



**HAL**  
open science

**Sélection d'écotypes bactériens pathogènes et non-pathogènes par la plante en relation avec la différenciation en espèces génomiques chez *Agrobacterium* spp.**

Perrine Portier

► **To cite this version:**

Perrine Portier. Sélection d'écotypes bactériens pathogènes et non-pathogènes par la plante en relation avec la différenciation en espèces génomiques chez *Agrobacterium* spp.. Ecologie, Environnement. Université Claude Bernard - Lyon I, 2004. Français. NNT: . tel-00350502

**HAL Id: tel-00350502**

**<https://theses.hal.science/tel-00350502>**

Submitted on 6 Jan 2009

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

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

**THESE**

Présentée devant

L'UNIVERSITÉ CLAUDE BERNARD – LYON 1

pour l'obtention du

**DIPLÔME DE DOCTORAT**

(arrêté du 25 Avril 2002)

par

**Madame Perrine Portier**

---

**Sélection d'écotypes bactériens pathogènes et non-pathogènes par  
la plante en relation avec la différenciation en espèces génomiques  
chez *Agrobacterium* spp.**

---

présentée et soutenue publiquement le 22 Septembre 2004

Directeur de thèse : Mr Xavier Nesme

Jury :

Mr Johannes Geiselmann

Mr Charles Manceau

Mr Astolfo Zoina

Mr Timothy Vogel

Mr Christophe Mougel

Mr Xavier Nesme

## Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse, Xavier Nesme pour m'avoir encadrée au cours de quatre années de thèse.

Et un grand merci à Denis Faure qui m'a si efficacement encadrée durant mon DEA.

Je remercie également tous ceux qui ont pu m'aider de près ou de loin dans ce travail. Je ne peux citer tout le monde, j'en oublierai beaucoup trop, je vais me contenter de quelques dédicaces spéciales :

Pour ma famille et ma belle famille

Mes amis pas toujours proches géographiquement mais proches par le cœur, notamment Blandine. Et merci à Anne pour son aide précieuse dans les moments durs.

Pour tous ceux du laboratoire d'Ecologie microbienne de Lyon et plus particulièrement Catherine Lerondelle, Marie-Andrée Poirier et David Chapulliot pour leur précieuse aide technique et tous les étudiants pour le reste (En vrac : l'ambiance de la salle étudiants, les batailles d'eau ou de glace, les barbecues et repas de Noël, les soirées, les préparations de pots de thèse, tous ces moments qui font avancer la recherche dans la joie et la bonne humeur).

Pour toute l'équipe du laboratoire LAPM de Grenoble (feu Plasticité et Expression des Génomes Microbiens) qui m'a accueillie pour mon DEA.

Un grand merci à tous ceux qui m'ont encouragée dans la voie de la recherche :

Daniel de Montmollin, Michel Blot, Guy Lemperière, Anders Barfod, Kjeld Ingvorsen, Joergen Bungaard, Stuart Baker.

Une pensée pour Michel Blot qui nous a quitté en Septembre 2002.

Et une dédicace spéciale à Aymeric pour le week-end survie à Arrèche.

Mais surtout un très grand merci à Pierre pour tout ce qu'il a fait et fait toujours pour moi, et à Tristan pour son sourire rigolo.

