2008 ; Geffroy *et al*, 2000).

Ce travail permet de connaître un gène qui contrôle un paramètre de la résistance quantitative chez le peuplier et c'est un gène de la famille des NBS-LRR. Ceci a déjà été remarqué chez le riz (Wisser *et al*, 2005). Cette résistance quantitative est contrôlée par un gène NBS-LRR et a été contournée par différentes souches de *Mlp* (Dowkiw *et al*, 2010) comme pour les résistances qualitatives.

Ceci suggère que les mécanismes de résistance quantitatifs et qualitatifs ont des caractéristiques communes. On peut les placer toutes les deux dans des interactions gène R-effecteur.

Les résistances quantitatives ne seraient simplement qu'une forme affaiblie des gènes R. C'est l'une des hypothèses formulée par Poland *et al*, (2009). Dans le cas de  $R_{US}$ , il ne s'agit pas



**Figure 62** : Les loci *MER*, *R*<sub>1</sub> et *R*<sub>US</sub> localisés sur le génome de Nisqually-1 V2.

A : Carte génétique fine de l'individu S9\_2 *P. deltoides* (Zhang *et al*, 2001). B : séquence du génome V2, réarrangement local. C : carte génétique consensus élaborée avec les 336 premiers individus et la première extension de la famille 54, les marqueurs génétiques en bleu sont spécifiques du père 101-74, en violet ils sont spécifiques de la mère 73028-62 et en noir ils sont communs aux deux parents. L'intervalle où se trouve *MER* est indiqué en vert, l'intervalle où se trouve *R*<sub>1</sub> est indiqué en rose, l'intervalle où se trouve *R*<sub>US</sub> est indiqué en bleu.

d'un gène de résistance qui a été partiellement contourné par *Mlp* : en effet, le gène n'était pas auparavant en présence de ce pathogène.

### 3.4 Le locus *MER*

Les séquences des 11 bandes AFLP développées par Zhang *et al*, (2001) pour cartographier l'individu S9-2 *P. deltoides* porteur du locus *MER*, ont été alignées sur le chromosome 19 du génome de Nisqually-1 V2 ; ceci permet de préciser la localisation de ce locus sur la séquence du génome (figure 62), ce qui est nouveau. Les marqueurs AFLP E47G14 et E39F39r qui encadrent *MER* dans un intervalle de 1 cM, délimitent une région d'environ 390 Kb sur le chromosome 19, mais cet intervalle ne tient qu'à une seule séquence AFLP de 471 pb qui est sur un gène NBS-LRR, et donc multi-locus.

Indirectement, par l'intermédiaire de la séquence du génome, la carte consensus du pedigree 73028-62 *P. deltoides* x 101-74 *P. trichocarpa* a été alignée avec la carte de l'individu S9-2. Cet alignement permet de visualiser la localisation des 3 loci *MER*,  $R_1$  et  $R_{US}$  par rapport à la séquence du génome (figure 62).

La localisation du locus *MER* semble différente de celle du locus  $R_{US}$ , mais elle ne tient qu'à la comparaison indirecte entre les deux locus via la séquence du génome. La localisation du locus *MER* pourrait correspondre avec celle du locus  $R_1$  ; l'intervalle dans lequel se trouve *MER* chevauche celui de  $R_1$ . Toutefois il est à souligner que cette analyse de la localisation des deux loci provenant de l'espèce *P. deltoides* est faite avec l'intermédiaire d'une séquence provenant de l'espèce *P. trichocarpa*.

Le marqueur RGAm4-1 qui a été développé sur la séquence du clone BAC *P. deltoides*, et localisé à 0,6 cM du locus *MER* (Lescot *et al*, 2004), a été cartographié génétiquement chez l'individu 101-74 *P. trichocarpa*. Sur la carte génétique consensus, il est placé entre les loci  $R_{US}$  et  $R_1$ . Ce marqueur ne fonctionnait pas chez le 73028-62. Encore une fois, les résultats soulignent le fort niveau de polymorphisme qui existe dans cette zone du génome. Aujourd'hui les données sont trop fragmentaires pour émettre une hypothèse quant à la localisation précise du locus *MER*.



## 4 Cluster de gènes R

### 4.1 Un super cluster de gène R

Les gènes NBS-LRR ou les gènes STK sont organisés en cluster sur le chromosome 19. Le super cluster de gènes NBS-LRR regroupe plusieurs clusters, avec une alternance entre clusters des trois catégories différentes de gènes : les BNLs, les TNLs et les STKs, mais il n'y a pas de présence de gènes CNL. Chaque cluster a des caractéristiques différentes en termes de nombre de gènes, de polymorphismes et de redondances entre les gènes, et les zones inter-géniques ainsi que les ETs

Le super cluster mesure près de 3 Mb avec plus d'une centaine de gènes putatifs : 27 gènes STK, 45 gènes BNL, et 31 gènes TNL. C'est l'un des plus gros caractérisé à ce jour, comparé au cluster du locus Dm3 chez la laitue (Kuang *et al*, 2004), et au cluster du locus B4 chez le haricot (Geffroy *et al*, 2009). La diversité nucléotidique des gènes observée peut vraisemblablement permettre au peuplier de posséder un large réservoir d'allèles différents. Cette diversité serait une adaptation de l'organisme pour répondre au mieux à l'évolution des pathogènes. En effet, les espèces pérennes comme le peuplier, sont plus exposées aux pathogènes avant leur reproduction et évoluent plus lentement. En effet, leur cycle de reproduction est relativement long (entre 6 et 8 ans pour arriver à l'âge de reproduction), ce qui est très désavantageux pour pouvoir s'adapter à l'évolution rapide des pathogènes comme *Mlp*.

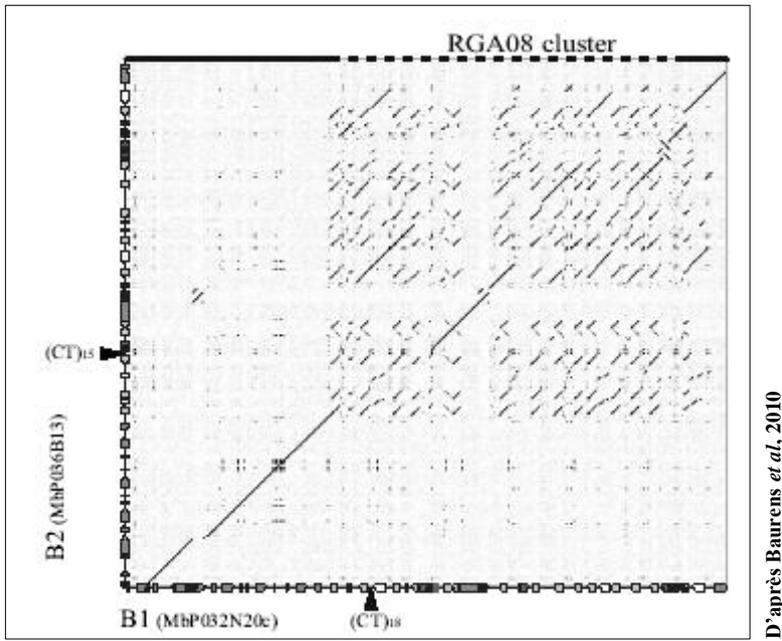
### 4.2 L'évolution des gènes NBS-LRR

#### 4.2.1 Les différents clusters

L'alternance des clusters de gènes TNLs et de BNLs montre qu'il y a eu certainement des réarrangements chromosomiques ancestraux à l'intérieur de cette extrémité télomérique. Ces réarrangements se sont produits après la duplication entière du génome, il y a 65 millions d'années. En effet, les gènes des différents clusters ne se retrouvent que rarement à l'extérieur de cette zone génomique. Ensuite, les gènes ont évolué à l'intérieur des clusters.

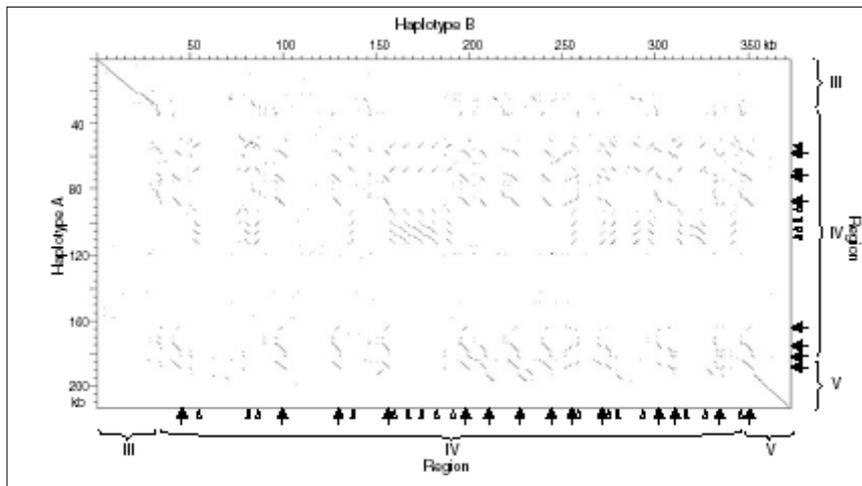
##### 4.2.1.1 Les clusters de TNL

Les gènes TNL semblent évoluer par des duplications en tandem car les gènes d'un cluster sont dans la même orientation. Ces duplications génèrent de nouveaux gènes à l'intérieur du même cluster. La vitesse d'évolution de chaque cluster semble différente car ils ont un



D'après Baurens *et al.*, 2010

**Figure 63** : Comparaison entre deux haplotypes au locus RGA08 chez la banane.



D'après Kuang *et al.*, 2005

**Figure 64** : Comparaison entre deux haplotypes au locus *RI* chez la pomme de terre.

nombre variable de gènes. Chaque cluster de TNL représente un clade particulier ou un petit nombre de clades.

#### 4.2.1.2 Les clusters de BNL

Pour les gènes BNL, les mécanismes évolutifs sont plus difficiles à définir ; en effet, les différents clusters ne sont pas clairement définis par clade et des gènes proches en séquence sont parfois placés dans des clusters différents. Néanmoins, les gènes dans un même cluster sont dans la même orientation ; ils semblent donc évoluer eux aussi par des duplications en tandem.

Les BNLs du chromosome 19 sont plus semblables entre eux, qu'avec ceux situés sur les autres chromosomes (Germain *et al*, 2010).

#### 4.2.2 Clade 1 des TNL

Le cluster de gènes TNL du clade 1 montre, au centre du cluster, une région très variable sur les séquences inter-géniques et une forte redondance des gènes entre les deux haplotypes. A l'extérieur du cluster, il y a la présence de régions génomiques orthologues. Ce type de répartition pour des clusters de gènes NBS-LRR a déjà été remarqué chez d'autres espèces telle que la banane et la pomme de terre (figure 63 ; Baurens *et al*, 2010 ; figure 64 ; Kuang *et al*, 2005).

Dans le cluster de gène dont fait partie  $R_{US}$ , le polymorphisme se concentre dans les zones inter-géniques. Au centre des clusters, les recombinaisons se font certainement entre les gènes de résistance non orthologues : c'est la seule possibilité de recombinaison homologue. En effet, l'homologie entre les gènes est très importante (Cf. figure 48, page IV-94). Ces recombinaisons conduisent à la création de nouveaux allèles. La recombinaison non orthologue inter-génique est un moyen pour le peuplier de créer rapidement des allèles très différents, en réponse à l'évolution rapide de l'agent pathogène.

Mais il y a aussi des recombinaisons non homologues qui aboutissent à des gènes atypiques qui sont constitués d'une suite particulière de domaines. Le gène  $R_{TNT14}$  possède un domaine TIR, puis un domaine NBS partiel et, enfin, de nouveau un domaine TIR. D'autres gènes ne possèdent que le domaine NBS ou que des LRR. Est-ce que ce sont des gènes qui sont exprimés et traduits et dont la protéine peut avoir une action, ou bien, est-ce que ce sont des pseudogènes ?

Le meilleur exemple est que le seul individu recombinant d'un côté de  $R_{US}$ , le 66303763, est certainement une recombinaison non homologue : il porte les deux allèles du marqueur INRA\_303 (Cf. figure 51, page IV-98). Cela ne change rien à la localisation du locus ; en effet,



l'individu 66303763 est de phénotype  $r_{US}$ , le gène n'est donc pas sur la partie de l'haplotype  $hR_{US}$  dont il a hérité (Cf. figure 50, page IV-97). Il montre un exemple de recombinaison non orthologue qui peut se produire dans des zones très fortement hétérozygotes.

Avec énormément de polymorphisme entre les deux haplotypes, on peut se demander où se situent les événements de recombinaison, mis à part dans les gènes de résistance. Des recombinaisons inégales doivent se produire entre les deux haplotypes, notamment au niveau des gènes de résistance : c'est-à-dire une recombinaison entre des allèles de différents gènes. Dans cette hypothèse, cela créera des nouveaux gènes avec une séquence chimérique. Dans certains cas, des recombinaisons peuvent créer des nouveaux gènes avec des ordonnancements différents entre les domaines TIR, NBS et LRR ou des gènes partiels. Les espèces pérennes sont plus sujettes aux attaques par des agents pathogènes avant leur reproduction et avec leur longue durée de génération, il est donc impossible pour eux d'égaliser le taux d'évolution d'un agent pathogène qui passe par plusieurs générations chaque année. Cette caractéristique peut avoir favorisé l'expansion des gènes de résistance.

#### **4.2.3 Le cluster de BNL où se trouve $R_1$**

La taille du cluster où se trouve  $R_1$  est différente entre l'haplotype  $hR_{US}$  et Nisqually-1, due à un nombre de gènes différents mais aussi aux ETs.

On remarque qu'en dessinant des couples d'amorces sur Nisqually-1 dans la zone de  $R_1$ , nous obtenons plus de marqueurs qui fonctionnent sur l'haplotype  $hr_1$  que sur l'haplotype  $hR_1$ .

### **4.3 Les éléments transposables**

Chez la banane, les ETs sont plus présents dans les clusters de gène de résistance que sur les régions voisines (Baurens *et al*, 2010). De plus, la zone d'intérêt possède des éléments transposables, qui pourraient être un moteur supplémentaire d'événements de duplication en tandem.



**Partie VI**  
**PERSPECTIVES GENERALES**



## Partie VI

### PERSPECTIVES GENERALES

<b>1</b>	<b>Validation fonctionnelle du gène candidat <math>R_{US}</math>.....</b>	<b>125</b>
<b>2</b>	<b>Compréhension des mécanismes de cette résistance .....</b>	<b>126</b>
2.1	Etude de l'expression du gène candidat .....	126
2.2	Domaine NLS .....	126
2.3	Domaine LRR.....	127
2.4	Etude des éléments transposables.....	127
2.5	Génotype [ $R_{US}/R_{US}$ ] .....	127
<b>3</b>	<b>L'interaction peuplier-<i>Mlp</i> .....</b>	<b>128</b>
<b>4</b>	<b>Etude du locus <math>R_{US}</math> dans les populations naturelles de <i>P. trichocarpa</i> .....</b>	<b>128</b>
<b>5</b>	<b>Vers le clonage positionnel de <math>R_1</math>.....</b>	<b>129</b>
5.1	Développement de nouveaux marqueurs génétiques .....	129
5.2	Clonage physique de l'allèle $R_1$ .....	129
<b>6</b>	<b>Y a-t-il d'autres gènes proches des loci responsables de résistance à <i>Mlp</i> ? .....</b>	<b>130</b>
6.1	QTL .....	130
6.2	Etude des individus porteurs d'une recombinaison proche des loci de résistance	131
<b>7</b>	<b>L'inventaire des gènes de résistance à <i>Mlp</i> .....</b>	<b>131</b>
7.1	Les marqueurs utiles.....	131
7.2	Le locus <i>MER</i> .....	131
7.3	Caractériser de nouveaux loci de résistance .....	132
<b>8</b>	<b>Séquençage de la région génomique couvrant les gènes <math>R_1</math> et <math>R_{US}</math></b>	<b>132</b>
<b>9</b>	<b>Gestion des peupleraies cultivées.....</b>	<b>133</b>
9.1	A court terme .....	133
9.2	A long terme .....	133



# 1 Validation fonctionnelle du gène candidat $R_{US}$

Pour vérifier que le gène candidat  $R_{nTNL98}$  est bien le gène responsable de la résistance, il est nécessaire d'obtenir une preuve irréfutable qui se fera par l'intermédiaire d'une transformation génétique.

Une première stratégie serait de réprimer cet allèle pour un individu de génotype [ $R_{US}$ ]. L'obtention d'un phénotype  $r_{US}$  serait la preuve de la réussite du clonage de  $R_{US}$  et surtout de la fonction de celui dans la résistance. Une technique pour réprimer cet allèle serait l'ARN interférence (ARNi). Mais avec la redondance importante du clade de gènes dont fait partie le gène candidat, il se pourrait que l'ARNi puisse réprimer d'autres gènes. Réprimer l'allèle candidat spécifiquement semble donc compliqué.

Les perspectives pour la validation fonctionnelle se sont tournées alors vers l'introduction de l'allèle candidat par transgénèse dans des génotypes sensibles qui portent donc de gros sores. L'obtention d'un individu de génotype [ $r_{US}$ ] dans lequel a été introduit l'allèle candidat par transgénèse et dont le phénotype serait  $R_{US}$ , serait la preuve de la réussite du clonage positionnel du locus  $R_{US}$ . La transformation génétique chez le peuplier est possible mais elle reste difficile et est dépendante du génotype qui est utilisé.

Pour atteindre ce but, l'idéal serait de faire la transformation génétique avec un fragment d'ADNg natif, c'est-à-dire avec la séquence du gène avec ses introns et son promoteur. Dans ce but, des fragments de 10 Kb utilisés pour le séquençage des clones BAC porteurs de  $R_{US}$  (04J18 et 97F12), ont été sélectionnés pour la transformation génétique d'un individu descendant de la famille 54 de génotype et phénotype  $r_{us}$ .

Des difficultés techniques ont été rencontrées à l'INRA d'Orléans pour la transformation génétique chez un descendant de la famille 54 et pour l'intégration d'un fragment de 11 Kb dans un vecteur binaire (Courteille, 2009). C'est pourquoi, la transformation génétique devrait se faire sur un individu qui n'est pas de la famille 54 mais sensible à  $Mlp$ , et que le transgène utilisé devrait être une séquence synthétisée correspondant au CDS du gène candidat avec le promoteur natif.

Pour étudier les génotypes dans lesquels on a introduit l'allèle candidat  $R_{US}$ , les marqueurs développés sur le CDS du gène spécifiquement et utilisés pour l'étude de l'expression, seront utiles pour vérifier l'expression de l'allèle introduit.

Ferrier-Cana *et al*, (2005) ont montré des modifications post-transcriptionnelles chez un gène de type CNL qui donne plusieurs transcrits alternatifs. Dans ce but il serait intéressant de



récupérer l'ADNc du gène candidat afin de vérifier le CDS, et donc obtenir une annotation parfaite du gène candidat. En effet les EST qui ont été utilisés pour l'annotation ne montre pas toujours une très forte identité, il s'agit donc d'indices mais pas de preuve de la structuration en intron et en exon (Cf. § IV-1.3.1). La partie la plus problématique est l'extrémité 3' du gène. Une démarche de 3' RACE serait donc utile pour s'assurer de cette annotation. Mais ce genre de démarche n'est pas forcément utilisable sur l'ensemble du gène du fait de sa forte redondance avec les autres gènes du même clade.

## 2 Compréhension des mécanismes de cette résistance

### 2.1 Etude de l'expression du gène candidat

Pour étudier l'expression du gène candidat, il serait intéressant d'étudier le promoteur du gène candidat pour connaître qu'est-ce qui induit son expression. Premièrement *in silico*, pour connaître de probables inducteurs agissant sur le promoteur. Deuxièmement en étudiant l'expression du gène candidat après d'autres traitements que l'inoculation de *Mlp*. Troisièmement en étudiant l'expression du gène chez le tabac avec son promoteur plus ou moins découpé pour en déterminer expérimentalement les régions promotrices importantes dans l'induction du gène.

Il serait intéressant de comparer l'expression du gène candidat avec d'autres gènes du même clade de TNL. En effet, comme montré précédemment, l'unique comparaison avec l'allèle  $r_{US}$  n'est pas une solution, elle doit être faite avec plusieurs gènes du même clade. Est-ce que seuls les gènes NBS-LRR nécessaires à la résistance sont transcrits quand il y a inoculation ou est-ce que de nombreux gènes NBS-LRR sont transcrits lorsque que la plante perçoit une attaque d'un agent pathogène et cela quelque soit l'agent pathogène ? De plus, pour connaître ce qu'il y a d'important dans les inducteurs de l'expression du gène, sa localisation cellulaire et tissulaire, la comparaison avec plusieurs autres gènes proches en séquence permettrait d'apprendre un peu plus sur la résistance liée au locus  $R_{US}$ .

### 2.2 Domaine NLS

Quand il sera obtenu la preuve irréfutable que cet allèle candidat est bien responsable de la résistance par transformation génétique, il serait intéressant de faire une autre transformation génétique en utilisant le gène candidat sans le domaine NLS, ce qui permettrait de savoir si la localisation cellulaire est importante dans le mécanisme de résistance. Cette démarche est rendue



possible par la présence, dans le gène candidat, du codon d'initiation de la traduction (ATG) présent dans les gènes du même clade qui ne porte pas ce domaine NLS. Il serait donc possible de synthétiser un ADNc du gène candidat sans le domaine NLS et de comparer le résultat phénotypique de ces deux transformations.

## 2.3 Domaine LRR

La comparaison entre la séquence de l'allèle candidat avec les séquences de tous les autres gènes du clade 1 des gènes TNL pourraient donner des indices intéressants. Notamment sur le domaine LRR, il serait intéressant de trouver les SNPs qui sont spécifiques à l'allèle  $R_{US}$ . Mais ceci n'est pas une certitude, en effet la résistance de l'allèle  $R_{US}$  est peut être due à un enchaînement particulier des différents domaines.

La synthèse de différents CDSs porteurs de SNPs avant de les introduire par transgénèse permettrait de connaître les zones du gène qui sont cruciales dans le mécanisme de résistance.

## 2.4 Etude des éléments transposables

Il faut aussi s'interroger sur l'effet des éléments transposables présents dans le cluster des gènes de résistance. Par exemple, Hayashi & Yoshida, (2009) ont montré qu'un ET à LTR avait une action régulatrice sur un gène NBS-LRR chez le riz. Ainsi le gène  $Pit^{K59}$  était surexprimé en présence de cet ET en amont du gène.

Pour la zone génomique où se trouve  $R_{US}$  et  $R_1$ , on n'y relève que des éléments transposables. Il est donc possible qu'il y ait un effet de ces éléments transposables sur la régulation des gènes conférant la résistance à  $Mlp$ .

Toutefois, il est à souligner que l'étude sur les éléments transposables présentée dans cette thèse n'est pas exhaustive. Elle ne relève que d'alignements des séquences des clones BAC avec la base d'éléments répétés RepBase. L'étude des ETs à proximité des loci  $R_1$  et  $R_{US}$  demande donc à être approfondie par une annotation précise. Elle permettra de préciser à la fois l'importance de ces ETs dans l'évolution de ce super cluster de gène de résistance ainsi que la possible action sur la régulation de ces gènes.

## 2.5 Génotype [ $R_{US}/R_{US}$ ]

Un croisement entre deux individus descendants du père 101-74 a été réalisé, dans le but d'obtenir des individus homozygotes pour l'allèle  $R_{US}$ , donc de génotype [ $R_{US}/R_{US}$ ]. Quel va être le phénotype de ces individus : est-ce que l'action de deux allèles  $R_{US}$  est additive ?



Pour étudier ces descendants, il serait intéressant de les génotyper avec quelques marqueurs génétiques pour définir les trois classes d'individus :  $\frac{1}{4}$  d'individus [ $R_{US}/R_{US}$ ],  $\frac{1}{2}$  d'individus [ $R_{US}/r_{US}$ ],  $\frac{1}{4}$  d'individus [ $r_{US}/r_{US}$ ]. Mais pour que ce génotypage soit correct, il est nécessaire d'utiliser des marqueurs co-dominants ou des marqueurs dominants dans les deux phases pour séparer correctement les trois classes. La comparaison entre le phénotype des individus des classes [ $R_{US}/R_{US}$ ] et [ $R_{US}/r_{US}$ ] permettra de déterminer l'action de deux allèles  $R_{US}$  ensemble.

### **3 L'interaction peuplier-*Mlp***

Lorsque le gène candidat sera validé, il restera à démontrer la nature exacte de l'interaction peuplier – *Mlp*. Une expérience de double hybride peut être envisagée pour piéger à partir du gène  $R_{US}$  des protéines d'avirulence. Mais si l'interaction correspond au modèle de garde, il n'y aura certainement pas de résultat. En effet, peu d'interaction entre gène R et gène Avr a pu être démontrée. Il est aussi intéressant de les étudier pour comprendre l'interaction et l'évolution entre les gènes de résistance du peuplier et les gènes d'avirulence de *Mlp*. Avec la séquence de *Mlp* d'autres voies sont ouvertes. Par le clonage des gènes de virulence/avirulence, l'étude de l'interaction pourra commencer.

## **4 Etude du locus $R_{US}$ dans les populations naturelles de *P. trichocarpa***

Il s'agit d'étudier le polymorphisme de la zone où le locus  $R_{US}$  est localisé dans des populations naturelles de *P. trichocarpa*. Une collection de plus de 300 individus est disponible à l'INRA d'Orléans. Cette population a été phénotypée ce qui permettra d'associer phénotype et génotype. Cette démarche avait été envisagée dans le cas où la carte génétique n'avait pas permis de déterminer un seul gène. Mais étant donnée la redondance des gènes, il est à craindre que les marqueurs autour de  $R_{US}$  amplifieront d'autres locus. A l'inverse, beaucoup de marqueurs n'amplifieront certainement pas tous les individus de la population naturelle. Pour ces derniers marqueurs, le séquençage de produit d'amplification PCR ne sera pas exploitable, il serait possible seulement d'exploiter la présence ou l'absence d'amplification. Avec la structure particulière de cette région, on peut s'attendre à ne pas avoir un faible déséquilibre de liaison ; le génotypage avec des marqueurs de la région de  $R_{US}$  ne donnera sûrement que peu de résultats intéressants.



Mais il serait intéressant d'utiliser les marqueurs dessinés sur le gène candidat  $R_{nTNL98}$  pour obtenir un indice complémentaire sur la responsabilité de  $R_{US}$  dans la résistance à *Mlp* et/ou connaître la présence d'un tel gène de résistance dans une population naturelle. Avec la recombinaison génique qui est importante dans ce cluster de gènes, il n'est pas évident qu'il y ait des résultats pour n'importe quel marqueur dessiné dans ce gène. A l'inverse, si il y a une zone précise dans le gène qui est important dans la résistance, peut-être qu'un marqueur précis dessiné dans cette région du gène donnera un résultat intéressant et important. Dans cette idée, il est peut être utile d'élaborer le maximum d'amorces spécifiques des différentes régions de ce gène notamment sur la région variable du domaine LRR.

## 5 Vers le clonage positionnel de $R_1$

Ce travail de thèse a permis de localiser plus finement le locus  $R_1$  dans un intervalle de 0.66 cM. Ce travail de thèse a donc permis de se rapprocher du locus  $R_1$ .

### 5.1 Développement de nouveaux marqueurs génétiques

L'intervalle où se trouve le locus  $R_1$  est de 0,66 cM et il est possible de l'améliorer : 14 individus possèdent une recombinaison dans l'intervalle de ce locus.

Le développement de nouveaux marqueurs génétiques sur la séquence du clone BAC 97G09 de l'haplotype  $hR_{US}$  du 101-74 et sur les séquences des clones BAC de l'haplotype  $hr_{US}$ , bientôt disponibles (Cf. § VI-8), permettra peut-être de développer des nouveaux marqueurs génétiques pour le 73028-62. Certains marqueurs disponibles marchent sur les deux haplotypes de la mère 73028-62. Le séquençage de produit PCR permettra certainement de caractériser des SNPs utiles pour le développement de nouveaux marqueurs génétiques. Le clonage des produits d'amplification PCR sera sûrement nécessaire afin d'avoir des résultats de séquençage de bonne qualité.

Cependant il est fort peu probable que le travail sur l'espèce *P. trichocarpa* aboutisse à l'identification d'un gène de résistance hérité de *P. deltoides*.

### 5.2 Clonage physique de l'allèle $R_1$

Une première solution serait d'utiliser la PCR sur de l'ADNg avec la Taq polymérase long range pour amplifier le gène chez le 73028-62. Mais il faudrait pouvoir posséder des amorces spécifiques entourant le gène (donc dans des zones inter-géniques), ce qui paraît particulièrement difficile à obtenir. En effet dans la zone de  $R_1$  le polymorphisme étant très



important, je doute qu'à partir des séquences de l'espèce *P. trichocarpa* de Nisqually-1 ou du 101-74 nous puissions obtenir ces amorces entourant le gène chez le 73028-62 de l'espèce *P. deltoides*.

Pour obtenir la séquence du gène  $R_1$ , il sera sûrement nécessaire de développer des ressources propres à l'individu 73028-62, afin d'obtenir des séquences de grande taille. La stratégie idéale serait de construire une banque BAC de cet individu, mais cette démarche est longue et coûteuse. Pour accélérer cette démarche, il est possible de ne pas organiser la banque BAC en clones individuels mais de l'organiser en pool de plusieurs centaines de clones qui seront criblés par PCR avec les marqueurs flanquants le locus (Isidore *et al*, 2005). Les pools positifs seront alors organisés en clones individuels afin de trouver le clone porteur du gène d'intérêt. Il n'est donc pas nécessaire d'organiser complètement la banque BAC en plaque 384.

Par ailleurs le développement d'une ressource BAC chez l'espèce *P. deltoides* porteuse de plusieurs gènes de résistance qualitative et de QTLs, serait une grande aide dans le développement de la thématique qui vise à décrire le cluster de gènes de résistance et ainsi comprendre son fonctionnement. Enfin l'approche de génomique comparative inter-espèce avec des séquences des clones BAC amènera probablement des éléments intéressants pour la description des évènements évolutifs des gènes de résistance de l'extrémité du chromosome 19.

## **6 Y a-t-il d'autres gènes proches des loci responsables de résistance à *Mlp* ?**

### **6.1 QTL**

L'analyse de QTL a été améliorée durant ma thèse (Cf. figure 65, page V-117). Il montre la présence de différents QTLs contrôlant différents paramètres quantitatifs sur le LG XIX chez le 101-74. La question est de savoir s'il n'y a pas un autre gène qui est présent dans la région génomique de  $R_1$  et  $R_{US}$ . En effet, il y a un nombre considérable de gènes NBS-LRR qui sont susceptibles de contrôler un des paramètres quantitatifs de la résistance à *Mlp*.

Il est possible de renouveler cette démarche avec les dernières cartes génétiques qui permettrait aussi de mieux préciser la localisation des différents QTL.

Il serait intéressant d'analyser en QTL les données pour l'ensemble des individus de la famille 54. Mais ces individus ont été phénotypés dans différentes expériences qui ne donnent pas exactement les mêmes résultats. Nous pouvons le voir dans la figure 35 (Cf. page III-77) où la taille des sores n'est pas parfaitement équivalente entre les différentes expériences et donc les



limites entre le classement des différents individus sont différentes. Il y a donc une difficulté pour analyser correctement les données de différentes expériences de phénotypage ensemble.

## **6.2 Etude des individus porteurs d'une recombinaison proche des loci de résistance**

Dans le cas où il y a d'autres loci impliqués dans les différents paramètres quantitatifs de la résistance à *Mlp*, l'étude des individus porteurs d'une recombinaison proche des loci est le seul moyen d'étudier ces différents loci séparément.

S'il y a différents loci, il est possible que leur spécificité par rapport aux virulences soit différente, ces individus recombinants permettront de le révéler en les étudiant après inoculation avec différentes souches.

# **7 L'inventaire des gènes de résistance à *Mlp***

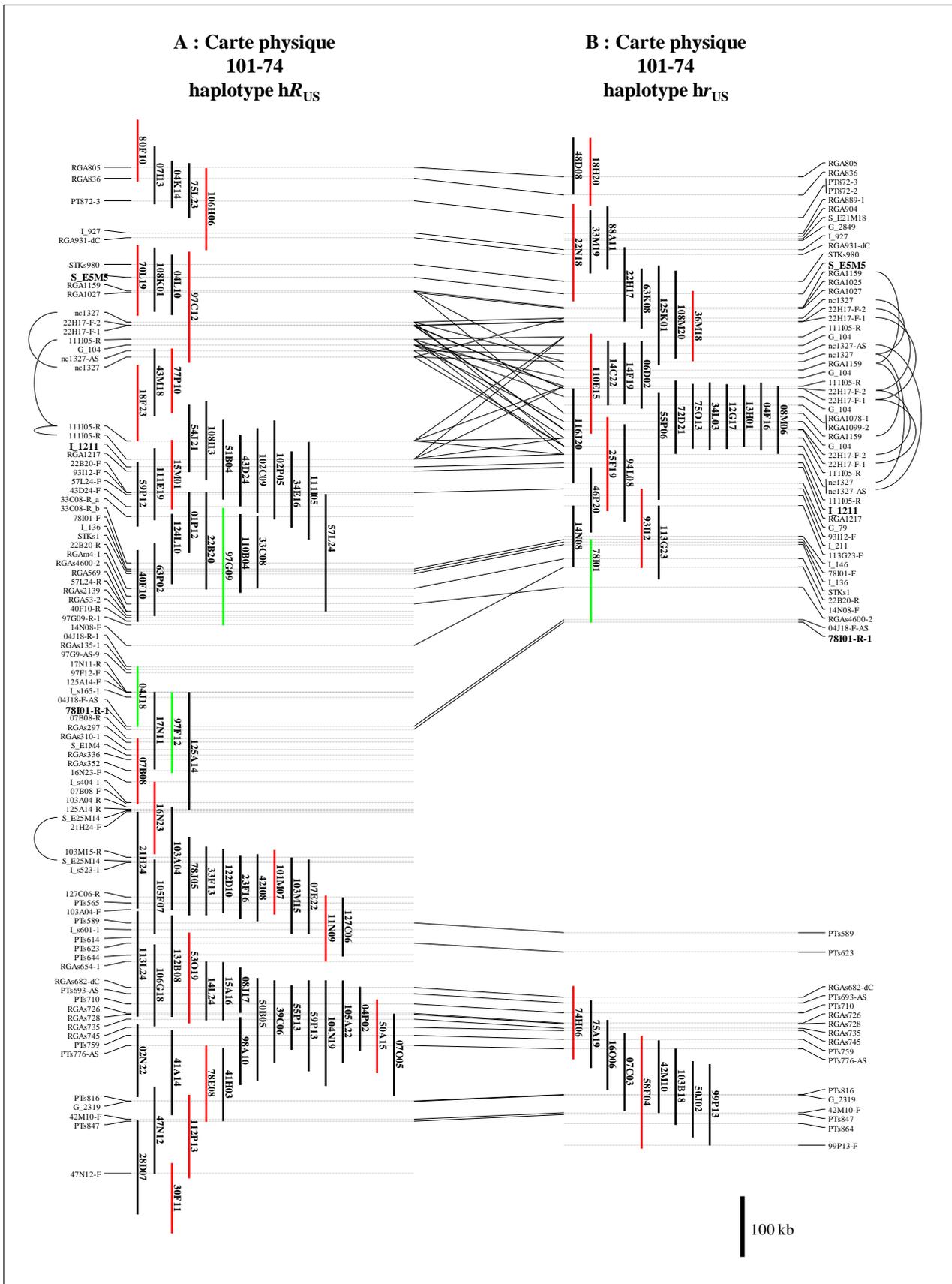
## **7.1 Les marqueurs utiles**

Avec les marqueurs développés et le potentiel de gène de résistance sur l'extrémité du chromosome 19, il serait maintenant plus facile de découvrir et de cloner d'autres gènes dans cette région génomique.

Dans ce super cluster, il serait préférable d'utiliser les marqueurs présents entre les clusters de gènes. En effet, ces régions montrent une complexité moins importante, ce qui donne plus de chance aux marqueurs de fonctionner sur beaucoup d'individus et moins de chance d'être multi-locus. Ils permettront donc un meilleur travail de routine pour plusieurs pedigrees à la fois. Des marqueurs sont déjà disponibles mais ils peuvent être améliorés avec les nouvelles séquences bientôt disponibles (Cf. § VI-8). La sélection des marqueurs SSR par leur simplicité d'utilisation mais aussi développer des marqueurs où il y a des InDels de taille intéressant et variable chez les différents haplotypes. Après quoi, il sera possible d'identifier les autres gènes de résistance dans un cluster précis en utilisant les marqueurs dans ce cluster.

## **7.2 Le locus *MER***

De nombreux marqueurs développés dans la région génomique de  $R_1$  seraient intéressants pour améliorer la cartographie génétique du locus *MER* qui semble être dans le même cluster de BNLs que  $R_1$  (Cf. § V-3.4). Il reste 1 cM pour plus de 500 individus et donc il reste des



**Figure 65** : Les clones BAC en cours de séquençage.  
 A : carte physique de l'haplotype hR<sub>US</sub>, B : carte physique de l'haplotype hr<sub>US</sub>.  
 Les clones BAC en vert sont séquencés, les clones BAC en rouge sont en cours de séquençage.

recombinants à génotyper dans le pedigree où S9\_2 est un des parents ; il y a moyen de progresser dans la localisation du locus *MER*. Les marqueurs génétiques de la cartographie fine du 73028-62 qui sont transférables pour la cartographie de l'individu S9\_2 porteur du locus *MER* serait intéressant pour connaître les relations entre le locus *MER* et le locus *R*<sub>1</sub>. L'élaboration de marqueurs sur les séquences du génome de Nisqually-1, maintenant disponible, peut aussi être utile.

### **7.3 Caractériser de nouveaux loci de résistance**

Il se pose également la question de l'intérêt des gènes hérités du peuplier noir qui, eux, ont longtemps co-évolué avec la rouille. Une population F<sub>1</sub> issue d'un croisement entre deux individus *P. nigra* est disponible à l'INRA d'Orléans dans laquelle de la variabilité dans les paramètres quantitatifs a été observé. Il serait intéressant d'utiliser les marqueurs développés dans cette étude pour faire une analyse approfondie de QTLs dans cette région génomique.

De même, il serait intéressant de chercher les marqueurs transférables à la cartographie des différentes familles du plan factoriel 9 x 9 (Villar *et al*, 1996).

## **8 Séquençage de la région génomique couvrant les gènes *R*<sub>1</sub> et *R*<sub>US</sub>**

Comme la carte physique est développée pour la région des loci *R*<sub>1</sub> et *R*<sub>US</sub>, le séquençage de clones BAC de la région génomique couvrant les deux loci permettra d'obtenir une meilleure qualité de séquence que celle de Nisqually-1 obtenue en WGS. Les erreurs d'assemblage causées par l'hétérozygotie ne devraient pas se produire avec les clones BAC, et les problèmes dus à la redondance seront minimisés par la stratégie BAC à BAC. Une sélection de 24 clones BAC, représentant au mieux les deux haplotypes du 101-74, est en cours de séquençage (figure 65). Cette sélection regroupe près de 1,7 Mb de l'haplotype h*R*<sub>US</sub> et 1,3 Mb pour l'haplotype h*r*<sub>US</sub>. Elle correspond à une partie majeure du super cluster de gène NBS-LRR. Ce séquençage sera très intéressant pour estimer le niveau d'hétérozygotie apparemment important entre les deux haplotypes.

Le séquençage de ces BAC permettra de vérifier les incertitudes qui existent dans les zones 2 et 4 décrites dans la figure 44, page III-86.

L'annotation complète et expertisée des gènes NBS-LRR permettrait de savoir s'il s'agit de gènes ou de pseudo-gènes. L'utilisation des ESTs disponibles dans les bases de données doit



constituer un appui pour cette tâche pour identifier les gènes fonctionnels. Le clusterigie des ces EST par clade pour les aligner avec les gènes permettra une annotation plus rapide de ces gènes.

La comparaison des gènes des deux haplotypes du 101-74 et de ceux de Nisqually-1 permettra d'émettre des hypothèses plus approfondies quant à l'évolution des gènes de résistance à trois échelles différentes : celle du gène, celle des clusters et celle des clusters à l'intérieur du super cluster.

## 9 Gestion des peupleraies cultivées

### 9.1 A court terme

Il faudra encore beaucoup de temps pour parvenir à une réponse en termes de gestion durable. A très court terme, on ne peut qu'évoquer des règles de bon sens : éviter la présence conjointe de peupliers et de mélèzes propice au développement du champignon et diversifier les variétés utilisées en populiculture. Pour ce dernier but, l'inventaire des gènes de résistance est donc souhaitable (Cf. § VI-7.3), tant pour conforter les stratégies en sélection que pour prédire les gènes d'avirulence du parasite.

Ces gènes contournés n'ont pas un intérêt total dans une peupleraie cultivée, mais un intérêt partiel ; elle freine la contamination. De plus, introduire un gène déjà contourné ne déclenchera pas un nouveau contournement, par contre, elle permet de renforcer la sélection de souche de *Mlp* qui porte une virulence contournant ces gènes de résistance.

### 9.2 A long terme

A plus long terme, il faut essayer d'anticiper les mécanismes de contournement, avant de déployer de nouveaux gènes de résistance ; la compréhension des mécanismes de l'interaction entre l'hôte et l'agent pathogène permettra, je l'espère, de comprendre les mécanismes aboutissant aux contournements et donc prédire les gènes les plus adaptés à une gestion durable. Dans ce but, l'étude de gènes déjà contournés est indispensable pour en comprendre leurs contournements.

L'identification de des gènes d'avirulence de *Mlp* permettra d'étudier plus en profondeur, ce qui différencie une souche contournante et une autre.



# ABBREVIATIONS

ADNc : <u>A</u> DN <u>c</u> omplémentaire	m : <u>m</u> ètre
AFLP : <u>a</u> mplified <u>f</u> ragment <u>l</u> ength polymorphism	MAMP : <u>m</u> icrobe- <u>a</u> ssociated <u>m</u> olecular <u>p</u> attern <i>Mlp</i> : <i>Melampsora larici-populina</i>
ARNi : <u>A</u> RN <u>i</u> nterférence.	NCBI : National Center for Biotechnology Information
AS-PCR : <u>a</u> llele <u>s</u> pecific <u>P</u> CR	NBS : <u>n</u> ucleotide <u>b</u> inding site
BAC : <u>b</u> acterial <u>a</u> rtificial <u>c</u> hromosome	NLS : <u>n</u> uclear <u>l</u> ocalization <u>s</u> ignal
BES : <u>B</u> AC <u>e</u> nd sequence	QTL : <u>q</u> uantitative <u>t</u> rait <u>l</u> ocus
BNL : <u>B</u> ED <u>N</u> BS <u>L</u> RR	PAMP : <u>p</u> athogen- <u>a</u> ssociated <u>m</u> olecular <u>p</u> attern
BSA : <u>b</u> ulked <u>s</u> egregant <u>a</u> nalysis	PASA : <u>P</u> CR <u>a</u> llele- <u>s</u> pecific <u>a</u> mplification
CAPS : <u>c</u> leaved <u>a</u> mplified <u>p</u> olymorphic sequence	PRR : <u>p</u> attern <u>r</u> ecognition <u>r</u> ecceptor
CC : <u>c</u> oiled- <u>c</u> oil	PCR : <u>p</u> olymerase <u>c</u> hain <u>r</u> eaction
cM : <u>c</u> enti <u>M</u> organ	PTI : <u>P</u> AMP- <u>t</u> riggered <u>i</u> mmunity
CNL : <u>C</u> C <u>N</u> BS <u>L</u> RR	RLK : <u>r</u> ecceptor- <u>l</u> ike <u>k</u> inase
dCAPS : <u>d</u> erived <u>C</u> APS	RLP : <u>r</u> ecceptor- <u>l</u> ike <u>p</u> rotein
DAMP : <u>d</u> amage- <u>a</u> ssociated <u>m</u> olecular <u>p</u> attern	rpm : <u>r</u> otation <u>p</u> ar <u>m</u> inute
EST : <u>e</u> xpressed <u>s</u> equence <u>t</u> ag	SCAR : <u>s</u> equence <u>c</u> haracterized <u>a</u> mplified region
ET : <u>é</u> lément <u>t</u> ransposable	SNP : <u>s</u> ingle <u>n</u> ucleotide <u>p</u> olymorphism
ETI : <u>e</u> ffector- <u>t</u> riggered <u>i</u> mmunity	SSR : <u>s</u> imple <u>s</u> equence <u>r</u> epeat
ETS : <u>e</u> ffector- <u>t</u> riggered <u>s</u> usceptibility	STK : <u>s</u> erine - <u>t</u> hréonine <u>k</u> inase
GFP : <u>g</u> reen <u>f</u> luorescent <u>p</u> rotein	STS : <u>s</u> equence <u>t</u> ag <u>s</u> ite
ha : hectare	TNL : <u>T</u> IR <u>N</u> BS <u>L</u> RR
InDel : <u>i</u> nsertion/ <u>d</u> élétion	TIR : <u>T</u> oll- <u>I</u> nterleukin <u>r</u> ecceptor
JGI : <u>J</u> oint <u>G</u> enome <u>I</u> nstitute	UN : <u>u</u> redia <u>n</u> umber
LRR : <u>l</u> eucine <u>r</u> ich <u>r</u> epeat	US : <u>u</u> redia <u>s</u> ize
LP : <u>l</u> atent <u>p</u> eriod	WGS : <u>w</u> hole <u>g</u> enome <u>s</u> hotgun
LTR : <u>l</u> ong <u>t</u> erminal <u>r</u> epeat	

# BIBLIOGRAPHIE

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997.** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25: 3389-3402.
- Ameline-Torregrosa C, Cazaux M, Danesh D, Chardon F, Cannon SB, Esquerré-Tugayé MT, Dumas B, Young ND, Samac DA, Huguet T, Jacquet C. 2008.** Genetic dissection of resistance to anthracnose and powdery mildew in *Medicago truncatula*. *Molecular Plant–Microbe Interactions* 21: 61-69.
- Aravind L. 2000.** The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends in Biochemical Sciences* 25: 421-423.
- Arcade A, Labourdette A, Falque M, Mangin B, Chardon F, Charcosset A, Joets J. 2004.** BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics* 20: 2324-2326.
- Ashikawa I, Hayashi N, Yamane H, Kanamori H, Wu J, Matsumoto T, Ono K, Yano M. 2008.** Two adjacent nucleotide-binding site-leucine-rich repeat class genes are required to confer *Pikm*-specific rice blast resistance. *Genetics* 180: 2267-2276.
- Baurens FC, Bocs S, Rouard M, Matsumoto T, Miller RN, Rodier-Goud M, Mbeguie-A-Mbeguie D, Yahiaoui N. 2010.** Mechanisms of haplotype divergence at the RGA08 nucleotide-binding leucine-rich repeat gene locus in wild banana (*Musa balbisiana*). *BMC Plant Biology* 10: 149.
- Bent AF, Mackey D. 2007.** Elicitors, effectors, and R genes: the new paradigm and a lifetime supply of questions. *Annual Review of Phytopathology* 45: 399-436.
- Berthelot A, Chantre G. 2004.** La peupleraie française : un maillon essentiel de la filière bois. *Biofutur* 247: 20-23.
- Boller T, Felix G. 2009.** A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual Review of Plant Biology* 60: 379-406.
- Bradshaw HD, Stettler RF. 1993.** Molecular genetics of growth and development in *Populus*. I. Triploidy in hybrid poplars. *Theoretical and Applied Genetics* 86: 301–307.
- Bundock PC, Cross MJ, Shapter FM, Henry RJ. 2006.** Robust allele-specific polymerase chain reaction markers developed for single nucleotide polymorphisms in expressed barley sequences. *Theoretical and Applied Genetics* 112: 358-365.
- Cartwright DA, Troglio M, Velasco R, Gutin A. 2007.** Genetic mapping in the presence of genotyping errors. *Genetics* 176: 2521-2527.
- Cervera MT, Gusmão J, Steenackers M, Peleman J, Storme V, Vanden Broeck A, Van Montagu M, Boerjan W. 1996.** Identification of AFLP molecular markers for resistance against *Melampsora larici-populina* in *Populus*. *Theoretical and Applied Genetics* 93: 733-737.