## UNIVERSITE CLAUDE BERNARD - LYON I

### Président de l'Université

Vice-Président du Conseil Scientifique

Vice-Président du Conseil d'Administration

Vice-Présidente du Conseil des Etudes et de la Vie Universitaire

Secrétaire Général

**M. le Professeur D. DEBOUZIE**

M. le Professeur J.F. MORNEX

M. le Professeur R. GARRONE

M. le Professeur G. ANNAT

M. J.P. BONHOTAL

### SECTEUR SANTE

#### Composantes

UFR de Médecine Lyon R.T.H. Laënnec

UFR de Médecine Lyon Grange-Blanche

UFR de Médecine Lyon-Nord

UFR de Médecine Lyon-Sud

UFR d'Odontologie

Institut des Sciences Pharmaceutiques et Biologiques

Institut Techniques de Réadaptation

Département de Formation et Centre de Recherche en Biologie Humaine

Département de Formation à la Recherche et à l'Evaluation Pédagogiques

Directeur : M. le Professeur D. VITAL-DURAND

Directeur : M. le Professeur X. MARTIN

Directeur : M. le Professeur F. MAUGUIERE

Directeur : M. le Professeur F.N. GILLY

Directeur : M. O. ROBIN Maître de Conférences

Directeur : M. le Professeur F. LOCHER

Directeur : M. le Professeur L. COLLET

Directeur : M. le Professeur P. FARGE

Directeur : M. le Professeur M. LAVILLE

### SECTEUR SCIENCES

#### Composantes

UFR de Physique

UFR de Biologie

UFR de Mécanique

UFR de Génie Electrique et des Procédés

UFR Sciences de la Terre

UFR de Mathématiques

UFR d'Informatique

UFR de Chimie Biochimie

UFR STAPS

Observatoire de Lyon

Institut des Sciences et des Techniques de l'Ingénieur de Lyon

Département de 1<sup>er</sup> cycle Sciences

IUT A

IUT B

Institut de Science Financière et d'Assurances

Directeur : M. le Professeur J.L. VIALLE

Directeur : M. le Professeur H. PINON

Directeur : M. le Professeur H. BEN HADID

Directeur : M. le Professeur A. BRIGUET

Directeur : M. le Professeur P. HANTZPERGUE

Directeur : M. le Professeur M. CHAMARIE

Directeur : M. le Professeur M. EGEE

Directeur : M. le Professeur J.P. SCHARFF

Directeur : M. le Professeur R. MASSARELLI

Directeur : M. le Professeur R. BACON

Directeur : M. le Professeur J.P. PUAUX

Directeur : M. J.C. DUPLAN Maître de Conférences

Directeur : M. le Professeur M. ODIN

Directeur : M. le Professeur G. MAREST

Directeur : M. le Professeur J.C. AUGROS

<b>INTRODUCTION</b>	<b>5</b>
<b>I. DIVERSITÉ DES AGROBACTÉRIES EN RELATION AVEC LE CONCEPT D'ESPÈCES ÉCOLOGIQUES</b>	<b>7</b>
<b>Introduction</b>	<b>7</b>
<b>I. A. Les agrobactéries, agents de la Galle du Collet</b>	<b>8</b>
I. A. 1. Caractéristiques	8
I. A. 1. a. Position taxonomique	8
I. A. 1. b. Caractéristiques cellulaires	8
I. A. 1. c. Structure du génome	8
I. A. 2. Taxonomie des agrobactéries	9
I. A. 3. Mécanismes de Pathogénie	11
I. A. 3. a. Conséquences pour la plante	11
I. A. 3. b. Plasmide Ti et pathogénie	12
I. A. 3. c. Mécanismes de virulence	13
I. A. 3. d. Contrôle de la maladie	14
I. A. 4. Écologie des agrobactéries	15
I. A. 4. a. Le concept d'opine	15
I. A. 4. b. Survie des agrobactéries dans le sol	15
I. A. 4. c. Structure des populations	16
<b>I. B. Genres, espèces et souches bactériens</b>	<b>18</b>
I. B. 1. La classification chez les procaryotes	18
I. B. 1. a. Genre	19
I. B. 1. b. Souches	19
I. B. 2. Le concept d'espèces évolutionnaires	20
I. B. 3. Le concept de cohésion des espèces	20
I. B. 3. a. Hybridations ADN-ADN	21
I. B. 3. b. Définition actuelle de l'espèce bactérienne	22
I. B. 3. b. Les méthodes prometteuses pour la taxonomie des bactéries	23
I. B. 4. Concilier les définitions d'espèces eucaryotes et procaryotes	24
I. B. 4. a. Le sexe chez les bactéries	24
I. B. 4. b. mécanismes de régulation des transferts de gènes	25
I. B. 4. c. Barrière génétique et isolement sexuel	26
I. B. 5. Concept d'espèces écologiques	27
I. B. a. Ecotypes	27
I. B. 5. b. Caractérisation des écotypes	27