- Cervera MT, Storme V, Ivens B, Gusmão J, Liu BH, Hostyn J, Van Slycken J, Van Montagu M, Boerjan W. 2001.** Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) on AFLP and microsatellite markers. *Genetics* 158: 787-809.
- Cervera MT, Storme V, Soto A, Ivens B, Van Montagu M, Rajora OP, Boerjan W. 2005.** Intraspecific and interspecific genetic and phylogenetic relationships in the genus *Populus* based on AFLP markers. *Theoretical and Applied Genetics* 111: 1440-1456.
- Chalhoub B, Belcram H, Caboche M. 2004.** Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size. *Plant Biotechnology Journal* 2: 181-188.
- Chini A, Grant JJ, Seki M, Shinozaki K, Loake GJ. 2004.** Drought tolerance established by enhanced expression of the CC-NBS-LRR gene, ADR1, requires salicylic acid, EDS1 and ABI1. *Plant Journal* 38: 810-822.
- Chisholm ST, Coaker G, Day B, Staskawicz BJ. 2006.** Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* 124: 803-814.
- Coleman C, Copetti D, Cipriani G, Hoffmann S, Kozma P, Kovács L, Morgante M, Testolin R, Di Gaspero G. 2009.** The powdery mildew resistance gene *REN1* co-segregates with an NBS-LRR gene cluster in two Central Asian grapevines. *BMC Genetics* 10: 89.
- Courteille A. 2009.** Contribution à la mise au point de la transformation génétique sur les individus d'une famille *Populus deltoides* x *P. trichocarpa*, afin de valider un gène candidat pour la résistance à la rouille (*Melampsora larici-populina*). Mémoire de stage : INRA Orléans, 32 pages.
- Dangl JL, Jones JD. 2001.** Plant pathogens and integrated defence responses to infection. *Nature* 411: 826-833.
- Dixon MS, Jones DA, Keddie JS, Thomas CM, Harrison K, Jones JD. 1996.** The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* 84: 451-459.
- Dodds PN, Lawrence GJ, Catanzariti AM, Ayliffe MA, Ellis JG. 2004.** The *Melampsora lini* AvrL567 avirulence genes are expressed in haustoria and their products are recognized inside plant cells. *Plant Cell* 16: 755-768.
- Dowkiw A. 2003.** Analyse génétique de la résistance et de la tolérance de peuplier hybrides *Populus deltoides* x *Populus trichocarpa* à la rouille foliaire à *Melampsora larici-populina*. Mémoire de thèse : Université d'Orléans, 177 pages.
- Dowkiw A, Husson C, Frey P, Pinon J, Bastien C. 2003.** Partial Resistance to *Melampsora larici-populina* Leaf Rust in Hybrid Poplars: Genetic Variability in Inoculated Excised Leaf Disk Bioassay and Relationship with Complete Resistance. *Phytopathology* 93: 421-427.
- Dowkiw A, Bastien C. 2004.** Characterization of Two Major Genetic Factors Controlling Quantitative Resistance to *Melampsora larici-populina* Leaf Rust in Hybrid Poplars: Strain Specificity, Field Expression, Combined Effects, and Relationship with a Defeated Qualitative Resistance Gene. *Phytopathology* 94: 1358-1367.

- Dowkiw A, Voisin E, Bastien C. 2010.** Potential of Eurasian poplar rust to overcome a major quantitative resistance factor. *Plant Pathology* 59: 523-534.
- Drost DR, Novaes E, Boaventura-Novaes C, Benedict CI, Brown RS, Yin T, Tuskan GA, Kirst M. 2009.** A microarray-based genotyping and genetic mapping approach for highly heterozygous outcrossing species enables localization of a large fraction of the unassembled *Populus trichocarpa* genome sequence. *Plant Journal* 58: 1054-1067.
- Eckenwalder JE. 1996.** Systematics and evolution of *Populus*. Edité par Stettler RF, Bradshaw HD Jr, Heilman PE, Hinckley TM : Biology of *Populus* and its implications for management and conservation, Part I, Chapter 1. NRC Research Press, National Research Council of Canada, Ottawa. Pages 7–32.
- Faigón-Soverna A, Harmon FG, Storani L, Karayekov E, Staneloni RJ, Gassmann W, Más P, Casal JJ, Kay SA, Yanovsky MJ. 2006.** A constitutive shade-avoidance mutant implicates TIR-NBS-LRR proteins in *Arabidopsis* photomorphogenic development. *Plant Cell* 18: 2919-2928.
- Ferrier-Cana E, Macadré C, Sévignac M, David P, Langin T, Geffroy V. 2005.** Distinct post-transcriptional modifications result into seven alternative transcripts of the CC-NBS-LRR gene JA1tr of *Phaseolus vulgaris*. *Theoretical and Applied Genetics* 110: 895-905.
- Flor HH. 1971.** Current status of the gene-for-gene concept. *Annual Review of Phytopathology* 9: 275-296.
- Frey P, Pinon J. 2004.** La rouille du peuplier : un pathosystème modèle. *Biofutur* 247: 28-32.
- Gaudet M, Fara AG, Sabatti M, Kuzminsky E, Scarascia Mugnozza G. 2007.** Single-reaction for SNP Genotyping on Agarose Gel by Allele-specific PCR in Black Poplar. *Plant Molecular Biology Reporter* 25: 1-9.
- Gaudet M, Jorge V, Paolucci I, Beritognolo I, Scarascia Mugnozza G, Sabatti M. 2008.** Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. *Tree Genetics & Genomes* 4: 25–36.
- Geffroy V, Sévignac M, De Oliveira JC, Fouilloux G, Skroch P, Thoquet P, Gepts P, Langin T, Dron M. 2000.** Inheritance of partial resistance against *Colletotrichum lindemuthianum* in *Phaseolus vulgaris* and co-localization of quantitative trait loci with genes involved in specific resistance. *Molecular Plant–Microbe Interactions* 13: 287-296.
- Geffroy V, Macadré C, David P, Pedrosa-Harand A, Sévignac M, Dauga C, Langin T. 2009.** Molecular analysis of a large subtelomeric nucleotide-binding-site-leucine-rich-repeat family in two representative genotypes of the major gene pools of *Phaseolus vulgaris*. *Genetics* 181: 405-419.
- Giraut L. 2005.** Caractérisation d'une banque d'ADN en grands fragments de peuplier : *Populus trichocarpa* 101-74. Mémoire de stage : INRA d'Evry, 19 pages.
- Hamez M, Dayanandan S. 2004.** Phylogeny of *Populus* (Salicaceae) based on nucleotide sequences of chloroplast trnT-trnF region and nuclear rDNA. *American Journal of Botany* 91: 1398-1408.
- Hammond-Kosack KE, Jones JD. 1997.** PLANT DISEASE RESISTANCE GENES. *Annual Review of Plant Physiology and Plant Molecular Biology* 48: 575-607.
- Hayashi K, Yoshida H. 2009.** Refunctionalization of the ancient rice blast disease resistance gene Pit by the recruitment of a retrotransposon as a promoter. *Plant Journal* 57: 413-425.

- Hulbert SH, Webb CA, Smith SM, Sun Q. 2001.** Resistance gene complexes: evolution and utilization. *Annual Review of Phytopathology* 39: 285-312.
- Isidore E, Scherrer B, Bellec A, Budin K, Faivre-Rampant P, Waugh R, Keller B, Caboche M, Feuillet C, Chalhouh B. 2005.** Direct targeting and rapid isolation of BAC clones spanning a defined chromosome region. *Functional & Integrative Genomics* 5: 97-103.
- Jansson S, Douglas CJ. 2007.** *Populus*: a model system for plant biology. *Annual Review of Plant Biology* 58: 435-458.
- Jones JD, Dangl JL. 2006.** The plant immune system. *Nature* 444: 323-329.
- Jorge V, Dowkiw A, Faivre-Rampant P, Bastien C. 2005.** Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytologist* 167: 113-127.
- Kawchuk LM, Hachey J, Lynch DR, Kulcsar F, van Rooijen G, Waterer DR, Robertson A, Kokko E, Byers R, Howard RJ, Fischer R, Pruffer D. 2001.** Tomato Ve disease resistance genes encode cell surface-like receptors. *Proceedings of the National Academy of Sciences, USA* 98: 6511-6515.
- Kelleher CT, Chiu R, Shin H, Bosdet IE, Krzywinski MI, Fjell CD, Wilkin J, Yin T, DiFazio SP, Ali J, Asano JK, Chan S, Cloutier A, Girn N, Leach S, Lee D, Mathewson CA, Olson T, O'connor K, Prabhu AL, Smailus DE, Stott JM, Tsai M, Wye NH, Yang GS, Zhuang J, Holt RA, Putnam NH, Vrebalov J, Giovannoni JJ, Grimwood J, Schmutz J, Rokhsar D, Jones SJ, Marra MA, Tuskan GA, Bohlmann J, Ellis BE, Ritland K, Douglas CJ, Schein JE. 2007.** A physical map of the highly heterozygous *Populus* genome: integration with the genome sequence and genetic map and analysis of haplotype variation. *Plant Journal* 50: 1063-1078.
- Klevebring D, Street NR, Fahlgren N, Kasschau KD, Carrington JC, Lundeberg J, Jansson S. 2009.** Genome-wide profiling of populus small RNAs. *BMC Genomics* 10: 620.
- Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W, Martin F. 2008.** Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Molecular Biology* 66: 619-636.
- Konieczny A, Ausubel FM. 1993.** A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant Journal* 4: 403-410.
- Kuang H, Woo SS, Meyers BC, Nevo E, Michelmore RW. 2004.** Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* 16: 2870-2894.
- Kuang H, Wei F, Marano MR, Wirtz U, Wang X, Liu J, Shum WP, Zaborsky J, Tallon LJ, Rensink W, Lobst S, Zhang P, Tornqvist CE, Tek A, Bamberg J, Helgeson J, Fry W, You F, Luo MC, Jiang J, Robin Buell C, Baker B. 2005.** The R1 resistance gene cluster contains three groups of independently evolving, type I R1 homologues and shows substantial structural variation among haplotypes of *Solanum demissum*. *Plant Journal* 44: 37-51.

- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA. 1987.** MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174–181.
- Laurans F, Pilate G. 1999.** Histological Aspects of a Hypersensitive Response in Poplar to *Melampsora larici-populina*. *Phytopathology* 89: 233-238.
- Lawrence GJ, Finnegan EJ, Ayliffe MA, Ellis JG. 1995.** The *L6* gene for flax rust resistance is related to the Arabidopsis bacterial resistance gene *RPS2* and the tobacco viral resistance gene *N*. *Plant Cell* 7: 1195-1206.
- Lefèvre F, Pichot C, Pinon J. 1994.** Intra- and interspecific inheritance of some components of the resistance to leaf rust (*Melampsora larici-populina* Kleb.) in poplars. *Theoretical and Applied Genetics* 88: 501-507.
- Lefèvre F, Goué-Mourier MC, Faivre-Rampant P, Villar M. 1998.** A Single Gene Cluster Controls Incompatibility and Partial Resistance to Various *Melampsora larici-populina* Races in Hybrid Poplars. *Phytopathology* 88: 156-163.
- Lescot M, Rombauts S, Zhang J, Aubourg S, Mathé C, Jansson S, Rouzé P, Boerjan W. 2004.** Annotation of a 95-kb *Populus deltoides* genomic sequence reveals a disease resistance gene cluster and novel class I and class II transposable elements. *Theoretical and Applied Genetics* 109: 10-22.
- Martin GB, Frary A, Wu T, Brommonschenkel S, Chunwongse J, Earle ED, Tanksley SD. 1994.** A member of the tomato Pto gene family confers sensitivity to fenthion resulting in rapid cell death. *Plant Cell* 6: 1543-1552.
- Masle JP. 2005.** Cartographie fine d'un facteur de résistance quantitative à *Melampsora larici-populina* chez le Peuplier. Mémoire de stage : INRA Orléans, 91 pages.
- McHale L, Tan X, Koehl P, Michelmore RW. 2006.** Plant NBS-LRR proteins: adaptable guards. *Genome Biology* 7: 212.
- Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND. 1999.** Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant Journal* 20: 317-332.
- Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. 2003.** Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* 15: 809-834.
- Michaels SD, Amasino RM. 1998.** A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. *Plant Journal* 14: 381-385.
- Michelmore RW, Paran I, Kesseli RV. 1991.** Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences, USA* 88: 9828-9832.
- Michelmore RW, Meyers BC. 1998.** Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Research* 8: 1113-1130.
- Moroldo M, Paillard S, Marconi R, Fabrice L, Canaguier A, Cruaud C, De Berardinis V, Guichard C, Brunaud V, Le Clainche I, Scalabrin S, Testolin R, Di Gaspero G, Morgante M, Adam-Blondon AF. 2008.**

A physical map of the heterozygous grapevine 'Cabernet Sauvignon' allows mapping candidate genes for disease resistance. *BMC Plant Biology* 8: 66.

**Neff MM, Turk E, Kalishman M. 2002.** Web-based primer design for single nucleotide polymorphism analysis. *Trends in Genetics* 18: 613-615.

**Newcombe G, Bradshaw HD Jr, Chastagner GA, Stettler RF. 1996.** A major gene for Resistance to *Melampsora medusae* f. sp. *deltoidae* in a hybrid poplar pedigree. *Phytopathology* 86: 87-94.

**Newcombe G. 1998.** Association of Mmd1, a Major Gene for Resistance to *Melampsora medusae* f. sp. *deltoidae*, with Quantitative Traits in Poplar Rust. *Phytopathology* 88: 114-121.

**Padmanabhan M, Cournoyer P, Dinesh-Kumar SP. 2009.** The leucine-rich repeat domain in plant innate immunity: a wealth of possibilities. *Cellular Microbiology* 11: 191-198.

**Paillassa E. 2006.** Quelles pertes de croissance pour 7 cultivars de peuplier face aux attaques de la rouille E4 du mélèze. *Forêt-Entreprise* 168: 60-63.

**Pakull B, Groppe K, Meyer M, Markussen T, Fladung M. 2009.** Genetic linkage mapping in aspen (*Populus tremula* L. and *Populus tremuloides* Michx.). *Tree Genetics & Genomes* 5: 505-515.

**Paolucci I, Gaudet M, Jorge V, Beritognolo I, Terzoli S, Kuzminsky E, Muleo R, Mugnozza GS, Sabatti M. 2010.** Genetic linkage maps of *Populus alba* L. and comparative mapping analysis of sex determination across *Populus* species. *Tree Genetics & Genomes* 6: 863-875.

**Pichot C, Teissier du Cros E. 1993a.** Susceptibility of *Populus deltoides* BARTR. To *Melampsora larici-populina* and *M. alii-populina*. I. Qualitative analysis of a 6 X 6 factorial mating design. *Silvae Genetica* 42: 179-188.

**Pichot C, Teissier du Cros E. 1993b.** Susceptibility of *Populus deltoides* BARTR. To *Melampsora larici-populina* and *M. alii-populina*. II. Quantitative analysis of a 6 X 6 factorial mating design. *Silvae Genetica* 42: 188-199.

**Pinon J, Frey P. 1997.** Structure of *Melampsora larici-populina* populations on wild and cultivated poplar. *European Journal of Plant Pathology* 103:159-173.

**Pinon J, Valadon A. 1997.** Comportement des cultivars de peupliers commercialisables dans l'Union Européenne vis-à-vis de quelques parasites majeurs. *Annales des Sciences Forestières* 54: 19-38.

**Pinon J, Frey P. 2005.** Interactions between poplar clones and *Melampsora* populations and their implications for breeding for durable resistance. Edité par McCracken AR, Pei MH : Rust disease of willow and poplar. CAB International, Wallingford, UK. Pages 139-154.

**Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ. 2009.** Shades of gray: the world of quantitative disease resistance. *Trends in Plant Science* 14: 21-29.

**Rafiqi M, Bernoux M, Ellis JG, Dodds PN. 2009.** In the trenches of plant pathogen recognition: Role of NB-LRR proteins. *Seminars in Cell & Developmental Biology* 20: 1017-1024.

- Ren C, Xu Z, Sun S, Lee MK, Wu C, Scheuring C, Zhang HB. 2005.** Genomic DNA libraries and physical mapping. Edited par Meksem K & Kahl G : The handbook of plant genome mapping. WILEY-VCH. Pages 175-213.
- Scalabrin S, Troglio M, Moroldo M, Pindo M, Felice N, Coppola G, Prete G, Malacarne G, Marconi R, Faes G, Jurman I, Grando S, Jesse T, Segala C, Valle G, Policriti A, Fontana P, Morgante M, Velasco R. 2010.** Physical mapping in highly heterozygous genomes: a physical contig map of the Pinot Noir grapevine cultivar. *BMC Genomics* 11: 204.
- Schuler GD. 1997.** Sequence mapping by electronic PCR. *Genome Research* 7: 541-550.
- Simons G, Groenendijk J, Wijbrandi J, Reijans M, Groenen J, Diergaarde P, Van der Lee T, Bleeker M, Onstenk J, de Both M, Haring M, Mes J, Cornelissen B, Zabeau M, Vos P. 1998.** Dissection of the *fusarium* I2 gene cluster in tomato reveals six homologs and one active gene copy. *Plant Cell* 10: 1055-1068.
- Song WY, Wang GL, Chen LL, Kim HS, Pi LY, Holsten T, Gardner J, Wang B, Zhai WX, Zhu LH, Fauquet C, Ronald P. 1995.** A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270: 1804-1806.
- Stirling B, Newcombe G, Vrebalov J, Bosdet I, Bradshaw HD Jr. 2001.** Suppressed recombination around the MXC3 locus, a major gene for resistance to poplar leaf rust. *Theoretical and Applied Genetics* 103: 1129-1137.
- Tabor GM, Kubisiak TL, Klopfenstein NB, Hall RB, McNabb HS. 2000.** Bulked Segregant Analysis Identifies Molecular Markers Linked to *Melampsora medusae* Resistance in *Populus deltoides*. *Phytopathology* 90: 1039-1042.
- Tameling WI, Elzinga SD, Darmin PS, Vossen JH, Takken FL, Haring MA, Cornelissen BJ. 2002.** The tomato R gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. *Plant Cell* 14: 2929-2939.
- Tameling WI, Vossen JH, Albrecht M, Lengauer T, Berden JA, Haring MA, Cornelissen BJ, Takken FL. 2006.** Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause autoactivation. *Plant Physiology* 140: 1233-1245.
- Tan X, Meyers BC, Kozik A, West MAL, Morgante M, St Clair DA, Bent AF, Michelmore RW. 2007.** Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in *Arabidopsis*. *BMC Plant Biology* 7: 56.
- Thomas CM, Jones DA, Parniske M, Harrison K, Balint-Kurti PJ, Hatzixanthis K, Jones JD. 1997.** Characterization of the tomato Cf-4 gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in Cf-4 and Cf-9. *Plant Cell* 9: 2209-2224.
- Tuskan GA, Gunter LE, Yang ZK, Yin TM, Sewell MM, DiFazio SP. 2004.** Characterization of microsatellites revealed by genomic sequencing of *Populus trichocarpa*. *Canadian Journal of Forest Research* 34: 85-93.

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313: 1596-1604.

Van der Biezen EA, Jones JD. 1998. The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Current Biology* 8: R226-7.

Villar M, Lefèvre F, Bradshaw HD Jr, Teissier du Cros E. 1996. Molecular genetics of rust resistance in poplars (*Melampsora larici-populina* Kleb/*Populus* sp.) by bulked segregant analysis in a 2 x 2 factorial mating design. *Genetics* 143: 531-536.

Wicker T, Yahiaoui N, Keller B. 2007. Illegitimate recombination is a major evolutionary mechanism for initiating size variation in plant resistance genes. *Plant Journal* 51: 631-641.

Wisser RJ, Sun Q, Hulbert SH, Kresovich S, Nelson RJ. 2005. Identification and characterization of regions of the rice genome associated with broad-spectrum, quantitative disease resistance. *Genetics* 169: 2277-2293.

Woolbright SA, DiFazio SP, Yin T, Martinsen GD, Zhang X, Allan GJ, Whitham TG, Keim P. 2007. A dense linkage map of hybrid cottonwood (*Populus fremontii* × *P. angustifolia*) contributes to long-term ecological research and comparison mapping in a model forest tree. *Heredity* 100: 59-70.

Yang S, Zhang X, Yue JX, Tian D, Chen JQ. 2008. Recent duplications dominate NBS-encoding gene expansion in two woody species. *Molecular Genetics and Genomics* 280: 187-198.

Yin T, Zhang X, Huang M, Wang M, Zhuge Q, Tu S, Zhu LH, Wu R. 2002. Molecular linkage maps of the *Populus* genome. *Genome* 45: 541-555.

Yin TM, DiFazio SP, Gunter LE, Riemenschneider D, Tuskan GA. 2004a. Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. *Theoretical and Applied Genetics* 109: 451-463.

Yin TM, DiFazio SP, Gunter LE, Jawdy SS, Boerjan W, Tuskan GA. 2004b. Genetic and physical mapping of *Melampsora* rust resistance genes in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence. *New Phytologist* 164: 95-105.

**Yin T, Difazio SP, Gunter LE, Zhang X, Sewell MM, Woolbright SA, Allan GJ, Kelleher CT, Douglas CJ, Wang M, Tuskan GA. 2008.** Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. *Genome Research* 18: 422-430.

**Zhang J, Steenackers M, Storme V, Neyrinck S, Van Montagu M, Gerats T, Boerjan W. 2001.** Fine Mapping and Identification of Nucleotide Binding Site/Leucine-Rich Repeat Sequences at the *MER* Locus in *Populus deltoides* 'S9-2'. *Phytopathology* 91: 1069-1073.

**Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D. 2004.** Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Molecular Genetics and Genomics* 271: 402-415.

# ANNEXE

**Annexe 1** : Composition des solutions (les concentrations indiquées sont les concentrations initiales)

**Premix PCR** (pour 50 ml) :  $MgCl_2$  (25 mM, Applied Biosystems) 5760  $\mu$ l, 10X PCR buffer 9000  $\mu$ l (Applied Biosystems), dNTP (10 mM, Applied Biosystems) 3600  $\mu$ l, Tampon de charge 18 000  $\mu$ l,  $H_2O$  13640  $\mu$ l.

**Tampon de charge** (pour 50 mL) : 30g sucrose 76 à 100 mg de rouge de crésol ("indicator grade" Aldrich chemical Company) qsp 50 ml d' $H_2O$ . Mélanger plusieurs heures et filtrer avec une membrane à 45  $\mu$ m.

**FM glycérolé** : LB bouillon 25 g/l,  $KH_2PO_4$  13 mM,  $K_2HPO_4$  36 mM,  $C_6H_5Na_3O_7 \cdot 2H_2O$  17 mM,  $(NH_4)_2SO_4$  6,7 mM, Glycérol 6%.

**2YT** (pour 1l): Bacto-Tryptone 16g, Yeast Extract 10g, NaCl 5g (Fw = 58,44 g/mol), qsp 1 l d' $H_2O$ .

**GET** (pour 10 ml) : 250  $\mu$ l de glucose 2 M, 250  $\mu$ l de Tris HCl 1 M pH = 8, 200  $\mu$ l d'EDTA 0,5 M pH = 8, 9,3 ml d' $H_2O$ .

**NaOH/SDS** (pour 20 ml) : 400  $\mu$ l de NaOH 10 M, 1 ml de SDS 20 %, 18,6 ml d' $H_2O$ .

**AcK 5/3 M** (pour 20 ml): 12 ml d'AcK 5 M, 5,7 ml d'acide acétique glacial, 2,3 ml d' $H_2O$ , Le pH a été ajusté entre 4,8 et 5,2.

**TBE 10X** : Tris 0,89 M, Acide Borique 0,89 M, EDTA 20 mM.

**Annexe 2** : Motifs MEME utilisés pour trouver les gènes NBS-LRR sur les séquences.

Motif MEME développé sur le génome entier du peuplier (Kohler *et al*, 2008) et sur *A. thaliana* (Meyers *et al*, 2003).

- Pour les gènes TNL

NLS (Kohler *et al*, 2008) : MQKEKRKQSKDEENDSSSRKRRKADLKKP

TIR-1 (Kohler *et al*, 2008) : YDVFLSFRGEDTRNNFTDHLTYALCQAGIHTFRDD

TIR-2 (Kohler *et al*, 2008) : ELPRGEEISPHLWKAIQESRISIIVFSKDYASPWCLDELVKI

TIR-3 (Kohler *et al*, 2008) : CKKxTGQIVLPVFYDVDPDVRKQTGSFA

TIR-4 (Kohler *et al*, 2008) : VQRWRDALTEAANLSGWD

TN linker TNL (Kohler *et al*, 2008) :

ANGHEAKFIQEIVEDVLYKLNxxxPHVPKHLVGMDSRVEYINSWLQMGTHTDVC

NBS TNL P-loop (Kohler *et al*, 2008) : VGIYGMGGIGKTTIAKVVYNQIRxQ

NBS TNL RNBS-A (Kohler *et al*, 2008) : VRLQEQLLSDILKER

NBS TNL Kinase 2 (Kohler *et al*, 2008) : IKDRLCRKKVLIVLDDV

NBS TNL RNBS-B (Kohler *et al*, 2008) : WFGPGSRIITTRDK

NBS TNL RNBS-C (Kohler *et al*, 2008) : YEVEELNDDDALQLFSWHAFKNDQP

NBS TNL GLPL (Kohler *et al*, 2008) : VVDYANGLPLALEVLGSSLYGR

NBS TNL RNBS-D (Kohler *et al*, 2008) : EEKKIFLDIACFFIGMKKDYV

NBS TNL MHDV (Kohler *et al*, 2008) : IVSDDKLWMHDLLRDMGREIV

NL linker TNL (Kohler *et al*, 2008) : EPGKRSRLWTHEDVVLMDNGTEKVEGIFLD

LRR TNL Motif 1 (LDL) (Meyers *et al*, 2003) : MDLSYSRNLKELPDLNATNLERLDLSYCSSLVELPSSI

LRR TNL Motif 4 (end) (Meyers *et al*, 2003) : LHWLDLKGCRKLVSLPQLPDSLQYLDAHGCELETVACP

Cterm TNL (Kohler *et al*, 2008) : EIPGWFNHQGWGSSISFQ

- Pour les gènes BNL

BED (Kohler *et al*, 2008) : SNDPFWNYVEKMDDGSMKCKFCGHLFANGTSSISRIKWHLSGHGVA

BN linker (Kohler *et al*, 2008) : LPTSSTKLVGQAFEENKKVIWSLLMDDEV

NBS BNL P-loop (Kohler *et al*, 2008) : STIGIYGMGGVGGKTTLLQHIHNELLQRPD

NBS BNL RNBS-A (Kohler *et al*, 2008) : HVYWVTVSQDFINRLQNLIKRLGLDLS

NBS BNL Kinase 2 (Kohler *et al*, 2008) : RKKQKWILILDDLWNN

NBS BNL RNBS-B (Kohler *et al*, 2008) : VPLKGCKLIMTTRSERVC

NBS BNL RNBS-C (Kohler *et al*, 2008) : PLSEGEAWTLFMEKLGHDIAL

NBS BNL GLPL (Kohler *et al*, 2008) : RIAKDVARECAGLPLGITMAGSLRGV

NBS BNL RNBS-D (Kohler *et al*, 2008) : QQCLLYCALFPEDHRIEREELIGYLIDEG

NBS BNL MHDV (Kohler *et al*, 2008) : KMHDLIRDMAIQILQENSQGM

Cterm BNL (Kohler *et al*, 2008) : EWEHPNAKDVLRFVFEF

- pour les gènes CNL

NBS CNL P-loop (Kohler *et al*, 2008) : IPIVGMGGVGGKTTLAQLVYND

NBS CNL RNBS-A (Kohler *et al*, 2008) : KIWVCVSQDFDV<sub>x</sub>RIQKAILE  
NBS CNL Kinase 2 (Kohler *et al*, 2008) : LKGKKFLLVLDDVWN  
NBS CNL RNBS-B (Kohler *et al*, 2008) : DGARGSKILVTTRNE<sub>x</sub>VA  
NBS CNL RNBS-C (Kohler *et al*, 2008) : VH<sub>x</sub>LG<sub>x</sub>LSEEDCWSLFKK<sub>x</sub>AF  
NBS CNL GLPL (Kohler *et al*, 2008) : EIAKKCKGLPLAAKTLGGLLR  
NBS CNL RNBS-D (Kohler *et al*, 2008) : RCFAYCAIFPKDYEI<sub>x</sub>KEELVRLWMAEGF  
NBS CNL MHDV (Kohler *et al*, 2008) : FKMHDLVHDLAQSIA  
LRR CNL Motif 1 (LDL) (Meyers *et al*, 2003) : IGNLVHLRYLDLSYTGITHLPYGLGNLKKLIYLN  
LRR CNL Motif 8 (end) (Meyers *et al*, 2003) : LHTITIWNCPKLKKLPDGICF  
Cterm CNL (Kohler *et al*, 2008) : GEDWPKIAHIPNIYI

### Annexe 3 : Les couples d'amorces utilisés en cartographies génétiques locales.

Couple d'amorce	Séquence utilisée	type de locus	Température d'hybridation (°C)	Taille 101-74 (h <sub>R</sub> ,h <sub>T</sub> )	Taille 73028-62 (h <sub>R</sub> ,h <sub>T</sub> )	CAPS ou dCAPS endonucléase de restriction	Taille Nisqually-1	Amorce F	Amorce R
SCAR_E1M4	SCAR	BSA-AFLP	62	189-	-1220		NT	GAATTC AACCACTAATATCTTGCT	TCAAGCTGAGGACTGTTTTGG
SCAR_E21M18	SCAR	BSA-AFLP	62	-1603	-		NT	GGTTGAATACAGCCAAATTTTACG	TGGAGATGGCTGTGGTAGTG
SCAR_E6M5	SCAR	BSA-AFLP	62	192-	-		NT	ATAACCAATGGCCAAACCCAC	TTTAAATGTGCAAACTGTGTCAAC
SCAR_M03/04	SCAR	BSA-RAPD	58	-	-1468		NT	TGAGGAAGCAGATGATATGGCAG	TTTGCATGGTACTGATTTCTCC
GCPC104	Biblio	SSR	56	110 125,160,180	125,160,180		NT	AAGGCTTATAGAAATGTGGCA	TCCTTGATGTTGTTCACAG
GCPC107	Biblio	SSR	52	185 174	-171		NT	TTCAATTATGAGTACGGCT	GCTTTATATCCCGATGTTG
GCPC2319	Biblio	SSR	56	139 130	130		139	TTCAATCATCAGATTTCTGCA	CCTATTGGGTTTAGGGTTT
GCPC2849	Biblio	SSR	56	-1100	-1110		99	ATGTCACACCAAAAAGAGG	AATACCTTTACCCGTGGATT
GCPC79	Biblio	SSR	54	-1250	-1100		230	TTACTTCGACAAGCTTTCT	TAAACACAAAAACAGGGGA
ORPM263	Biblio	SSR	56	283 250	270-		283, 250	AGCACATCTTCGAGCATGA	TGTAGCAATTTGCCAAATCA
PMGC667	Biblio	SSR	54	181,198,212	176,178		207	CATTCGTCAGTAGTAAAGC	GGTAAAGCTACCTCTGCATC
INRA1211	NisV1	SSR	58	231 219	202 236		221	AGAACATATAAATTTCTCAC	GGATATTTCAACATGTGTCC
INRA1593	NisV1	SSR	60	-	-1155		149	CTACTTGTGATGTAATCTCAGGA	ATTCCACGATATATGCGACAG
INRA1944	NisV1	SSR	60	-1130	750-		138	GCCTTCGCTATAAGTCCAAAC	TTTATCCAACTCATCTCCAAA
INRA211	NisV1	SSR	56	-1153	140-		153	TTCTTTTATGCTTTTGACAATCTT	TTGTGACCAAGTAGGACA
INRA2244	NisV1	SSR	58	180-	-		199	CTCAATGCTCAACTTAATTTATCACA	TTCCAAAGCTCTCTCAATAAC
INRA2272	NisV1	SSR	56	150 130	85 85		125	CCCGGTGACTTCTCACATA	TTCTTTATGTTATTTATAAATCAGTCG
INRA2607	NisV1	SSR	58	125 110	-		130	TTGCTTCTTCTTGATGTGTTG	TGGACCAATGACAAGTGTGTA
INRA2770	NisV1	SSR	58	209 266	-		209, 266	GAGAGGAGAGAGTAGTACG	TCATTTGCTACATTAACAATTCAC
INRA2879	NisV1	SSR	58	-1100	500-		108	AGAAAGCTACCCCTGTGACACT	TTTTATTGGGCTTTTGTGTG
INRA465	NisV1	SSR	58	137-	-		137	CTCTTTTGTGACAGCATTGT	TTAGATATAACCCGACAGCCAC
INRA524	NisV1	SSR	58 : 60	140 140,175	-1110		103	CGTTACAGTGAACACTTATCCCTT	AAGTTGCTCAAACTCAAGCC
INRA789	NisV1	SSR	58	110-	-		NT	TGATTTGGGAAATGTTCTCTG	GGACCATTAGTTAAAGCCAAA
INRA927	NisV1	SSR	60	209 235	-1203		NT	TAACCAACCTCCACAATGC	GGATTTGCTCTGTATACCG
INRA979	NisV1	SSR	50	212 222	228 196		230	AATACACTAGACAAGATACATCC	TTAATTTGATAACCAAGTGC
INRA_s165-1	NisV1	SSR	58	188-	-		NT	TGATATTAACAAGCTTTTCAACC	AAGGGCTGAGAAAACCCATA
INRA_s404-1	NisV1	SSR	58 : 60	200-	205 175		NT	TACAAGGGGGAGAGGGAGAG	GCCACGAAATTTGTAAGA
INRA_s601-1	NisV1	SSR	58	210-	-		240	CGSTTTCACCGATCAAAAAT	CGSTTATCAACTGCGGATA
INRA_s748-2	NisV1	SSR	57	181 181	-1162		NT	TGCCCTACACACGACGTA	TTGACGGCAGAGACTCTGGA
nc1432-1	NisV1	STS intG	62	-1477	-		477	CAGACGTGAGCGTGTATGG	GCAGCAAGTTTGAGCAGAA
nc2079	NisV1	STS intG	64	469-	-		469	ATGCGTGGGGCGGTGT	GCAGCAAGTTTGAGCAGAA
nc799-1	NisV1	STS intG	55	-	-1493		NT	CAATGTTCTGACCCCTTTTGAAT	GGGATGATCATGAGAGCTACC
PT1610-1	NisV1	STS G	60	-	-11137		1137	ATGACTTTAAAAAGTAGCATTACG	AATTCACACAAATGCTTTGTT
PT1939-4	NisV1	STS G	60	-1987	-		987	TCAAAAGTTTTGACGTTGTAGG	AGTTGGATCCAAATTCOAATG
PT872-2	NisV1	STS G	64	-11180	-		NT	GGTTGGSCACAATGTCACT	GCTAACATCTGGACAACAGGAC
PTs589 <sup>1</sup>	NisV1	STS G	58	1001=639+362 1001	1001 1001	CAPS <i>HindIII</i>	NT	TCACCACCTTCTGTGTATTT	ACCTCAGCTTCTATTGTTCTG
RGA1159	NisV1	STS RGA	64	-1139	-		NT	GAACTGATGATCATCTATTGGCTAT	GGGCTGTTCTGACCCG
RGA53-2	NisV1	STS RGA	60	1055-	-		1055	TCATTTAAGTTCATAACGATAACTCT	GCTTTCTCCACTCTTCCACC
RGA569	NisV1	STS RGA	58	77-	77,120 77		75	GAAAGTGAATGGCAATGGAA	ATGCATCTGATGCTGTGAATA
RGA805	NisV1	STS RGA	60	886 1093	1093-		NT	GTTAGGATCGTCAATTTTATGTTCC	GCAGCAATAAGCTTTCGAA
RGA836 <sup>2</sup>	NisV1	STS RGA	60	1069=505+564 1069	1069 1069	CAPS <i>HaeIII</i>	NT	TTTACTTTAGGSCCTGTTACTG	AGTTTCCATGAAACAAAGTCCA
RGAs135-1	NisV1	STS RGA	62	708-	-		737	CAAAAGATATCGACTGTGAGTTG	CATCATCTCGAAAAGTGTGGATT
RGAs177-1	NisV1	STS RGA	60	-	-1992		992	GCTGTTGCAATTAACCTGACC	CATCATCTCGAAAAGTGTGGATT
RGAs237-1	NisV1	STS RGA	58	-	-1763		NT	CGTGGTAAATACCCAGAGTTCA	CATCATCTCGAAAAGTGTGGATT
RGAs237-4	NisV1	STS RGA	62	-	-1614		614	CAAAATGAATGAAAGACGAGTGTGTC	CATCATCTCGAAAAGTGTGGATT
RGAs2701-1	NisV1	STS RGA	56	-	470-		470	ACAAATTTCCATTGTGCTTTT	CATCATCTCGAAAAGTGTGGATT
RGAs283-2	NisV1	STS RGA	60	-	-1867		NT	CTTCAGAGCTGTCAAGATTAAGTT	AGAAGCATATCTTTGAGAAGACA
RGAs297	NisV1	STS RGA	60	1014-	-		1014	TTTTGACGAGCTCGAGTTTA	GGAAACAAATGATGAGCTCAACC
RGAs336	NisV1	STS RGA	58	1554-	-		1554	CTGACTCCGCCCTGTATAC	TTCTCTTTCAGCTCCACCTACT
RGAs352	NisV1	STS RGA	60	356-	-		356	GCSTTGTGAGTACTATTCTTTC	AACATTTGTGCAAGGAAACGAC
INRA_920-3	NisV2	SSR	55	250-	-1270		NT	AAAGAAAACCAAGACTTAAA	TTTCACTCTCTTGCATAA
PT2387	NisV2	STS G	55	732-	-		NT	CAAAACACACGCAATTAATAGTCA	GCACAGGAGCTGTGTAAGT
RGAm4-1	MER	STS RGA	52	1118-	-		1118	GACTCTCAATCTCTATTCCACTGA	CATCATCTCGAAAAGTGTGGATT

### Annexe 3 (suite) :

Couple d'amorce	Séquence utilisée	Type de locus	Température d'hybridation (°C)	Taille 101-74 (h <sub>R</sub> , h <sub>R</sub> , r <sub>101</sub> )	Taille 73028-62 (h <sub>R</sub> , h <sub>R</sub> , r <sub>1</sub> )	Taille Nisqually-1	Amorce F	Amorce R
04J18-R-1	BES	STS intG	58 : 60"	301-	-1398	NT	AACCTTAATCTTCAGCTTTCTGC	ATCTGTGAAGTATTGCTGGGAAG
08K24-F	BES	STS intG	55	981\450,615	615,800\800	NT	ACCCATTCTTAAAAGGTGCG	TGTAATTGCAGATGACTCC
113G23-F	BES	STS intG	58	-1540	-	540	AGCTTGGACAATATTGCTTTT	AGATCACCTTGGAGATGCATT
125A14-F	BES	STS intG	60	419-	-	-	TTCAAACAATATTCTCCGTTCC	CCAGGGTCTTTAGCTTTACC
14F23-F	BES	STS intG	55	-1617	400\400	NT	AAGCCAATTCATGGCTTCA	ATGTGGCTACATTGAAGCAG
14N08-F	BES	STS intG	60 : 60"	340\308	-1308	NT	TATCTTTGCCACAACCGTAGAG	CTAAGGTGGTGCAGTAGGATAG
17N11-R	BES	STS intG	58	468-	370	-	CAGTCTTAAAAATTCGTCAAGC	GATGGAAGGATTACAGTTGGTC
22H17-F-2	BES	SSR	56	425\288	-	288	CACAAAACAAGAGATTCAATCG	CCTGGACTCAACTATCTAAAATGC
33C08-R	BES	STS intG	60	515,773-	-	-	ACTCTAACGAGATCCATCAGCA	TGTGTTTTGGTACATTGGTCA
42B22-R	BES	STS intG	58	269-	-	269	CTAGTTTCCAAAAGGCAACCT	TTCCTCGAAGCTGAACAAAA
42M10-F	BES	STS intG	60	275\347	-	284	CTTGGACTCGAGTTAATCAAGG	TTGGTTACCCCTATATAATTTGTT
43D24-F	BES	STS intG	64	641-	-	-	AAACCACCCTGCCTAAAGTT	TACTCTTCTTGCCTGTGCT
47N12-F	BES	STS intG	60	384-	-	384	CTTGTACGCCAGTCTAGCTTCT	GAGCTCTATCAGATGCAATCA
57L24-F	BES	STS intG	60	431-	-	-	TGGATCCAAGAAATGTTAGTTT	TGCTTTTTCGGTAGAATACATCA
57L24-R	BES	STS intG	58	345-	-	345	TCTCTCTAAAAGTATTGGATCTAAGAGG	GATTTATTGGTGAATGACTGCTC
97F12-F	BES	STS intG	60	572-	-	-	ATAAGGCTTAGAATCAACAAGC	ATCTCGATCCTCAGTCAATTC
97G09-R-1	BES	STS intG	60	625-	-	-	CGAGGCTGTGATTAGCAGATTA	CAATGTTTGTGGAGCTCACT
04J18-F-AS	pPCR	STS intG	58	161\218	-	NT	TTCAAATCCCTAATCCAGTTGA	CATCTTGTCAAGCTCTATATGCTG
97G9-AS-11	pPCR	STS intG	58	-	343-	NT	TTAGAAAACAAGAAAGATATAGAATT	CGTGCAATGCTAGGAAATATAATA
97G9-AS-9	pPCR	STS intG	60	156-	-	NT	AGAAAACAAGAAAGATATATAAAGAATC	TTTGTAGATCTATTCCGCATCC
PTs623-AS	pPCR	STS G	58	-1803	-	NT	GGGGAAGAATATGCTGTCAAGA	CTGCTTACCAATTCATGAAGTCGA
PTs693-AS	pPCR	STS G	60	265\211	-	NT	CCCGTCTCTTTTATCTTCAAC	CGAATCAACATGTTGAGCTGGG TCTTCCACGCTTAACCTCA
PTs776-AS	pPCR	STS G	60	416\142	-	NT	CAAAACAACCTCCCACTCTCTC	GGGGTTGGGAGAATTTTCA
RGAs931-dC <sup>33</sup>	pPCR	STS RGA	53	350=324+26\350	350\350	NT	ATCTATAGTGTCTTCTCTAAAGGGTA	TTACATGGTAAATCTAATGT
RGAs682-dC <sup>84</sup>	pPCR	STS RGA	53	151=128+23\151	151\151	NT	TTTTATATCTCTCTCCCATC	CATAATTCAGCACCATCTACTT
RGAs735-AS	pPCR	STS RGA	59	-1336	-	NT	GTCAGAAACAGTTTTGCTATTA	TGAAAACAGATGAAGAGGACA

Colonne couple d'amorce : #1 CAPS *Hind*III, #2 CAPS *Hae*III, #3 dCAPS *Rsa*I #4 dCAPS *Mse*I.

Colonne sequence utilisée : Biblio : couple d'amorce issue de la bibliographie (Tuskan *et al*, 2004), NisV1 : développer sur le génome de Nisqually-1 V1, NisV2 : développer sur le génome de Nisqually-1 V2, MER : développer sur le BAC proche du locus *MER* (Lescot *et al*, 2004), BES : développer sur un BES du 101-74, pPCR : développer sur un produit d'amplification PCR, BAC : développer sur un BAC du 101-74.

Colonne type de locus : SRR : marqueur microsatellite, STS intG : marqueur STS intergénique, STS G : marqueur STS génique, STS RGA : marqueur STS sur un gène de résistance.