I. B. 5. c. Espèces et écotypes	28
<b>Conclusion et hypothèses de travail</b>	<b>29</b>
<b>II. ANALYSE DES POPULATIONS D'AGROBACTÉRIES ASSOCIÉES AUX PLANTES</b>	<b>31</b>
<b>II. 1. Analyse des populations d'<i>Agrobacterium</i> pathogènes isolées de tumeurs de peuplier □ intérêt de l'approche génomique.</b>	<b>31</b>
II. 1. 1 Introduction	31
II. 1. 2. Matériel et méthodes	32
II. 1. 2. a. Modèle végétal	32
II. 1. 2. b. Provenances des populations bactériennes analysées	32
II. 1. 2. c. Souches bactériennes et analyses AFLP.	33
II. 1. 3. Résultats	33
II. 1. 3. a. Identification des espèces génomiques	33
II. 1. 3. b. Diversité intra-spécifique	34
II. 1. 4. Discussion	34
II. 1. 5. Conclusion	36
<b>II. 2. Génétique des populations d'<i>Agrobacterium</i> associées à <i>Prunus persica</i> X <i>P. amygdalus</i> (GF677) □ épidémiologie génomique.</b>	<b>38</b>
II. 2. 1. Introduction	38
II. 2. 2. Matériel et méthodes	39
II. 2. 2. a. Isolement, purification, tests biochimiques et de pathogénie, culture et conservation des agrobactéries	39
II. 2. 2. b. Souches bactériennes	41
II. 2. 2. c. Analyses moléculaires	42
II. 2. 2. d. Analyses phylogénomiques	43
II. 2. 3. Résultats	44
II. 2. 3. a. Identification des espèces génomiques	44
II. 2. 3. b. Identification des clades dans les espèce génomique	45
II. 2. 4. Discussion	47
II. 2. 4. a. Structures génétiques intra-spécifiques	47
II. 2. 4. b. Epidémiologie moléculaire	48
II. 2. 4. c. Découverte d'une nouvelle espèce d' <i>Agrobacterium</i>	50
II. 2. 4. d. Succession des populations colonisant les rhizosphères	51
II. 2. 4. e. Association entre agrobactéries et GF677	51
II. 2. 4. f. Souches et écotypes	53
II. 2. 5. Conclusion	53
II. 3. 1. Introduction	55

II. 3. 2. Matériel et méthodes	56
II. 3. 2. a. Sol modèle	56
II. 3. 2. b. Plantes modèles et conditions de culture	56
II. 2. 3. c. Souches bactériennes	57
II. 2. 3. d. Traitement des isolats	57
II. 3. 2. e. Analyses moléculaires et analyses phylogénomiques	58
II. 3. 2. f. Séquençage du gène rrs	58
II. 3. 3. Résultats	59
II. 3. 3. a. Analyse des populations du sol	59
II. 3. 3. b. Analyse des populations piégées par les rhizosphères	59
II. 3. 4. Discussion	62
II. 3. 5. Conclusion	65

### **III. DEMASKING GENOME REGIONS INVOLVED IN *AGROBACTERIUM* SPECIES DELINEATION USING EXPERIMENTAL AND PREDICTED AFLP FRAGMENTS 69**

III. 1. Introduction	69
III. 2. Material and methods	71
III. 2. a. Bacterial strains	71
III. 2. b. Molecular analysis	71
III. 2. e. Phylogenomic analyses	73
III. 2. f. AlignK analysis	73
III. 2. g. Predictive AFLP	73
III. 3. Results	73
III. 3. a. AFLP analyses	73
III. 3. b. Consensus AFLP markers of genomic species	74
III. 3. c. Predictive AFLP	74
III. 3. d. Detected versus predicted AFLP fragments in strain C58	74
III. 3. e. Consensus AFLP markers: identification and localisation	75
III. 4. Discussion	76
III. 5. Conclusion	80

### **CONCLUSION ET PERSPECTIVES 81**

### **ANNEXE I : PROTOCOLES D'ISOLEMENT, PURIFICATION, TESTS BIOCHIMIQUES ET CONSERVATION DES AGROBACTÉRIES 84**

### **ANNEXE II : MILIEUX DE CULTURE ET D'ISOLEMENT**

<b>POUR <i>AGROBACTERIUM</i></b>	<b>86</b>
<b>ANNEXE III : PROTOCOLES ET SOLUTIONS POUR LES TESTS BIOCHIMIQUES</b>	<b>88</b>
<b>ANNEXE IV : PROTOCOLE AFLP</b>	<b>91</b>
<b>ANNEXE V : ANALYSE DES DONNÉES AFLP</b>	<b>95</b>
<b>REFERENCES BIBLIOGRAPHIQUES</b>	<b>111</b>



# Introduction

# Introduction

À l'interface entre le sol et les racines des plantes existe une niche écologique très spécifique, la rhizosphère (Gobat *et al.*, 2003). Celle-ci correspond à la fraction de sol adhérente aux racines. Dans la rhizosphère, la plante excrète de nombreux composés organiques, modifiant profondément ce milieu. Ceci en fait une niche écologique de choix, comparé au sol nu relativement pauvre en nutriments, pour les bactéries qui y vivent en grand nombre. Parmi ces bactéries colonisant les rhizosphères, certaines sont potentiellement pathogènes. Les bactéries du genre *Agrobacterium*, agents de la Galle du Collet (Conn, 1942) en sont un exemple. Si ces bactéries acquièrent, par conjugaison (Genetello *et al.*, 1977), le plasmide Ti (pour *Tumor inducing*, ou plasmide inducteur de tumeur), elle peuvent devenir phytopathogènes, provoquant la Galle du Collet (Conn, 1942). Cette maladie résulte du transfert à la plante d'un morceau d'ADN, l'ADN-T, porté par le plasmide Ti, induisant une tumeur productrice d'opines dans les tissus racinaires de la plante (Thomashow *et al.*, 1980). Les opines, composés carbonés, sont utilisées comme source de carbone par les agrobactéries ayant provoqué la tumeur, car le plasmide Ti porte aussi les gènes de dégradation de ces opines (Montoya *et al.*, 1977).

La relation entre les agrobactéries et les plantes est en fait une histoire à trois partenaires : la plante, le plasmide Ti et la bactérie.

Alors que les mécanismes d'infection de la plante par le plasmide Ti sont bien documentés (Zambryski, 1992), et que les liens entre les bactéries et le plasmide Ti ont été étudiés (Teyssier-Cuvelle, 1999 ; 2004), l'écologie de la bactérie indépendamment de la présence du plasmide Ti, ainsi que ses liens avec la plante, sont des domaines où peu de données sont disponibles. Or, si le plasmide Ti est le responsable de la galle du collet, il est véhiculé par les agrobactéries. Une meilleure connaissance de l'écologie de ces bactéries paraît nécessaire afin de déterminer les facteurs influençant ces populations, et par là, l'épidémiologie de la Galle du Collet.

Les travaux de thèse de Cécile Ponsonnet (1994) et de Christophe Mougel (2000), ont suggéré que les plantes pouvaient jouer un rôle dans la structure génétique des populations d'agrobactéries associées aux plantes.

D'autre part les agrobactéries (au sens large) appartiennent aussi bien à l'espèce *Rhizobium rhizogenes* (Sawada *et al.*, 1993) qu'au genre *Agrobacterium*, qui

comprend au moins une douzaine d'espèces différentes, dont certaines, celles du biovar1 (espèces G1 à G9), sont génétiquement très proches (Popoff *et al.*, 1984).

Dans une première partie, après analyse de la bibliographie, nous introduisons les différents concepts d'espèce bactérienne et ainsi que les connaissances actuelles concernant la taxonomie et l'écologie des agrobactéries.

A partir de ces données collectées par C. Ponsonnet et C. Mougel, notre point de départ a été de déterminer si les plantes pouvaient être impliquées dans la sélection différentielle d'espèces de bactéries dans le genre *Agrobacterium*. La reprise de ces données, ainsi que l'analyse d'une nouvelle épidémie de Galle du Collet, nous a permis de préciser les effets de la plante sur la structure des populations d'agrobactéries. Nous avons complété ces études par une analyse des agrobactéries non-pathogènes (sans plasmide Ti) naturellement associées aux rhizosphères de différentes plantes modèles. Ceci nous a permis de montrer la puissance de l'AFLP comme outil de suivi épidémiologique et de confirmer un certain effet des plantes sur les populations d'agrobactéries.

Dans la troisième partie de ce travail de thèse, nous avons tiré partie de la disponibilité de la séquence complète du génome de la souche modèle C58 (souche type de l'espèce G8) et de l'analyse AFLP d'une large gamme de souches de référence, pour identifier de nombreux marqueurs génomiques impliqués dans la définition des espèces G1, G4 et G8 du biovar1, et identifier par simulation de l'AFLP sur le génome de la souche C58, les gènes et les fonctions concernés par ces marqueurs génomiques.

Partie I  
Étude bibliographique

Diversité des agrobactéries en relation avec le  
concept d'espèces écologiques

# I. Diversité des agrobactéries en relation avec le concept d'espèces écologiques

## Introduction

Les agrobactéries sont des bactéries naturellement présentes dans les sols. Elles appartiennent aux genres *Agrobacterium* (Conn, 1942), qui comprend une douzaine d'espèces réparties en différents biovars (Popoff et al., 1984) et *Rhizobium* (Sawada et al., 1993). Si elles acquièrent le plasmide Ti (*Tumor inducing*), elles peuvent devenir pathogènes, provoquant la Galle du Collet (Smith and Townsend, 1907). Cette maladie se traduit par l'apparition d'une tumeur au niveau du collet de la plante. Cette tumeur est provoquée par le transfert à la plante d'une portion d'ADN du plasmide Ti. Du fait de leur importance économique, ces bactéries ont été étudiées d'un point de vue phytopathologique et surtout elles ont beaucoup intéressé le domaine de la biotechnologie. En effet, le plasmide Ti, qui porte les gènes de virulence chez *Agrobacterium*, a permis la construction des plantes génétiquement modifiées consistant en l'insertion de gènes d'intérêt dans l'ADN du plasmide Ti pour être transféré au génome de la plante (Bevan, 1984 ; Valentine, 2003). Depuis 2001 la séquence complète de la souche modèle C58 est disponible (Wood et al. 2001 ; Goodner et al., 2001).