Colonne taille d'amplification : NT : non testé, ND : non déterminé, - : pas d'amplification.

## Annexe 4 : Les couples d'amorces utilisés en cartographie physique.

Couple d'amorce	Séquence utilisée	type de locus	Température d'hybridation (°C)	Taille 101-74 (h <sub>1</sub> R <sub>1</sub> h <sub>2</sub> h <sub>3</sub> )	Taille 73028-62 (h <sub>1</sub> R <sub>1</sub> h <sub>2</sub> )	Taille Nisqualiy-1	Amorce F	Amorce R	
SCAR_E25M14	SCAR	BSA-AFLP	58	299	ND	299	GATCCAAAGGTTTCGATGC	TCTCGCTGTACCTACTAAAGGTC	
SCAR_E4M2	SCAR	BSA-AFLP	60	217		217	GTGATCTATGCGTTCATTCC	GCCACATGCTGTAGTGAAA	
SCAR_E5M5	SCAR	BSA-AFLP	56	188		188	TGTCCAACCTAATCAGGACAAC	GACTCTAACCTAGTTTTGTTCG	
INRA_136	NisV1	SSR	58	140	140	145	140	AAATTAAGGCATATTGAGAGGATG	TCCTTCGCTCCCTCGTAAAG
INRA_146	NisV1	SSR	52	-170		-170	170	ATATTTTTAGACCGCGCTCTC	ATCGATAAATATCCTATTGTGTT
INRA_s523-1	NisV1	SSR	56	190,425	425	-145	NT	TCAAACCTTTTCAATGAGGTCC	ATTGGGATGGATTGATTGAA
nc1327	NisV1	STS intG	60	529	529	-	NT	TGTAAGCACGGGAGTTTTT	TGGGTCTCTGGGAGATTC
PT872-3	NisV1	STS G	62	1085	1184	1085	1085	GGATGTTGCCACTTCCTC	GCTAACATCTTGGACAACGAGGAC
PTs565	NisV1	STS G	60	1007	ND	1007	1007	CACCCTGATGAGAAATACGTA	ACCACCAGAATAGTCCATGACAC
PTs614	NisV1	STS G	58	94	11ND	941	NT	TTCACCCGGCTAATGAAGAAG	TTTCGAAAGTGGTTCATCACC
PTs623	NisV1	STS G	58	868	868	868	868	GGGGAAGAATATGCTGCTCAAGA	ATAAGCTTCGCCATTGTATGC
PTs644	NisV1	STS G	56	1049	ND	1049	NT	TTGATAATCCAATCGAGCTTT	ATGCAAGSGTATTGACCTGTT
PTs693	NisV1	STS G	58	1070	1070	1070	1070	CCCGCTGCTTTTATCTTCAAC	CTTTGGCTACCAATCAAAAGC
PTs710	NisV1	STS G	60	900	900	900	919	GAGATTCCATTGACAGGGTTT	AAATCAGATGCAAGGGAGAAA
PTs759	NisV1	STS G	58	1024	1024	1024	NT	CTCGCCATTTAGAATTGACTT	GCTAGTGGACCAAAATTAACAA
PTs776	NisV1	STS G	58	948	948	938	NT	CAACCAACCTCCACTCTCTC	AACGTACCGAATTCAGCGTAAA
PTs816	NisV1	STS G	58	1030	1030	1030	NT	AATTTCAACAGCCCATACATC	CCCATTTGCATTACGCAATA
PTs847	NisV1	STS G	60	1046	1046	-	1046	ACTGCTGAGTGAAGCAACA	TGAATCCAGATTGATTTGAT
PTs864	NisV1	STS G	58	-1150	1042	1042	1042	ACTGTAGGTTGAAGCACAACA	ATTGTAGTATGATGGCAATG
RGA1025	NisV1	STS RGA	60	-987		999	NT	GCATGACGTAATTAAGTATG	TGCCATTAATTCACCAAAAG
RGA1027	NisV1	STS RGA	62	1214	1214	1214	NT	GAATAAATGAGCTAAACTCTAAAATAAC	GTCTTTTACCGTGGTGTATG
RGA1078-1	NisV1	STS RGA	52° 90"	1297	1297	-	1297	TTCAGTTATATATGATGATTAGAGTGT	TTGTTTATGTTGATATGTCAG
RGA1099-2	NisV1	STS RGA	52° 90"	-1370		-	1370	AAATGGTTACACAATAGTATAGA	ATTCATAAATGAGCATGGATAC
RGA1217	NisV1	STS RGA	62	1124	1124	1124	NT	TCCTGACCTCAAAACAATCTT	TGGAGATTGCTCCCTGAC
RGA806	NisV1	STS RGA	60	1128	1128	1128	NT	GGTCAAGAATTTAAAGTCATTAAGG	GGACACTGTTGGATGTCC
RGA889-1	NisV1	STS RGA	60	-641		641	NT	TGGAGTTAGGATCTTTCTATAGTCC	AACAGTATGCTAGTCCCAAAATC
RGA904	NisV1	STS RGA	60	-500		-	NT	TGATTTGGGATCTCTCAAGTAAT	ACGATGCTACCCATCTTATCC
RGA929	NisV1	STS RGA	60	1094		NT	NT	TTTTTAGCATGCTTAATGCTCTCTC	GGGTGTTAGTGACTCTTTGTG
STKs1	NisV1	STS RGA	60	95	1951	-	NT	TGATTAACAATCTCAATCCCAT	GACGTGCAATACCCATTTTTG
RGAs2139	NisV1	STS RGA	58	110	-	-	110	GAAATGAAATGGCAATGGAA	AACGCTGGTTCAATGATTTAC
RGAs310-1	NisV1	STS RGA	60	1081	ND	1081	NT	TTGATTAACAGAGTAAACCGATGGA	CATCATCTCGAAAGTGTGATT
RGAs4600-2	NisV1	STS RGA	60	679	679	-	NT	CCAGTTTCAATGACAGTCTTTG	AGAAGCATCTCTTTGAGAAAGACA
RGAs654-1	NisV1	STS RGA	58	828	ND	-	NT	GGATGGTAAAGCACTTGAATTAAT	AAGGCATGCCCACTAAAAGC
RGAs682	NisV1	STS RGA	60	588	588	588	588	ACCAACCGCAACTTGAATCT	AACCAACGAAAGTGAAGAA
RGAs726	NisV1	STS RGA	60	1016	1016	1016	NT	TATGTAGGAATCTGTGGACTG	AAGCGAGTCTAATTTGATGGC
RGAs728	NisV1	STS RGA	60	1055	1055	1055	NT	CCATGCTGAAGTCTGTGAAGC	GAGAGATTGAGGAAGCAATGG
RGAs735	NisV1	STS RGA	60	1154	1154	1154	NT	CATAGCTCATGACCACTTTG	TGGAACAGATGAAAGAGGACA
RGAs745	NisV1	STS RGA	58	766	766	766	NT	GCCTATTGGAATGGGGATAGAT	GAGTGGTCTTCCCGAAT
RGAs931	NisV1	STS RGA	60	856	856	856	857	TCCCACAGAAATACATGATTTG	GGTACTATGAAATCAATAGTATGGAC
STKs980	NisV1	STS RGA	60	1189	1189	1189	1189	TGATGACAAAGAGATGGAGAA	TGGACGACAAAGCAACTTACC
nc1327-AS	pPCR	STS intG	60	452	452	-	452	TGTAAGCACGGGAGTTTTT	GCTGAAAAGGATTTTACTTAGGAAG
04J18-F-3	BES	STS intG	58	333	333	345	-	CAATAGCTCATGGCCAAGTAAG	GACCACATTTCAAATATACCCTAAA
07B08-F	BES	STS intG	58	531	ND	531	531	ACTCATGGTCTCAACATTGCT	GTGGGTGATGTAGCTGGTAAA
07B08-R	BES	STS intG	58	342	ND	349	349	TAGTAAAGACTCTCACGGACCA	CGCTAAAAGACTACTGCTGTGA
103A04-F	BES	STS intG	58	378	ND	378	378	ATCCCACTGGAACAGATTCC	ATCTTGATCTGCTGGCTAGA
103A04-R	BES	STS intG	58	363	ND	363	363	CAATCATGCTGCAAGTATGAA	TTGTCAGGATGATTTGGTG
103M15-R	BES	STS intG	60	238	238	-	238	CTTGAATGAAATGAGACAAATGCT	CAAGACGATCCACTATATCAGGA
111105-R	BES	STS intG	54	616	616	616	616	ATCGATTTCCCATTTGCTCTT	TTCTTTGCTAAAAGGTGCAAAA
125A14-R	BES	STS intG	60	421	ND	421	422	CTTCGTTGCTGGAAGATTATG	CAAAGCAGAAAGTCTCATC
127C06-R	BES	STS intG	58	234	ND	234	-	ATATCCCCTCATGACAGCA	TAAATTCATCAACATTTGGG
16N23-F	BES	STS intG	58	139	-	-	151	TCTCCACTTAATAAATCACGTTAAA	GGGAAATGAGATGCTTACTAGT
21H24-F	BES	STS intG	58	233	-	-	234	TTTTGTGCGAAGAAATGTTT	AAAACCTTTTATGCTCCCTCT
22B20-F	BES	STS intG	58	600	-	600	-	TGSAAAACAAGTCCGATTAAGT	GAAAGAGTTCAGGAGTTTCA
22B20-R	BES	STS intG	60	308	308	308	308	GAACATCTCAATGAAAATCGCAC	AACATGTTACAAGCAGCAATG
22H17-F-1	BES	SSR	58	175	175,212	212	175 et 212	TGATGCCACCAATAAAGCAAT	TGGCACTAATGATGATGCTGT
40F10-R	BES	STS intG	60	276	-	276	276	AGACAGACAAATGSAATGGAAT	ATGTTTCCCAACAATCTCTTAC
57L24-F	BES	STS intG	60	431	-	-	-	TGGATCCAAGAAATGTTAGTTT	TGCTTTTGGTGAATACATCA
78I01-F	BES	STS intG	62	536	535	537	535	ACTCAAGTCACTGGTTAATCTG	TTCACTTTACATGACAGATGCT
78I01-R-1	BES	STS intG	58	328	328	-	-	ATTGCGAAATTAATCCAGAA	CTCTCTCCATCAACACTTCTCT
93I12-F	BES	STS intG	58	253	253	NT	NT	TTCCTTAACAATACGATACCAGA	ATGAGCGGCTGATATAATACGG
99P13-F	BES	STS intG	58	-164		164	-	CTCAAAAGCTTAAACTATTAGGTGA	ACGAGTTCGAATCACCATT

## Annexe 5 : Les couples d'amorces développés pour la cartographie fine de $R_{US}$ .

Couple d'amorce	Séquence utilisée	Type de locus	Température d'hybridation (°C)	Taille 101-74 ( $R_{US}$ ) ( $R_{US}$ ) ( $R_{US}$ )	Taille 73028-62 ( $R_{US}$ ) ( $R_{US}$ ) ( $R_{US}$ )	Taille Nisqualiy-1	Amorce F	Amorce R
RGAR22-1	BAC	STSRGA	60	785-	-1785,900	-	AACTCTGACAAGAGCCATTGTG	CATCATCTCGAAAAGTGTGGATT
RGAR38-1	BAC	STSRGA	58	466-	-	-	AAAGAATGTAGGGTCCATGCAA	CATCATCTCGAAAAGTGTGGATT
RGAR73-1	BAC	STSRGA	58	326-	-1326	600	CCAGCTCAATATTACCAGCACT	CATCATCTCGAAAAGTGTGGATT
RGAR73-4	BAC	STSRGA	60	1079-	-11079	1192	GGATCACCATGATTGTCGAGTA	CATCATCTCGAAAAGTGTGGATT
RGAR100-1	BAC	STSRGA	60	680-	-1680	-	GAGGCGTGGACTGAAATTTTG	CATCATCTCGAAAAGTGTGGATT
RGAR100-4	BAC	STSRGA	58	892-	-	-	TTTCTGGTCTAATCTTCAA	CATCATCTCGAAAAGTGTGGATT
RGAR176-1	BAC	STSRGA	60	500,998-	-1325,500	500	GGCTGGGTCAATGAAGATAAC	CATCATCTCGAAAAGTGTGGATT
RGAR18-2	BAC	STSRGA	60	1300-	400,500/400,500,1300	-	GAAAGTGAATGGCAATGGAA	CCATAAACTAATGATAAACCGGTAA
RGAR32-1	BAC	STSRGA	60	756-	-	-	ATGTGGGTCCATGTGATC	CATGTCATGCATAGCCATAAT
RGAR70-1	BAC	STSRGA	56:20"	437-	-	-	GCGTTGTTGAGGTACTATTCTTTC	ATGGAATGACGATAATGTCAGC
RGAR70-2	BAC	STSRGA	58	380/380	-	380	GCGTTGTTGAGGTACTATTCTTTC	TATCCAGCGCCCTGCTATT
RGAR96-2	BAC	STSRGA	58	597-	-	-	GCGTTGTTGAGGTACTATTCTTTC	TGTACCACAATAGGACGAACA
RGAR96-4	BAC	STSRGA	64:20"	584-	-	-	GCGTTGTTGAGGTACTATTCTTTC	GGACGAACACTATCACCGTAGA
RGAR138	BAC	STSRGA	58	376IND	376	376	GCGTTGTTGAGGTACTATTCTTTC	TATTGGTATGGCAGTTTGTGG
RGAR172	BAC	STSRGA	54	310-	-	310	GCGTTGTTGAGGTACTATTCTTTC	GTCATGCTAAGTATATGAGTGCACA
INRA_303-1	NisV2	SSR	57	142/388	-	NT	CAAAGAAAAACAACATGG	GCCTAGCCACATTGATT
INRA_R79	BAC	SSR	58	217-	-	-	AAAAACAGTACTAAAACATTCTAACCATTTG	ATGATTACACAATTATTGATGGATAG
INRA_R84	BAC	SSR	56	108-	-	-	TTTTATCAACAATCCTTAAATTAACACAA	AAACTTGGAAAACATCTGGAG
INRA_R91	BAC	SSR	56:60"	160/175	-1360	-	TAAGCTATTGGTTGGAATGGT	TTGTAGATAATCTCATGTAGAAAATCTCTTT
ncR109	BAC	STSRGA	58	463/463	-	-	TTGAATCTTTGACCGACATCAC	TTCTCGAATCATCTCTCCAT
ncR115	BAC	STSRGA	58	386/370	-	-	AGCTTCGAAGGCTGTAATTGAA	TTGTGATCTGGCCTTATGATTG
ncR125	BAC	STSRGA	58	351IND	351	330	CTGTCCCTTTGCTCTCTCCCTA	CAAATGACCATTAGCAATCGAA
ncR87	BAC	STSRGA	58	236/236	236	NT	ATTAATTGGAACCTCGCTCATC	ACGACTCCATTGTTGACAGGTT
ncR89	BAC	STSRGA	58	242/242	242	NT	ATTTGATGACATGTGACAAGC	TCAGTACCGCTTGAGAAATGS
ncR109-AS-1	pPCR	STSRGA	58:20"	-1228	-	NT	GTGAAGAAGAGCCCTGCTACA	TTCTCGAATCATCTCTCCAT
ncR109-AS-2	pPCR	STSRGA	58:20"	288-	-	NT	TTGAATCTTTGACCGACATCAC	AGATCCTTTTATCAAAAAGTTGCG
ncR115-AS	pPCR	STSRGA	58:20"	152-	-	NT	AGCTTCGAAGGCTGTAATTGAA	CCCTCATTCTTTGATTTGACTC

**Annexe 6 : Les couples d'amorces développés pour la cartographie fine de  $R_1$ .**

Couple d'amorce	Séquence utilisée	Type de locus	Température d'hybridation (°C)	Taille 101-74 (h <sub>R1</sub> , h <sub>R2</sub> )	Taille 73028-62 (h <sub>R1</sub> , h <sub>R2</sub> )	Taille Mismajilly-1	Amorce F	Amorce R
INRA_858-1	NisV2	SSR	55	150	150	NT	CAAAAGTCAAGGCCAAAA	GGGCAACCAACATCTCTA
INRA_957-1	NisV2	SSR	55	100	100	NT	TAAAAGCCTCTCTGCTCTT	AAAAAGTCAATGCTGATG
INRA_729-1	NisV2	SSR	55	150\150,170	150	NT	ACGCATCAAAATGATCTGA	ACTGATGAAGGCATGG
INRA_812-4	NisV2	SSR	55	235\215	-	NT	TCACATTGTTGTTGATTGC	TTCTTGGCAGAAAGCTA
INRA_837-1	NisV2	SSR	55	230\230	-\150	NT	TCCTTGTAATGGGACAAAA	ACATAACACAGGCCACCTA
INRA_762-1	NisV2	SSR	55	-\155	-\180	NT	GCCTGGGTAAAACACT	GATCTAGTTTAAAGGGTCTGT
INRA_998-6	NisV2	SSR	55	-	-\250	NT	TGACGTGGCACAATATCA	ATAGATGGGGCTGAATC
INRA_819-1	NisV2	SSR	55	193\161	-\250	NT	GTTTTAATATGTTGGCTCA	AAATATTAACCGCTGGAC
INRA_863-3	NisV2	SSR	55	110-	125\125	NT	CGATGATGGGCTCTATCT	CGAGAGATCCAACGTGAA
INRA_888-1	NisV2	SSR	55	198-	150\150	NT	CAAGTCCCATAATCCCA	TGCATGTTATCCCTTTTCT
INRA_747-1	NisV2	SSR	55	-\151	160\160	NT	TTCTCTGGCTTGATAAACA	ATGCCATCACTTCCATGT
INRA_938-1	NisV2	SSR	55	-	164\164	NT	TGTTTTCTTTGGCTTTC	TGGCCATTGATGAGAAGTA
INRA_847-1	NisV2	SSR	55	170\185	195\165	NT	GGCATTGTTTAAATATCA	CATTATGAACCTGTTTTGG
INRA_943-1	NisV2	SSR	55	-\225	200\200	NT	GGTATGTTTGGAGTGACG	TCAACGTTTTACACTGCTTT
INRA_938-2	NisV2	SSR	55	275-	254\254	NT	TGCAAAATCAAGCTTTCAC	ACAAGTTTTTGGCATGG
RGA776	NisV2	STS RGA	55	165\165	-	165	TGAAGGGTCTGTTATTGGAG	TGATGGAGTACGAGACAGA
RGA775	NisV2	STS RGA	55	167\167	-	167	AGAGAGGTGGCTTTCCTGAT	CATGTTCTCCACATCATCCA
RGA807	NisV2	STS RGA	55	170\170	-	170	ACCTGAGGTATCTGAGGATGS	ATTTTCCAACCTCCTCAAGC
RGA811	NisV2	STS RGA	55	164\164	-\150	164	CCGCTCTTCTCTGTGCTA	GCCTTGCCTGAGCAGTAAT
RGA822	NisV2	STS RGA	55	165\165	165\165	165	ATCATGGCGAAAGTAAAAACA	TGTGATGACCTCCAGTTCT
RGA782	NisV2	STS RGA	55	190\172	172\172	172	CGGAACAATGAAGTACACACA	AATCTCTCGAGGCTCATCT
RGA784	NisV2	STS RGA	55	200\188	188\188	188	AGATACCTCAAGGCATGCAA	TTCCAACCTTCTCGAGCATC
RGA809	NisV2	STS RGA	55	159,210\159,210	210\210	159	GCATTGAGTTGGTTACCAGAA	ACCACACTGACTCCCACCA
RGA821	NisV2	STS RGA	55	228\228	228\228	228	GATTTGTGAAGGCTTTTGA	ACATTCACAGCAATTCGTT
nc826-1	NisV2	STS intG	55	941	941	941	CTGTAATGSCCTTTGAGCCTATG	ACTCTGCTGTAGAAGATCAG
nc825-3	NisV2	STS intG	55	986	986	986	TATTGACGGTCCAGATGTAAGG	TGCAGCTTCATAACCAAGCTAT
nc816-2	NisV2	STS intG	55	184\270	-	184	AATGTGAATGGGATGTTCTGT	TGCAATGTTAACGAGTGGAAAT
nc815-1	NisV2	STS intG	55	468\400	-	468	AGATCTCTAATGCCAAGGATGT	TGGCACATATCAGATTGTTGA

# ARTICLE



# ARTICLE

Article soumis à New Phytologist en novembre 2010.

## **Qualitative and quantitative resistances to leaf rust finely mapped within two NBS-LRR-rich genomic regions of chromosome 19 in poplar.**

Bresson Aloïs<sup>1</sup>, Jorge Véronique<sup>2</sup>, Dowkiw Arnaud<sup>2</sup>, Guerin Vanina<sup>2</sup>, Bourgait Isabelle<sup>2</sup>, Tuskan Gerald A<sup>3</sup>, Schmutz Jeremy<sup>4</sup>, Chalhoub Boulos<sup>1</sup>, Bastien Catherine<sup>2</sup>, Faivre Rampant Patricia<sup>1§</sup>.

<sup>1</sup>INRA, UMR1165, UMR INRA / Univ. Evry : Unité de Recherche en Génomique Végétale, Centre de recherche de Versailles-Grignon, Evry CEDEX, 91057, FRA

<sup>2</sup> INRA, UR0588, Unité de recherche Amélioration, Génétique et Physiologie Forestières, Centre de recherche d'Orléans, Orléans CEDEX 2, 45075, FRA

<sup>3</sup> Oak Ridge National Laboratory, P.O. Box 2008, MS-6422, Bldg. 1062, Rm 215, Oak Ridge, TN 37831-6422, USA

<sup>4</sup> Hudson Alpha Institute of Biotechnology, Huntsville, AL, USA

<sup>§</sup>Corresponding author : Phone : (33) 1 60 87 45 34, FAX : (33) 1 60 87 45 49

e-mail : faivre@evry.inra.fr

## Summary

$R_{US}$  is a major dominant gene controlling uredinia size during leaf rust infection, inherited from *Populus trichocarpa* while  $R_1$  is a gene governing a qualitative resistance, inherited from *P. deltoides*. Here, we report a reiterative process of concomitant fine-scale genetic and physical mapping guided by the *P. trichocarpa* genome sequence. The high-resolution linkage maps were developed using a *P. deltoides* x *P. trichocarpa* of 1415 progeny.  $R_{US}$  and  $R_1$  were mapped in a peritelomeric region of the chromosome 19. Markers closely linked to  $R_{US}$  were used to screen a BAC library constructed from the *P. trichocarpa* parent, heterozygous at the locus  $R_{US}$ . Two local physical maps were developed, one encompassing the  $R_{US}$  allele and the other spanning  $r_{US}$ . The alignment of the two haplo-physical maps showed structural differences between haplotypes. The genetic and physical maps were anchored to the genome sequence, revealing genome sequence mis-assembly. Finally, the  $R_{US}$  locus was localized within a 0.8 cM interval while  $R_1$  was localized upstream of  $R_{US}$  within a 1.1 cM interval. The alignment of the genetic and physical maps with the local reorder of chromosome 19 sequence indicated that  $R_{US}$  and  $R_1$  belonged to a genomic region rich in NBS-LRR and STK genes.

Key words: poplar rust, qualitative resistance, quantitative resistance, genetic mapping, physical mapping, positional cloning, R-gene cluster.

## Introduction

Leaf rust, caused by the fungus *Melampsora larici-populina* (*Mlp*), is the most devastating disease in European poplar plantations credited with premature defoliation, reduced growth

and 50% yield loss (Paillassa, 2006). Infections over successive years can result in the complete loss of economic value. Monoclonal plantations involving a limited number of cultivars have favoured the emergence of new strains and virulences in *Mlp* populations, and so far, 8 virulences have been defined (Pinon & Frey, 2004). Since 1984, no inter-specific *Populus deltoides* x *P. trichocarpa* or *P. deltoides* x *P. nigra* cultivars remained free from rust damages for more than five years after commercial deployment (Gomez *et al.*, 2000).

The genetics of poplar resistance to *Mlp* has been studied in intra- and inter-specific progenies involving *P. deltoides*, *P. trichocarpa* and *P. nigra* species and both qualitative and quantitative resistances are evidenced. Qualitative resistances were inherited only from *P. deltoides* (Pichot & Teissier du Cros, 1993a) and segregations observed in *P. deltoides* x *P. nigra* and *P. deltoides* x *P. trichocarpa* progenies revealed relatively simple genetic determinism (Lefèvre *et al.*, 1994; Lefèvre *et al.*, 1998; Dowkiw & Bastien, 2007). To date, *Mer* and *R<sub>1</sub>*, two loci controlling qualitative resistance to different *Mlp* strains, have been both mapped on the linkage group (LG) XIX of two different *P. deltoides* parental maps (Cervera *et al.*, 2001; Jorge *et al.*, 2005). *Mer* is overcome by the virulence factor 7 and *R<sub>1</sub>* by the virulence factor 1 identified in different *Mlp* strains (Pinon & Frey, 2004, Dowkiw & Bastien, 2004).

Quantitative resistances (QR) measured in field tests under natural infection and in leaf disc bioassay under artificial inoculation were identified in the three poplar species, *P. deltoides*, *P. trichocarpa*, *P. nigra*, and their inter-specific hybrids (Pichot & Teissier du Cros, 1993b; Lefèvre *et al.*, 1998; Legionnet *et al.*, 1999; Dowkiw & Bastien, 2004; Dowkiw & Bastien, 2007). In a reference *P. deltoides* x *P. trichocarpa* mapping progeny, QR measured by sporulation intensity (uredinia size) appears to be mostly controlled by two genetic factors. The first one, referred as *R<sub>US</sub>*, inherited from *P. trichocarpa* was mapped on a small unassigned LG containing 3 markers and spanning 28.3 cM by Jorge *et al.*, (2005). *R<sub>US</sub>* has

been shown to explain up to 80 % of the variation for three components of QR (latent period, size of uredinia, and number of uredinia) after inoculation with seven *Mlp* strains. Recently, Dowkiw *et al.*, (2010) identified several  $R_{US}$ -defeating strains which overcame  $R_{US}$  effect in different poplar genetic backgrounds. The other major locus is inherited from the *P. deltoides* parent and is associated either by pleiotropy or linkage with the qualitative resistance  $R_1$ . Three other defeated qualitative resistances inherited from different *P. deltoides* parents could also be associated to quantitative resistance to *Mlp* (Dowkiw & Bastien, 2007). Villar *et al.*, (1996) and Lefèvre *et al.*, (1998) first suggested the existence of key region involved in the control of *Mlp* rust resistance. Such genomic regions are frequently reported in plants and correspond to clusters of NBS-LRR genes (Meyers *et al.*, 2003). In poplar, clusters of those genes have been found in the vicinity of the *Mer* locus (Zhang *et al.*, 2001). Kohler *et al.*, (2008), Yang *et al.*, (2008) counted approximately 400 NBS-LRR genes in the genome v1.1 (Tuskan *et al.*, 2006), some of them being arranged in super clusters.

To further elucidate the relationships between  $R_I$  and  $R_{US}$  loci, we used an iterative process of genetic and physical mapping with the help of the complete genome sequence of *P. trichocarpa* Nisqually-1. At the beginning of our work we used the JAZZ assembly v1.1 (Tuskan *et al.*, 2006) and later the ARACHNE assembly v2.0 (<http://www.phytozome.org/>). For physical mapping, we constructed a deep coverage 15X BAC library for the individual *P. trichocarpa* 101-74 heterozygous at the  $R_{US}$  locus. The availability of BAC clones allowed the design of tightly closely linked markers through BAC end sequencing. An extension of the *P. deltoides* x *P. trichocarpa* segregating population up to 1415 individuals and its scoring for qualitative and quantitative resistance traits allowed to fine map the  $R_1$  and  $R_{US}$  resistance loci. Further characterization of the genomic regions surrounding  $R_1$  and  $R_{US}$  gave new insight on poplar genome organization. Optimal strategy to identify genes underlying  $R_1$  and  $R_{US}$  loci using a reference sequenced genome deprived from these alleles is discussed. Positional gene

cloning is still a major challenge in heterozygous species with duplicated genome and in complex regions as clusters of R-genes.

## **Materials and Methods**

### *Plant material*

The studied material was composed of (i) *P. deltoides* 73028-62, (ii) *P. trichocarpa* 101-74, (iii) exhaustive or selected sets of genotypes from four *Populus deltoides* x *P. trichocarpa* F<sub>1</sub> families generated by controlled hybridization with this male genitor (Table 1), and (iiii) the reference sequenced *P. trichocarpa* genotype ‘Nisqually-1’. The 73028-62 x 101-74 progeny comprises 1415 full-sib genotypes that were generated in two distinct years (1990, 2003). The set of genotypes generated in 1990 crosses already served as a mapping and QTL detection pedigree in a previous study (Jorge *et al.*, 2005). All genotypes were grown from cuttings under glasshouse rust-free conditions. Growth condition details can be found in Dowkiw *et al.*, (2003). *P. trichocarpa* 101-74 was later transferred in the dark to generate etiolated leaves for BAC library construction according to the protocol of Chalhoub *et al.*, (2004). DNA extractions for genotyping were made from the youngest leaves while rust resistance bioassays involved 3 cm diameter excised leaf-discs sampled from the fifth to the eighth fully expanded leaf below the apex. To ascertain the recombinant status of a few genotypes, both their phenotype and genotype were checked on leaves directly collected from original rootstocks in the nursery before rust epidemics.

### *Studied resistance segregating loci*

The  $R_1$ -mediated qualitative resistance to *Mlp* inherited from *P. deltoides* 73028-62 has been shown to segregate 1:1 in the 73028-62 x 101-74 F<sub>1</sub> progeny (Lefèvre *et al.*, 1998, Dowkiw *et*

*al.*, 2003). The  $R_{US}$ -mediated quantitative resistance inherited from *P. trichocarpa* 101-74 segregates independently from  $R_1$ , also with a 1:1 ratio, and generates clear bimodal distributions of genotypic means for uredinia size in the considered progeny (Dowkiw & Bastien, 2004). The 1:1 segregation pattern of  $R_{US}$  has been found in all hybrid progenies derived from 101-74 until now (Dowkiw & Bastien, 2007). In this work, we followed the nomenclature proposed by Dowkiw *et al.*, (2010):  $F_1$  individuals [ $R_{US}/r_{US}$ ] and [ $r_{US}/r_{US}$ ] will be designated [ $R_{US}$ ] and [ $r_{US}$ ]. The same rule was applied for the  $R_1$  locus.

#### *Complementary rust resistance assessments*

Genotypes from the 73028-62 x 101-74 progeny that were not previously characterized by Jorge *et al.*, (2005) as  $R_1$  vs.  $r_1$  and  $R_{US}$  vs.  $r_{US}$  were phenotyped using *Mlp* strains 93ID6 (virulences 3-4) and 98AG69 (virulences 1-3-4-5-7), respectively. This last strain has the advantage of being unaffected by any “residual” effect of  $R_1$  in this progeny (Dowkiw & Bastien, 2004).

The large number of genotypes studied imposed spreading them into three sub-experiments for each strain. Each experimental design consisted in five randomized complete blocks where each genotype was represented by one disk per block. Nisqually-1 was represented by three disks per block in one sub-experiment for each strain. Details of the inoculation procedure can be found in Dowkiw *et al.*, (2003). Temperature was set at 17°C. The mean inoculum pressure ranged between 132 and 343 urediniospores per leaf-disk. Uredinia size was measured using the 1-to-5 fast scoring method given in Dowkiw *et al.*, (2010).

The  $R_1$  phenotype was defined as the absence of any uredinia on any leaf disk after inoculation with strain 93ID6. Genotypes showing only one disk with uredinia during the first test had their phenotypes later validated in one to three complementary inoculation tests. The  $R_{US}$  and  $r_{US}$  phenotypes were defined graphically, by setting an upper (respectively, lower)

limit to the small (respectively large) mode of the bimodal distribution of clonal means observed for uredinia size (US) in each experiment. Genotypes falling between these two limits were defined as ambiguous and tested again with strain 98AG69 along with some reference genotypes for  $R_{US}$  and  $r_{US}$  phenotypes.

#### *Identification of markers linked to resistance locus through Bulk Segregant Analysis and AFLP screening*

A BSA-AFLP approach (Michelmore *et al.*, 1991) was carried out to identify markers closely linked to  $R_{US}$ . In order to limit the number of false positives and to overcome some problems linked to inheritance of dominant markers from the female parent, 5 independent pairs of  $R_{US}$  and  $r_{US}$  bulks were constituted (Table 1): The individuals from each bulk were also selected on their genotype at 3  $R_{US}/r_{US}$  flanking markers (Jorge *et al.*, 2005). Equal amounts of DNA from 10 genotypes were pooled in each bulk, and the AFLP procedure was applied as described in Jorge *et al.*, (2005). The polymorphic bands were further analysed at single genotype level. The confirmed bands were finally cloned and sequenced. Sequences were used for BLASTn alignment against genome sequence of *P. trichocarpa* Nisqually-1 v1.1 ([http://genome.jgi-psf.org/Poptr1\\_1/Poptr1\\_1.home.html](http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html)) to determine their physical location. SCAR markers were developed from the AFLP sequences and the help of genome sequence v1.1.

#### *Identification of PCR markers linked to resistance loci through genome sequence analysis*

First, 13 SSR markers were retrieved from former studies (Tuskan *et al.*, 2004; Yin *et al.*, 2004). SSR motifs were also searched in the region of the genome sequence surrounding the cloned AFLP markers. Second, STS primers inside and outside the gene spaces were also generated. STS in NBS-LRR genes were generated not only from the candidate genomic

region but also from genes located on other scaffolds but showing sequence similarity. In some cases, in order to develop specific primer pairs in NBS-LRR genes, a consensus primer was designed in the TIR or LRR domains and the specific primer was designed respectively in the 5' or 3' non coding flanking region. Third, primers were also designed in 3 TIR-NBS-LRR genes located on a published *P. deltoides* BAC sequence near the *Mer* locus (Lescot *et al.*, 2004).

Primer pairs were defined with the Primer 3 software using default parameters except for the primer length: 22 nucleotides and annealing temperature: 59°C (<http://fokker.wi.mit.edu/primer3>). The specificity of candidate primer pairs were checked against the genome sequence using both ePCR (Schuler, 1997) and BLASTn alignments. When necessary, CAPS, dCAPS, AS-PCR markers were also developed as described by Konieczny & Ausubel (1993) and Michaels & Amasino (1998), Bundock *et al.*, (2005) to recover polymorphism.

Primer pairs were tested for amplification and segregation on the two parents and six progenies of the mapping pedigree. The PCR reactions consisted in one denaturing step of 5 min at 94°C followed by 35 cycles of 20 sec at 94°C, 30 sec at variable temperature according to primer sequences, 40 sec at 72°C and a final elongation step of 7 min at 72°C.

#### *Construction of local genetic maps*

Genetic makers were first used to construct local genetic maps with 336 genotypes, 326 of them having been already used by Jorge *et al.*, (2005). The mapping work was then extended to 1415 individuals with a subset of markers for high resolution mapping. Markers with more than 40% of missing data and/or with more than 3.5% of genotyping errors were removed from the mapping analysis. The final order of the genetic markers was established manually starting with the physical order retrieved from the genome sequence v1.1. The raw data from

recombinant individuals in the target region were then inspected and the order was corrected to minimize number of recombination events. The map distances were calculated with MAPMAKER/EXP 3.0 using the Kosambi map function (Lander *et al.*, 1987). Linkage groups were drawn using MapChart v2.2 (Voorrips, 2002). The consensus genetic map of LG XIX was constructed using the map projection procedure implemented in Biomercator v2.1 of *P. deltoides* parental map on *P. trichocarpa* parental map (Arcade *et al.*, 2004).

*Construction of a P. trichocarpa (clone 101-74) BAC library, insert size, pools and BAC end sequences (BES)*

A BAC library was constructed from the *P. trichocarpa* male parent (clone 101-74) according to the procedure developed by Chalhoub *et al.*, (2004). Briefly, young etiolated leaves were used to isolate high molecular weight nuclear DNA, which was partially digested with *Hind*III and cloned into pIndigo BAC – vector (Epicenter). The library was arrayed in 384-well microtiter plates. The insert size of BAC clones was determined according Chalhoub *et al.*, (2004). Two types of BAC pools were prepared for PCR screening: plate pools and row pools. This genomic resource was completed by BAC-end sequencing in the same procedure as Tuskan *et al.*, (2006), Kelleher *et al.*, (2007) of the clones from the first 20 plates.

*Construction of physical maps*

PCR screening of plate pools from the BAC library representing 10X genome coverage were performed and if necessary PCR were also made with the remaining plate pools. Row pools corresponding to the positive plates were then screened and the 24 candidate clones of a positive row were individually tested for the identification of the positive BAC clones. The extremities of some positive BAC clones were sequenced and aligned onto the genome sequence by BLASTn. Concurrently, the BES of the first 20 plates were also aligned by

BLASTn (e-value= $10^{-50}$ ) on the genome sequence. All the selected alignments of BES in the region of interest manually inspected using the MEGA4 software (Tamura *et al.*, 2007). The construction of the BAC contigs and the orientation of BAC into the contigs were inferred from genetically mapped of newly markers designed in BES and/or from the alignment of the BES onto the genome sequence.

This approach allowed the construction of two physical maps, one for each haplotype. The haplotype carrying the  $R_{US}$  allele was named haplotype  $hR_{US}$  and the haplotype carrying the  $r_{US}$  allele was named haplotype  $hr_{US}$ .

#### *Domain identification of NBS-LRR genes in the target region*

The NBS-LRR genes MEME motifs established by Meyers *et al.*, (2003) and Kohler *et al.*, (2008) were used as tBLASTn queries to determine if gene models encoded NLS (Nuclear Localization Sequence), TIR, CC, BED, NBS or LRR domains.

## **Results**

#### *Complementary phenotyping for qualitative and quantitative resistances*

The segregations of both  $R_1$  and  $R_{US}$  fitted 1:1 ratios in the complementary set of 1079 genotypes from the 73028-62 x 101-74 progeny studied here. 702 genotypes could be defined as  $R_1$ , 645 as  $r_1$  ( $P > \chi^2$  1:1 = 0.120). Ambiguous genotypes for presence of qualitative resistance were all declared  $r_1$  after several inoculations with strain 93ID6 at high inoculum pressures. Graphical analysis of the clonal mean distributions for US allowed to declare 696 genotypes as  $R_{US}$ , 656 as  $r_{US}$  ( $P > \chi^2$  1:1 = 0.277), while 22 genotypes remained ambiguous for their  $R_{US}$  vs.  $r_{US}$  status after multiple testing with strain 98AG69. 348 genotypes were

categorized as  $R_1R_{US}$ , 337 as  $R_1r_{US}$ , 327 as  $r_1R_{US}$ , and 301 as  $r_1r_{US}$ , thus indicating independent segregation ratios for both resistance factors ( $P > \chi^2_{1:1:1:1} = 0.297$ ).

Nisqually-1 expressed high susceptibility with US performances close to those observed for  $r_{US}$  genotypes: average US scores in  $\text{mm}^2$  were  $0.1768 \pm 0.0002$ ,  $0.8169 \pm 0.0008$  and  $0.806 \pm 0.030$  for 210  $R_{US}$  genotypes, 235  $r_{US}$  genotypes and Nisqually-1, respectively.

#### *Identification of markers linked to resistance locus through BSA screening*

One-hundred-and-fourteen AFLP primers-enzyme combinations were tested to identify polymorphic bands between the 4 female parents and the male parent and between resistant and susceptible bulks derived from the 4 progenies. Twelve discriminating bands were found among parental clones and all the contrasting bulks. Seven polymorphic bands were confirmed by testing the segregation in 90  $F_1$  individuals. These bands together with two previously mapped AFLP markers (E4M2-7 and E5M5-7; Jorge *et al.*, 2005), and the RAPD marker M03/04\_480 (Villar *et al.*, 1996), were cloned and sequenced (Table S1). BLASTn alignments of the AFLP sequenced fragments onto the *Populus* genome sequence (v1.1) showed that 6 of them aligned with a cluster of NBS-LRR genes at the peritelomeric region of chromosome 19 (E5ME-7, E21M18, E23M2, E15M14\_a, E25M14 and E6M5), whereas 3 aligned to scaffold 117 (E1M4, E25M14 and M03/04\_480). Only 3 of these AFLP and the RAPD markers could be converted into SCAR genetic markers, the others were monomorphic.

#### *Genetic linkage maps of the $R_{US}$ and $R_1$ regions*

For the construction of *P. trichocarpa* genetic map surrounding the  $R_{US}$  region, 3 individuals showing a trisomic or triploid pattern were excluded. A total of 333 individuals were used to construct a local genetic map with 68 molecular markers expected on the LG XIX (5.4% of

missing data). The map resulted in 27 distinct loci in an interval bounded by two SSR markers, I\_524 and I\_2879 (Fig. 1a). The total length of the interval studied was 23.6 cM. Average map distances between loci were 0.87 cM, ranging between 0.3 and 3.6 cM. Marker segregation type and phase were indicated in S2. RGA<sub>m4-1</sub> designed from the *P. deltooides* BAC sequence near the *Mer* locus didn't amplify in our *P. deltooides* parent but it was located 0.6 cM away from *R<sub>US</sub>* on the *P. trichocarpa* map. Twenty-three robust markers were subsequently used for fine mapping using 1078 additional progenies. The fine genetic map of the target region was constructed with 91% of data after removing of 10 triploid or trisomic and 2 aneuploid individuals. The high-resolution map included 18 loci with *R<sub>US</sub>* spanning 10 cM (Fig. 1a). New recombination events allowed the separation of previous co-segregating markers. However, among the 23 markers, 4 groups of markers still showed co-segregation. The *R<sub>US</sub>* locus was mapped within a region of 0.5 cM flanked by the co-segregating makers RGA<sub>s135-1</sub> and Is<sub>165-1</sub> at 0.1 cM on one side and the marker 14N08-F at 0.4 cM on the other side.

A local genetic map for *R<sub>1</sub>* was constructed by using a set of 335 individuals after removing one individual showing trisomic/triploid pattern (17.4% of missing data). The local map covered 33.8 cM and consisted of 24 molecular markers distributed in 15 loci (Fig. 1c). Thirteen robust markers were subsequently used for fine mapping using 1413 individuals (19.7% of missing data), 2 individuals showing abnormal maternal meiosis pattern were removed. Finally, *R<sub>1</sub>* was located in a 1.9 cM interval, between two loci consisting on one side by the co-segregating markers I\_1211 and G\_79 (0.6 cM) and on the other side of marker I\_920 (0.9 cM) (Fig. 1 c).

#### *Comparative mapping and anchoring onto the genome sequence*

Thirteen markers (8 SSR, 1 RGA, 3 STS, 1 SCAR) were shared between the *P. trichocarpa* and *P. deltoides* parental maps constructed using 333 and 335 F<sub>1</sub> individuals, respectively. The marker order was perfectly conserved between the two maps (Fig. 1; Fig. S1). Discrepancies were nevertheless observed among genetic distances between some marker pairs, recombination rates being on average 1.4 higher in *P. deltoides* than in *P. trichocarpa* in this genomic region (Fig. S1). Interestingly, the recombination rate between the markers flanking *R<sub>I</sub>* (G\_79 and I\_920) was 2.3 fold greater in *P. deltoides* than in *P. trichocarpa*. Finally, the markers that were mapped both in *P. trichocarpa* and *P. deltoides* allowed us to determine that *R<sub>I</sub>* and *R<sub>US</sub>* are located at two different loci.

Sequences from all primers were subjected to BLASTn searches against the two versions of poplar genomic sequence (v1.1 and v2.0; Fig. 1c and Table S2). Twelve genetic markers had no hits on genome sequence v1.1 and seventeen had no hit on v2.0. For instance, the genetic markers 57L24-R and E6M5 had no homolog on v1.1 but are anchored on the superscaffold [i.e., chromosome] 19 of v2.0. The co-segregating markers E1M4/RGAs336 and PTs776-AS/PT1939-4/nc1432-1 had homologs on genome v1.1 but not on v2.0. Several major changes from the v1.1 to the v2.0 versions of the genome sequence were confirmed by our genetic maps. In the v2.0, the scaffold 117 bordered by the markers RGAs135-1 and 47N12-F was totally integrated in the chromosome 19 (Fig. S1 and S2). This integration was confirmed by the alignment of our *P. deltoides* genetic map (Fig. 1). The genetic mapping of markers derived from the scaffolds 4600, 2139, 2701 and 283 of genome sequence v1.1 confirmed their integration into the chromosome 19 in v2.0. The alignment of the *P. trichocarpa* genetic maps on the v1.1 revealed an inversion of one co-linear block bordering by the markers 04J18-R-1 and RGA1159 (Fig. S2). This inversion was also found with *P. deltoides* map, but was not found in the v2.0 of the improved assembly (Fig. 1). Some discrepancies in markers order are however still observed between the v2.0 genome sequence and the genetic maps

constructed here. Anchoring *P. trichocarpa* genetic markers on to v2.0 revealed a mis-assembly of 5 co-segregating markers (RGAs352-RGAs135-1). This mis-assembly was also retrieved after anchoring the *P. deltoides* map with the genome sequence. The closely linked markers RGAs2701-1 RGAs237-4 and RGAs237-1 appeared to be inverted at a long distance relative to their co-segregating markers. There were also 3 instances where marker position were inverted (PT2387, 47N12-F, 14F23-F). Finally, for several markers which were expected as multilocus, according to the results of BLASTn searches against the genome sequences v1.1 and v2.0, only 1 locus per marker could be unambiguously mapped. Nevertheless, in most cases, the order of the markers of *P. trichocarpa* and *P. deltoides* linkage maps was consistent with the order of the markers in the genome sequence.

#### *Haplotype physical maps around $R_{US}$ and $R_I$ loci and anchoring onto the genome sequence*

The 52, 224 BAC clones of the BAC library were organized in 136 plate pools and 2176 row pools. The insert size of random sampling of 91 BAC clones allowed the estimation of an average insert size of 150 kb. Three percent and 0.3% of the clones contain chloroplastic and mitochondrial DNA sequences, respectively. Based on an estimated genome size of 480 Mb, the BAC library was estimated to represent 15 *Populus* genome equivalents (15X). End sequencing was performed on 7,680 BAC clones from the first 20 plates. After removal of low quality sequences, a set of 13,536 useful BES were available (GenBank: HN280500-HN291979).

Ninety-five BAC clones were recovered by the PCR screening of our BAC library with 60 markers and end sequenced. Concurrently the 13,536 available BES were aligned onto the genome sequence, allowing integration of 20 additional BAC clones in the physical mapping work. Finally, 133 BES coming from 71 BAC were used to anchor the physical map onto the genome sequence (Fig. 2; Tables S3 and S4). Forty-five of them were used to derive new

markers (Table S2). These markers were used to orientate individual clones and to confirm the entire subcontig grouping. Twenty-one of them were used as genetic markers to ensure that the walk had not jumped to other regions of the genome (Fig. 2; Table S2). The physical map of the  $hR_{US}$  haplotype consisted of 76 BAC clones arranged in 2 super-contigs (Fig. 2; Fig. S3). The first, delimited by the markers RGA805 and 97G09-R-1, comprised 32 BAC and covered about 850 Kb. The second, delimited by the markers 4J18-1-R and 47N12-F, comprised 44 BAC and covered about 900 Kb. The physical map of  $hr_{US}$  haplotype comprised 39 BAC clones also distributed in 2 contigs. The largest one covered about 1 Mb and consisted of 30 BAC clones, 8 of them could not be orientated in the contig, because of the alignment of BES to multiple locations onto the genome sequence. The smallest one comprised 9 BAC clones and spanned around 350 kb. The gap observed between the two physical maps, was caused by the lack of BAC clones in our library.

Comparative analysis between the two haplo-physical maps revealed a longer physical map of at least 100 kb for  $R_{US}$  than for  $r_{US}$  between the markers I\_1211 and 78I01-R-1. The 1.7 Mb  $hR_{US}$  contig spans a genetic distance of 9.5 cM. The overall ratio of physical to genetic distance was about 180 kb/cM. However, the ratio was not uniform across the region. Ranging from 650 kb/cM between the markers RGAs352 and PTs589 to 35 kb/cM between the markers RGAs682-dc and RGAs735-AS.

Anchoring the genetic and physical maps onto the genome sequence v2.0 revealed a global conservation of marker order and physical distances (Fig. 2). However, a discrepancy was confirmed for a 200 kb region with 7 markers flanked by the markers RGAs135-1 and RGAs352, which was inverted and located 2 Mb further upstream than expected. In the second assembly, a 250 kb region corresponding to the top of scaffold 117 from v1.1 appears to be misassembled in v2.0 (Fig. S2).

Both physical and genome sequence showed complex regions of duplicated sequences between the markers S\_E5M5 and INRA1211. The physical length of this region is estimated to be 300 kb, where all the markers were multi-locus and the BES had multiple alignments. The complexity of this genomic area is also visible on the alignment of the two physical maps (Fig S3). Taking into account information coming from genetic, physical maps and alignment of the genome sequence v1.1 and v2.0; we showed that the large inversion initially identified on the genome assembly was corrected in v2.0 (Fig. S2).

#### *Genes in the interval containing the $R_{US}$ and $R_1$ loci*

The alignment of the consensus genetic map on the proposed local re-order allowed us to define homologous intervals for  $R_{US}$  and  $R_1$  onto the genome sequence (Fig. 3).  $R_1$  was located between the two markers G\_79 and I\_920 and  $R_{US}$  was located between Is\_165-1 and RGAs297 (Fig. 3). A tBLASTn alignment with MEME domains allowed us to precisely locate the candidate gene families in the two separate genomic regions.  $R_1$  was located in a region containing a cluster of complete or partial BED-NBS-LRR, TIR-NBS-LRR and 1 STK genes, whereas the  $R_{US}$  region only contained genes from the TIR-NBS-LRR class.

## **Discussion**

A previous mapping study has placed a major *Mlp* quantitative resistance locus (QRL), named  $R_{US}$ , on an unassigned LG of a *P. trichocarpa* parental map and a qualitative resistance locus  $R_1$  onto LG XIX of the *P. deltooides* parental map (Jorge et al, 2005). In the present study, we undertook comparative genetic mapping in order to precise location and linkage relationships between the QRL, identified in *P. trichocarpa* and the *R1* gene localized on the LG XIX of the *P. deltooides*. For that, it was determinant to score (phenotype) resistance conferred by the

QRL as a Mendelian classification. Thus, we took advantage of clear bimodal distribution of average uredinia size after inoculation with mono-uredinial *Mlp* strains to categorize clones from the mapping progeny in two classes showing contrasted levels of QR. This classification resulted in the accurate localization of  $R_{US}$  on LG XIX of the *P. trichocarpa* parent. Moreover, we could conclude from the alignment on the reference *P. trichocarpa* sequence of genetic markers common to both parental maps that  $R_1$  and  $R_{US}$  are two distinct loci,  $R_1$  being located between the  $R_{US}$  and the telomere.

In order to better precise the location of the two resistance loci, it was necessary to construct a fine-scale, integrated genetic and physical maps. For that purpose, a large segregating population of 1415 individuals and a deep genome coverage BAC library with an average insert size of 150 kb were created and used for an iterative process of physical mapping, fine-scale genetic mapping, and alignment to the reference genome sequence of *P. trichocarpa*. Genetic linkage mapping requires polymorphic markers that segregate within the mapping population in order to estimate recombination rate. Alternatively, physical mapping overcomes the need for segregating polymorphic loci and allows mapping in segments where recombination is suppressed. In addition, the availability of BAC clones allowed the design of tightly linked genetic and physical markers derived from BES. The combination of these approaches allowed resolving of both heterozygous status of our model, and the complex structure of the region of interest (Moroldo *et al.*, 2008). An ordering based on only the genome sequence would have resulted in establishing erroneous BAC contigs, due to miss-assembly caused by gene duplications, and to different genome organization between *P. trichocarpa* cv Nisqually-1 and *P. trichocarpa* 101-74 at these regions. Our study showed that the development of a fine and dense genetic map was crucial for ordering the BAC clones. Moreover, as the poplar genome is heterozygous, genetic markers allowed assignments of BAC clone to either  $hR_{US}$  or  $hr_{US}$  haplotypes. The presence of repeats and duplicated

sequences complicated the development of specific markers. NBS-LRR encoding gene clusters like the one studied here are particularly prone to such redundancy, both within a cluster and between different clusters (Tang *et al.*, 2010). Another constraint arose from the non-homogeneous pattern of heterozygosity between the two haplotypes of *P. trichocarpa* 101-74. The variability in both the level of redundancy and the heterozygosity led to a succession of intervals, some having high redundancy with moderate levels of heterozygosity, other showing the opposite situation, and finally, some having both high redundancy and high heterozygosity. Specific strategies were used for each type of interval. When the heterozygosity was high, designing markers in BES was generally efficient. When the redundancy was high, like in the large cluster of NBS-LRR encoding genes, intergenic areas were targeted for genetic marker development. On the whole, markers designed from BES were more successful than those derived from the genome sequence.

Comparison of the two haplotype physical maps revealed differences in their structural organization with at least 100 kb indel. Structural variations have been already reported in grape (Velasco *et al.*, 2007), rice (Zhou *et al.*, 2007) and suspected in poplar by Yin *et al.*, (2004) at vicinity of the *Mer* locus.

The alignment of the 101-74 integrated genetic and physical map with the poplar genome sequence also suggests a genomic inversion between Nisqually-1 and 101-74 (Fig. 1). However, the marker order was co-linear between the *P. deltoides* and *P. trichocarpa* 101-74 genetic maps reinforcing the hypothesis that the inversion is most likely due to an error in the sequence assembly of Nisqually-1. Indeed, it has been shown that automatic assembly of BAC contigs based on BAC fingerprints (Moroldo *et al.*, 2008, Scalabrin *et al.*, 2010) and shotgun sequence assembly (Vinson *et al.*, 2005) of highly heterozygous genome can lead to errors in the final output, where the high heterozygosity rate observed in the corresponding genomic area could affect the assembly quality. The construction in this study of a detailed

integrated genetic and physical map of the genome region encompassing the  $R_{US}$  and  $r_{US}$  alleles allowed us to determine a more precise organization of this complex genomic region.

The average relationship between physical and genetic distances in *Populus* was estimated to 200 kb/cM on average (Tuskan *et al.*, 2006). Substantial variation of this estimated rate was observed in this study along the cluster of resistance genes, ranging from 28 kb/cM to 600 kb/cM. Moreover, some intervals showed a complete suppression of recombination whereas others behaved like hot spots of recombination. Large variation of genetic to physical distances ratio has also been observed in barley in the vicinity of *Mla* (Wei *et al.*, 1999), in wheat around Lr10 (Stein *et al.*, 2000) and in poplar where a suppression of recombination occurred in the interval containing the *Mxc3* resistance gene (Stirling *et al.*, 2001). High rate of polymorphism between allelic sequences could explain suppression of recombination. Indeed, among the 13 genetic markers closely linked to  $R_{US}$ , all except one were dominant, suggesting an important sequence polymorphism between the two haplotypes.

As listed by Poland *et al.* (2009), several hypotheses about molecular mechanisms underlying the  $R_{US}$  locus could be suggested. The alignment of the *P. trichocarpa* local genetic map along the Nisqually-1 genomic sequence (v2.0 version) showed that the  $R_{US}$  locus colocalized in a peritelomeric region. This region comprises clusters of 83 potential NBS-LRR encoding genes, that one of the largest gene families in *Populus* genome (Kohler *et al.*, 2008, Yang *et al.*, 2008). Precisely, the closest flanking genetic markers define an interval of 116kb that spans four NBS-LRR coding genes and two other genes (1 potential NBS-LRR gene with no MEME motif and 1 gene with unknown function). Although the genotype containing  $R_{US}$  is not yet sequenced, these findings strongly suggest that  $R_{US}$  could be a member of an allelic series of the NBS-LRR genes. In different species, QRLs colocalizing with mapped known R-genes, some of which NBS-LRR-like genes, have been described in several plant species

(Ballini *et al.*, 2008; Tan *et al.*, 2008; Danan *et al.*, 2009; Wisser *et al.*, 2005; Geffroy *et al.*, 2009). QRL colocalizing with major R genes were also observed in poplar.  $R_I$ , which confers qualitative resistance defeated by virulence 1, shows “residual resistance” expressed by extended latent period and reduced uredinia size (Dowkiw *et al.*, 2003). Other R-genes, also inherited from *P. deltoides*, contribute significantly to QR observed in artificial inoculation tests (Dowkiw & Bastien, 2007). Consistently with the proposal of Jones & Dangl (2006), Zhang *et al.*, (2010) demonstrated that Effector-Triggered-Immunity associated to qualitative resistance participates to QR of *Arabidopsis* to *P. syringae*. Moreover, *Rcg1*, one of the rare map-based cloned QRL has been found to encode an NBS-LRR resistance gene (Broglie *et al.*, 2007). We cannot however completely exclude that the  $R_{US}$  belongs to another class of defense genes. The only QRL cloned, *pi21*, is recessive and is a proline-rich gene of unknown function that lacks similarity to any known defense-related major genes (Fukuoka *et al.*, 2009). Further characterization of  $R_{US}$  QRL and comparison of the corresponding sequences in different *Populus species* will give better insights in possible origins and evolution of disease resistances in this genus.

Less surprising is the fact that the alignment of the *P. deltoides* local genetic map along the v2.0 version showed that the  $R_I$  locus falls into a cluster of NBS-LRR encoding genes where a marker closely linked to the *Mer* locus is also present (Lescot *et al.*, 2004). In the same mapping pedigree, Lefèvre *et al.* (1998) suggested the presence of other R-genes at the close vicinity of  $R_I$  in order to explain segregations for compatibility to 93CV1 *Mlp* strain. Identification of rare recombinants by Dowkiw *et al.*, (2003) supported this hypothesis.

In order to identify which of these candidate genes encodes  $R_{US}$ , the full sequencing of one or two BAC clones that overlap the  $R_{US}$  locus will be initiated. It is expected that the number of genes present in the three haplotypes considered here (101-74  $hR_{US}$ , 101-74  $hr_{US}$  and Nisqually-1) may vary substantially as it is often observed for NBS-LRR gene clusters (Zhou

*et al.*, 2007). Moreover, the reference sequence from Nisqually-1 is a pseudomolecule that contains chimeric mixture of its maternal and paternal haplotypes. As such, the number of gene models proposed in the  $R_{US}$  region may not be complete or accurate. Finally, the observed co-linearity between *P. deltooides* and *P. trichocarpa* in the region should facilitate the future isolation of the  $R_1$  locus via comparative genome mapping.

## **Acknowledgements**

The authors thank I. Le Clainche, F. Bitton and C. Boussardon for technical assistance; J. Just and M. Villar for helpful discussions; P. Poursat and his collaborators at the Experimental unit of INRA Orléans for resistance assessments; Anne Françoise Adam Blondon for the insightful comments on the manuscript. This work was supported by INRA (Programme ECOGER) and the French Ministry for Agriculture and Fisheries (Programme 142, sous action 27).

## **References**

- Arcade A, Labourdette A, Falque M, Mangin B, Chardon F, Charcosset A, Joets J. 2004.** BioMercator: integrating genetic maps and QTL towards discovery of candidate genes. *Bioinformatics* **20**: 2324-2326.
- Ballini E, Morel JB, Droc G, Price A, Courtois B, Notteghem JL, Tharreau D. 2008.** A genome-wide meta-analysis of rice blast resistance genes and quantitative trait loci provides new insights into partial and complete resistance. *Molecular Plant-Microbe Interactions* **21**: 859-868.

**Broglie KE. et al. 2007.** E.I. du Pont de Nemours and Company, Pioneer Bred International, Inc.; University of Delaware. Polynucleotides and methods for making plants resistance to fungal pathogens, United States Patent 20060223102.

**Bundock PC, Cross MJ, Shapter FM, Henry RJ. 2006.** Robust allele-specific polymerase chain reaction markers developed for single nucleotide polymorphisms in expressed barley sequences. *Theoretical and Applied Genetics* **112**: 358-365.

**Cervera MT, Gusmão J, Steenakers M, Peleman J, Storme V, Vanden Broeck A, Van Montagu M, Boerjan W. 1996.** Identification of AFLP molecular makers for resistance against *Melampsora larici-populina* in *Populus*. *Theoretical and Applied Genetics* **93**: 733-737.

**Cervera MT, Storme V, Ivens B, Gusmão J, Liu BH, Hostyn J, Van Slycken J, Van Montagu M, Boerjan W. 2001.** Dense genetic linkage maps of three *Populus* species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) on AFLP and microsatellite markers. *Genetics* **158**: 787-809.

**Chalhoub B, Belcram H, Caboche M. 2004.** Efficient cloning of plant genomes into bacterial artificial chromosome (BAC) libraries with larger and more uniform insert size. *Plant Biotechnology Journal* **2**: 181-188.

**Danan S, Chauvin JE, Caromel B, Moal JD, Pellé R, Lefebvre V. 2009.** Major-effect QTLs for stem and foliage resistance to late blight in the wild potato relatives *Solanum*

*sparsipilum* and *S. spegazzinii* are mapped to chromosome X. *Theoretical and Applied Genetics* **119**: 705-719.

**Dowkiw A, Husson C, Frey P, Pinon J, Bastien C. 2003.** Partial Resistance to *Melampsora larici-populina* Leaf Rust in Hybrid Poplars: Genetic Variability in Inoculated Excised Leaf Disk Bioassay and Relationship with Complete Resistance. *Phytopathology* **93**: 421-427.

**Dowkiw A, Bastien C. 2004.** Characterization of Two Major Genetic Factors Controlling Quantitative Resistance to *Melampsora larici-populina* Leaf Rust in Hybrid Poplars: Strain Specificity, Field Expression, Combined Effects, and Relationship with a Defeated Qualitative Resistance Gene. *Phytopathology* **94**: 1358-1367.

**Dowkiw A, Bastien C. 2007.** Presence of defeated qualitative resistance genes frequently has major impact on quantitative resistance to *Melampsora larici-populina* leaf rust in *P. ×interamericana* hybrid poplars. *Tree Genetics & Genomes* **3**: 261–74.

**Dowkiw A, Voisin E, Bastien C. 2010.** Potential of Eurasian poplar rust to overcome a major quantitative resistance factor. *Plant Pathology* **59**: 523-534.

**Fukuoka S, Saka N, Koga H, Ono K, Shimizu T, Ebana K, Hayashi N, Takahashi A, Hirochika H, Okuno K et al. 2009.** Loss of function of a proline-containing protein confers durable disease resistance in rice. *Science* **325**: 998-1001.

**Geffroy V, Macadré C, David P, Pedrosa-Harand A, Sévignac M, Dauga C, Langin T. 2009.** Molecular analysis of a large subtelomeric nucleotide-binding-site-leucine-rich-repeat

family in two representative genotypes of the major gene pools of *Phaseolus vulgaris*. *Genetics* **181**: 405-19.

**Gumez JL, Lachaume R, Pinon J. 2000.** Historique de l'émergence de la race E4 de *Melampsora larici-populina*. Conditions de son développement en France dans les régions Nord-Pas-de-Calais et Picardie. Enseignements à en tirer. *Bulletin Trimestriel Centre de Populiculture du Hainaut* 3-76.

**Jones JD, Dangl JL. 2006.** The plant immune system. *Nature* **444**: 323-329.

**Jorge V, Dowkiw A, Faivre-Rampant P, Bastien C. 2005.** Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection. *New Phytologist* **167**: 113-127.

**Kelleher CT, Chiu R, Shin H, Bosdet IE, Krzywinski MI, Fjell CD, Wilkin J, Yin T, DiFazio SP, Ali J et al. 2007.** A physical map of the highly heterozygous *Populus* genome: integration with the genome sequence and genetic map and analysis of haplotype variation. *Plant Journal* **50**: 1063-1078.

**Kohler A, Rinaldi C, Duplessis S, Baucher M, Geelen D, Duchaussoy F, Meyers BC, Boerjan W, Martin F. 2008.** Genome-wide identification of NBS resistance genes in *Populus trichocarpa*. *Plant Molecular Biology* **66**: 619-636.

**Konieczny A, Ausubel FM. 1993.** A procedure for mapping Arabidopsis mutations using co-dominant ecotype-specific PCR-based markers. *Plant Journal* **4**: 403-410.

- Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg LA. 1987.** MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174–181.
- Lefèvre F, Pichot C, Pinon J. 1994.** Intra- and interspecific inheritance of some components of the resistance to leaf rust (*Melampsora larici-populina* Kleb.) in poplars. *Theoretical and Applied Genetics* **88**: 501-507.
- Lefèvre F, Goué-Mourier MC, Faivre-Rampant P, Villar M. 1998.** A Single Gene Cluster Controls Incompatibility and Partial Resistance to Various *Melampsora larici-populina* Races in Hybrid Poplars. *Phytopathology* **88**: 156-163.
- Legionnet A, Muranty H, Lefevre F. 1999.** Genetic variation of the riparian pioneer tree species *Populus nigra*. II. Variation In susceptibility to the foliar rust *Melampsora larici-populina*. *Heredity* **82**: 318-327.
- Lescot M, Rombauts S, Zhang J, Aubourg S, Mathé C, Jansson S, Rouzé P, Boerjan W. 2004.** Annotation of a 95-kb *Populus deltoides* genomic sequence reveals a disease resistance gene cluster and novel class I and class II transposable elements. *Theoretical and Applied Genetics* **109**: 10-22.
- Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW. 2003.** Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**: 809-834.

**Michaels SD, Amasino RM. 1998.** A robust method for detecting single-nucleotide changes as polymorphic markers by PCR. *Plant Journal* **14**: 381-385.

**Michelmore RW, Paran I, Kesseli RV. 1991.** Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proceedings of the National Academy of Sciences, USA* **88**: 9828-9832.

**Moroldo M, Paillard S, Marconi R, Fabrice L, Canaguier A, Cruaud C, De Berardinis V, Guichard C, Brunaud V, Le Clainche I et al. 2008.** A physical map of the heterozygous grapevine 'Cabernet Sauvignon' allows mapping candidate genes for disease resistance. *BMC Plant Biology* **8**: 66.

**Paillassa E. 2006.** Quelles pertes de croissance pour 7 cultivars de peuplier face aux attaques de la rouille E4 du mélèze. *Forêt-Entreprise* **168**: 60-63.

**Pichot C, Teissier du Cros E. 1993a.** Susceptibility of *Populus deltoides* BARTR. To *Melampsora larici-populina* and *M. alii-populina*. I. Qualitative analysis of a 6 X 6 factorial mating design. *Silvae Genetica* **42**: 179-188.

**Pichot C, Teissier du Cros E. 1993b.** Susceptibility of *Populus deltoides* BARTR. To *Melampsora larici-populina* and *M. alii-populina*. II. Quantitative analysis of a 6 X 6 factorial mating design. *Silvae Genetica* **42**: 188-199.

**Pinon J, Frey P. 2005.** Interactions between poplar clones and *Melampsora* populations and their implications for breeding for durable resistance. In: Pei M, McCracken AR, eds. Rust Diseases on Willow and Poplar. Wallingford, UK: CAB International, 139-154.

**Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ. 2009.** Shades of gray: the world of quantitative disease resistance. *Trends in Plant Science* **14**: 21-29.

*ology* **9**: 23.

**Scalabrin S, Troggio M, Moroldo M, Pindo M, Felice N, Coppola G, Prete G, Malacarne G, Marconi R, Faes G et al. 2010.** Physical mapping in highly heterozygous genomes: a physical contig map of the Pinot Noir grapevine cultivar. *BMC Genomics* **11**: 204.

**Schuler GD. 1997.** Sequence mapping by electronic PCR. *Genome Research* **7**: 541-550.

**Stein N, Feuillet C, Wicker T, Schlagenhauf E, Keller B. 2000.** Subgenome chromosome walking in wheat: a 450-kb physical contig in *Triticum monococcum* L. spans the Lr10 resistance locus in hexaploid wheat (*Triticum aestivum* L.). *Proceedings of the National Academy of Sciences, USA* **97**: 13436-13441.

**Stirling B, Newcombe G, Vrebalov J, Bosdet I, Bradshaw HD Jr. 2001.** Suppressed recombination around the MXC3 locus, a major gene for resistance to poplar leaf rust. *Theoretical and Applied Genetics* **103**: 1129–1137.

**Tamura K, Dudley J, Nei M, Kumar S. 2007.** MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**: 1596-1599.

**Tan MY, Hutten RC, Celis C, Park TH, Niks RE, Visser RG, van Eck HJ. 2008.** The *R<sub>Pi-mcd1</sub>* locus from *Solanum microdontum* involved in resistance to *Phytophthora infestans*, causing a delay in infection, maps on potato chromosome 4 in a cluster of NBS-LRR genes. *Molecular Plant-Microbe Interactions* **21**: 909-918.

**Tang P, Zhang Y, Sun X, Tian D, Yang S, Ding J. 2010.** Disease resistance signature of the leucine-rich repeat receptor-like kinase genes in four plant species. *Plant Science* **179**: 399-406.

**Tuskan GA, Gunter LE, Yang ZK, Yin TM, Sewell MM, DiFazio SP. 2004.** Characterization of microsatellites revealed by genomic sequencing of *Populus trichocarpa*. *Canadian Journal of Forest Research* **34**: 85–93.

**Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A et al. 2006.** The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596-1604.

**Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J et al. 2007.** A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One* **2**: e1326.

**Villar M, Lefèvre F, Bradshaw HD Jr, Teissier du Cros E. 1996.** Molecular genetics of rust resistance in poplars (*Melampsora larici-populina* Kleb/*Populus* sp.) by bulked segregant analysis in a 2 x 2 factorial mating design. *Genetics* **143**: 531-6.

**Vinson JP, Jaffe DB, O'Neill K, Karlsson EK, Stange-Thomann N, Anderson S, Mesirov JP, Satoh N, Satou Y, Nusbaum C et al. 2005.** Assembly of polymorphic genomes: algorithms and application to *Ciona savignyi*. *Genome Research* **15**: 1127-1135.

**Voorrips RE. 2002.** MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* **93**: 77-8.

**Wei F, Gobelman-Werner K, Morroll SM, Kurth J, Mao L, Wing R, Leister D, Schulze-Lefert P, Wise RP. 1999.** The Mla (powdery mildew) resistance cluster is associated with three NBS-LRR gene families and suppressed recombination within a 240-kb DNA interval on chromosome 5S (1HS) of barley. *Genetics* **153**: 1929-48.

**Wisser RJ, Sun Q, Hulbert SH, Kresovich S, Nelson RJ. 2005.** Identification and characterization of regions of the rice genome associated with broad-spectrum, quantitative disease resistance. *Genetics* **169**: 2277-93.

**Wu JL, Sinha PK, Variar M, Zheng KL, Leach JE, Courtois B, Leung H. 2004.** Association between molecular markers and blast resistance in an advanced backcross population of rice. *Theoretical and Applied Genetics* **108**: 1024-1032.

**Yang S, Zhang X, Yue JX, Tian D, Chen JQ. 2008.** Recent duplications dominate NBS-encoding gene expansion in two woody species. *Molecular Genetics and Genomics* **280**: 187-198.

**Yin TM, DiFazio SP, Gunter LE, Jawdy SS, Boerjan W, Tuskan GA. 2004.** Genetic and physical mapping of *Melampsora* rust resistance genes in *Populus* and characterization of linkage disequilibrium and flanking genomic sequence. *New Phytologist* **164**: 95-105.

**Yin T, Difazio SP, Gunter LE, Zhang X, Sewell MM, Woolbright SA, Allan GJ, Kelleher CT, Douglas CJ, Wang M et al. 2008.** Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. *Genome Research* **18**: 422-430.

**Zhang J, Steenackers M, Storme V, Neyrinck S, Van Montagu M, Gerats T, Boerjan W. 2001.** Fine Mapping and Identification of Nucleotide Binding Site/Leucine-Rich Repeat Sequences at the MER Locus in *Populus deltoides* 'S9-2'. *Phytopathology* **91**: 1069-1073.

**Zhang J, Lu H, Li X, Li Y, Cui H, Wen CK, Tang X, Su Z, Zhou JM. 2010.** Effector-triggered and pathogen-associated molecular pattern-triggered immunity differentially contribute to basal resistance to *Pseudomonas syringae*. *Molecular Plant-Microbe Interactions* **23**: 940-948.

**Zhou B, Dolan M, Sakai H, Wang GL. 2007.** The genomic dynamics and evolutionary mechanism of the *Pi2/9* locus in rice. *Molecular Plant-Microbe Interactions* **20**: 63-71.

## Figure legends

**Figure 1:** Local genetic maps of the target region and their alignment onto the scaffold 19 assembly v2.0.

Local genetic maps of *P. trichocarpa* derived from genotypic and phenotypic data for 333 and 1400 F<sub>1</sub> individuals (a) and *P. deltoides* (c) derived from genotypic and phenotypic data for 335 and 1413 F<sub>1</sub> individuals. Map distances in cM (Kosambi distances) are indicated on the left of linkage groups.

The *P. trichocarpa* and *P. deltoides* genetic maps were anchored on the genome sequence v2.0 (b) through ePCR and BLASTn searches using sequences of primers as queries and genome sequence as database.

SCAR markers coming from the AFLP-BSA approach were designated by the S\_AFLP primer combination names (E, *Eco*RI, M, *Mse*I) preceded by S\_.

Published SSR markers were named according Yin *et al.*, (2004), they begin by O\_, G\_ or P\_. SSR markers developed in this study begin by I\_ (for INRA).

Markers developed from NBS-LRR or STK genes were designated RGA, the letter “m” indicated that the marker was designed onto the BAC sequence located near the *Mer* locus (Lescot *et al.*, 2001).

Markers derived from gene models of genome sequence carried the prefix PT.

Markers derived from intergenic sequences begin by nc for non-coding sequences.

Markers derived from BES were named by plate&row letter&column number. The suffix F and R designed left (Forward end) and right (Reverse end) ends of BAC clones respectively.

The letter “s” in the marker names indicated that those markers derived from scaffold sequence v1.1.

AS, designated markers Allele Specific, dC dCAPS.

**Figure 2:** Alignments between the physical, genetic maps and the chromosome 19 sequence v2.0.

Local physical maps showing overlapping BAC clones of the haplotype  $hR_{US}$  (blue) (b) and the haplotype  $hr_{US}$  (green) (c). BAC clones were represented as vertical bars, light blue and light green bars designated BAC clones for which insert size were estimated, dark blue and dark green bars were used to show BAC clones for which insert size is unknown. BES represented by red rectangles show BLASTn alignment onto the chromosome 19 sequence, and orange rectangles indicated no BLASTn alignments on chromosome 19. BAC ends represented by black rectangles were found duplicated and they were not assigned with high confidence. Local genetic maps of *P. trichocarpa* derived from genotypic and phenotypic data for 335  $F_1$  individuals (c). The chromosome 19 sequence v2.0 (a) was aligned with physical map of  $hR_{US}$  haplotype. Red blocks represent gaps into the genome assembly; black blocks show sequences of the chromosome 19. Red and black connected lines designate alignments by BLASTn searches with BES or with marker primer sequences, respectively. Alignments of both BES and marker primer sequences are indicated by green connected lines. Markers note in italic designated markers for which no additional BAC clones were found in our BAC library.

**Figure 3:** Putatif resistance genes in the interval of  $R_{US}$  and  $R_1$

The consensus genetic map (a) between *P. trichocarpa* map of 1400  $F_1$  individuals and *P. deltooides* map of 1413  $F_1$  individuals were aligned onto the rearranged chromosome 19 v2.0 (b). Genetic distances in cM were noted on the left of the linkage groups. Markers in bold bordering  $R_{US}$  and  $R_1$  were mapped on consensus map and aligned onto the chromosome 19 sequence. Underlined markers were found only on the *P. trichocarpa* map. Markers in italic were found only on the *P. deltooides* map.

The gene models were designated by their number without the prefix POPTR\_0019s. n: NLS domain, T: TIR domain, B: BED domain, N: NBS domain and L: LRR domain. STK: Serine threonine kinase.

Discrepancies in the alignment between the consensus genetic map and the rearranged local scaffold 19 sequence could be due to length differences between individual genetic maps.

For marker details see Fig. 1 and S1.

### **Supporting information legends**

#### **Table S1**

Sequences of markers derived from the BSA approach.

#### **Table S2**

Primer pairs developed for genetic and physical marker development;

#### **Table S3**

BAC insert sizes and location of BES onto the genome sequence v2.0.

#### **Table S4:**

BAC end sequences in the genomic target region of physical maps.

#### **Fig. S1:**

Local genetic maps anchored onto the genome sequence v1.1.

#### **Fig. S2:**

Detail alignment of the two versions of the *Populus* genome v1.1 and v2.0 assemblies of the region encompassing the super cluster of resistance genes.

#### **Fig. S3:**

Alignment between the two haplotypic physical maps.

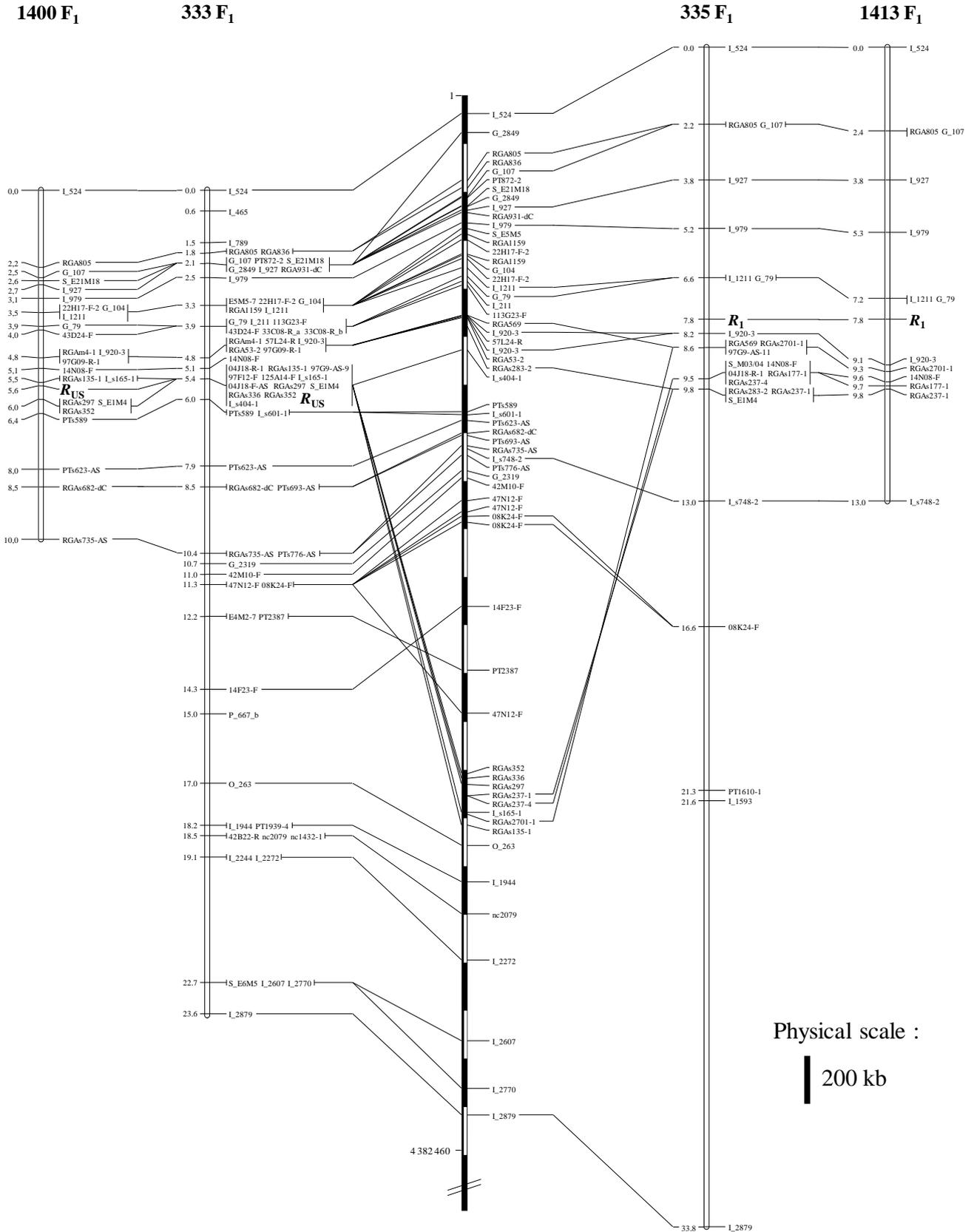
**Table 1** : List of *P. deltoides* x *P. trichocarpa* progenies material used in this study

Female parent	Male parent	Number of genotypes (selected phenotype)	Use	Source/Reference for phenotypes
73028-62	101-74	336 (all)	Local genetic map & High resolution mapping	Jorge et al. 2005
		1079 (all)	High resolution mapping	This study
		12 ( $r_1R_{us}$ )	BSA	Jorge et al. 2005
		12 ( $r_1r_{us}$ )		
		12 ( $r_1R_{us}$ )	BSA	Jorge et al. 2005
		12 ( $r_1r_{us}$ )		
M170-3	101-74	12 ( $R_{us}$ )	BSA	Dowkiw and Bastien (2007)
		12 ( $r_{us}$ )		
L155-079	101-74	11 ( $R_{us}$ )	BSA	Dowkiw and Bastien (2007)
		11 ( $r_{us}$ )		
L150-089	101-74	9 ( $R_{us}$ )	BSA	Unpublished data
		10 ( $r_{us}$ )		

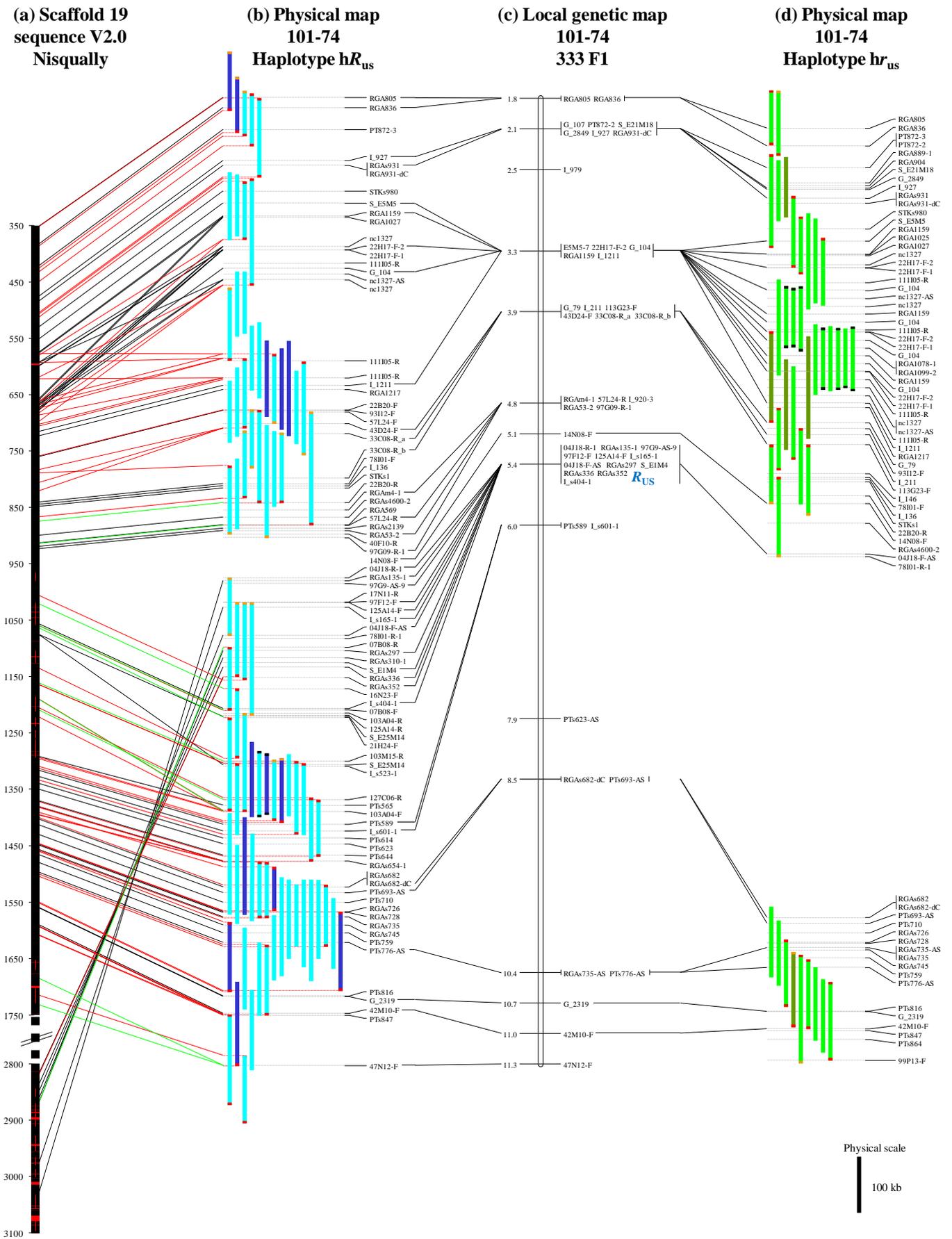
**(a) Local genetic maps  
*P. trichocarpa* (101-74)  
LG XIX**

**(b) Scaffold 19  
sequence v2.0  
*P. trichocarpa*  
(Nisqually-1)**

**(c) Local genetic maps  
*P. deltoides* (73028-62)  
LG XIX**



**Figure 1**



**Figure 2**

(a) Consensus genetic map  
 LG XIX  
*P. trichocarpa* (101-74 1400 F<sub>1</sub>)  
*P. deltoides* (73028-62 1413 F<sub>1</sub>)

(b) Local contig re-order  
 Scaffold 19 *P. trichocarpa*  
 (Nisqually-1) v2.0

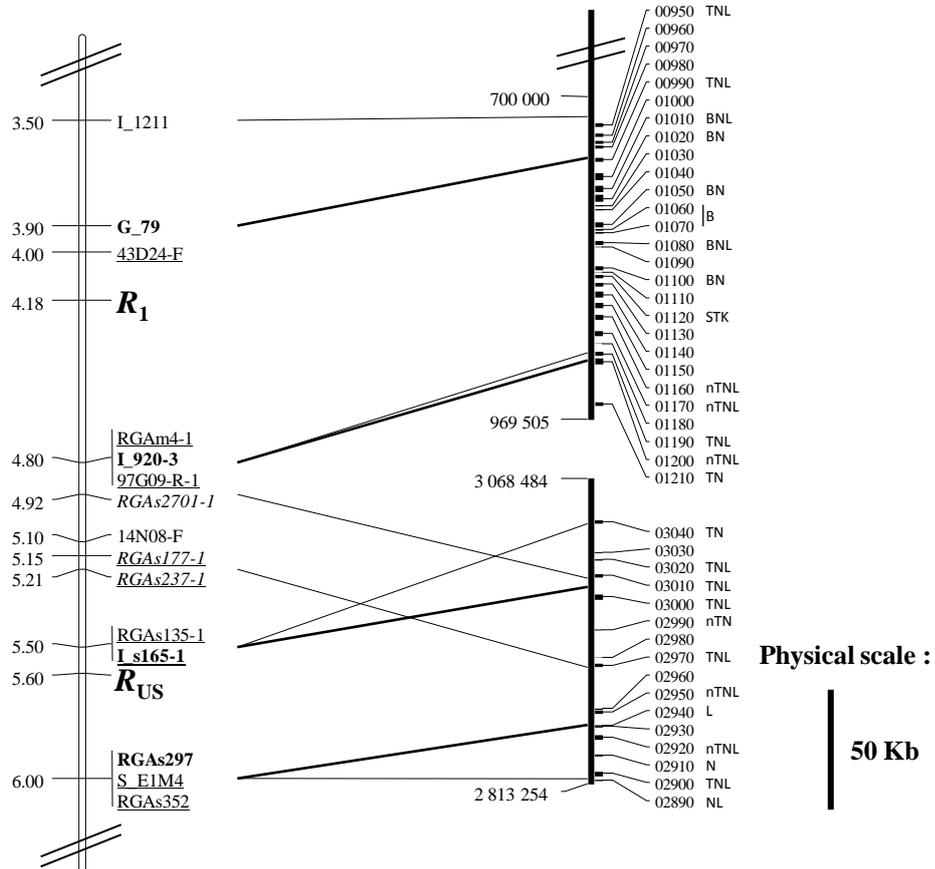


Figure 3

## Résumé

La rouille foliaire causée par le champignon *Melampsora larici-populina* (*Mlp*) est l'une des principales maladies affectant les peupleraies cultivées en Europe. Dans une population en ségrégation *Populus deltoides* x *P. trichocarpa*, un QTL majeur ( $R_{US}$ ) a été caractérisé. Ce locus, hérité de *P. trichocarpa*, contrôle la taille des sores lors de l'infection, paramètre quantitatif de la résistance à *Mlp*. Ce locus avait été cartographié sur l'extrémité du chromosome 19. Cette région génomique regroupe plusieurs dizaines de gènes (cluster) qui codent des protéines avec les domaines Nucleotide Binding Site (NBS) et Leucine Rich Repeats (LRR).

Cette étude porte sur la cartographie génétique et physique du locus  $R_{US}$ , avec l'aide du génome de *P. trichocarpa*. La première tâche a été d'ordonner correctement les séquences de l'extrémité du chromosome 19. Les cartes physiques locales pour chacun des haplotypes autour du locus  $R_{US}$  ont été construites à partir d'une banque d'ADN de grand fragment (BAC) de l'individu parent *P. trichocarpa*. Elles regroupent la plus grande partie du cluster de gènes NBS-LRR. La carte génétique fine a été construite avec une population en ségrégation de 2 144 individus. Le locus  $R_{US}$  a été cartographié sur une dans un intervalle de 0,14 cM. Le séquençage de clones BAC a permis d'identifier un seul gène NBS-LRR candidat.

Dans la même population, un second locus ( $R_1$ ) de résistance à *Mlp*, a été caractérisé sur la même extrémité du chromosome 19. Il contrôle une résistance qualitative, hérité de *P. deltoides*. La microsynchronie entre les deux espèces a été utilisée pour cartographier finement ce locus dans un intervalle de 0,66 cM. Cette cartographie comparée a permis aussi de déterminer que  $R_1$  et  $R_{US}$  sont deux loci indépendants.

## Abstract

Leaf rust caused by the fungus *Melampsora larici-populina* (*Mlp*) is one of the major diseases affecting poplar cultivation in Europe. In a segregating population *Populus deltoides* x *P. trichocarpa*, a major QTL ( $R_{US}$ ) has been characterized. This locus inherited from *P. trichocarpa*, controls the size of uredinia during the infection, a parameter of quantitative resistance to *Mlp*. This locus was mapped on the top of chromosome 19. This genomic region contains several tens of genes arranged in cluster that encode proteins with domains Nucleotide Binding Site (NBS) and Leucine Rich Repeats (LRR).

This study focuses on genetic and physical mapping of the  $R_{US}$  locus, with the help of the genome of *P. trichocarpa*. The first task was to order the correct sequence of the top of chromosome 19. Local physical maps for each haplotypes around the  $R_{US}$  locus were constructed from a DNA library of large fragment (BAC) of the individual parent *P. trichocarpa*. They comprise the largest part of the cluster of NBS-LRR genes. The fine genetic map was constructed with a segregating population of 2 144 individuals. The locus was mapped on  $R_{US}$  in an interval of 0.14 cM. The sequencing of BAC clones identified a single NBS-LRR candidate gene.

In the same population, a second locus ( $R_1$ ) of *Mlp* resistance was characterized at the same extremity of chromosome 19. It controls a qualitative resistance inherited from *P. deltoides*. Microsynteny between the two species was used to finely map this locus in an interval of 0.66 cM. This comparative mapping has also determined that  $R_1$  and  $R_{US}$  are two independent loci.