L'épidémiologie de la Galle du Collet (Nesme et al., 1997 ; Ponsonnet et al., 1994 ; Pionnat et al., 1999a) a montré la très grande diversité, au niveau des souches, des espèces et des genres, des agrobactéries impliquées dans une même épidémie. Cette diversité des agents pathogènes est une caractéristique très originale des épidémies de Galle du Collet qui ne se rencontre pas, à notre connaissance, dans d'autres modèles pathogènes végétaux ou animaux.

En épidémiologie, la définition des espèces bactérienne est cruciale, permettant de déterminer avec précision les microorganismes impliqués dans une épidémie. Les agrobactéries étant très diverses, l'épidémiologie de cette maladie requiert l'analyse de nombreux isolats (Nesme et al., 1997). Une question particulière réside dans la connaissance du rôle relatif des espèces bactériennes impliquées qui pourrait permettre de mieux connaître l'écologie de cet agent phytopathogène original.

Dans une première partie je présenterais les connaissances actuelles concernant les agrobactéries, agents de la Galle du Collet. Ensuite, je définirai les souches, espèces et genres bactériens tels que nous allons les utiliser dans la suite de ce travail de thèse.

## I. A. Les agrobactéries, agents de la Galle du Collet

### I. A. 1. Caractéristiques

#### I. A. 1. a. Position taxonomique

Le genre *Agrobacterium*, ainsi que l'espèce *Rhizobium rhizogenes*, appartiennent à la famille des *Rhizobiaceae*, des Alphaprotéobactéries. Elles sont parfois définies comme des *Rhizobium* qui ne fixent pas l'azote mais produisent d'autres types d'hypertrophies (Woese, 1987 ; Kersters and De Ley, 1984). Cependant, la plupart des agrobactéries (au sens large) sont des bactéries saprophytes du sol, ni pathogènes, ni symbiotiques.

#### I. A. 1. b. Caractéristiques cellulaires

Les agrobactéries ont la forme de batonnets de 0,6 – 1,0 µm x 1,5 – 3,0µm. Les cellules sont à Gram négatif et vont seules ou par paires. Elles ne forment pas de spores et sont mobiles grâce à des flagelles (de 1 à 6) péritriches. Elles sont chimioorganotrophes aérobies. Mais la plupart des souches d'agrobactéries sont capables de pousser sous une pression d'oxygène réduite (ce qui est le cas dans les tissus végétaux par exemple). La température optimale de croissance de ces bactéries se situe entre 25 et 28°C. Sur un milieu contenant des sucres elles produisent beaucoup d'exopolysaccharides, les colonies ont un aspect très muqueux. Elles produisent de la catalase et généralement aussi de l'uréase et de l'oxidase. Certaines souches regroupées dans le biovar1 produisent des 3-cétoglycosides (Moore *et al.*, 1988).

#### I. A. 1. c. Structure du génome

Le paradigme qui à longtemps prévalu pour l'organisation génomique des bactéries est issu des études menées sur *Escherichia coli* et *Bacillus subtilis*. L'organisation génomique de ces bactéries est assez simple : tous les gènes essentiels à la survie

Feature	Circular	Linear	pAtC58	pTiC58	Total
Size (bp)	2,841,490	2,075,560	542,779	214,233	5,674,062
G+C content (%)	59.4	59.3	57.3	56.7	58.13
Protein-coding genes					
Assigned function	1715	1286	333	141	3475 (64.1%)
Conserved hypothetical	710	353	128	45	1236 (22.8%)
Hypothetical	364	243	89	12	708 (13.1%)
Total	2789	1882	550	198	5419
Average ORF size (bp)	892	988	843	925	922
Coding (%)	87.9	89.9	85.4	85.5	88.3
Regulators (%)	7.7	10.4	11.8	5.1	9.0
ABC transport	47	80	20	6	153
RNA					
rRNA	2	2	0	0	4
tRNA	40	13	0	0	53
tmRNA	1	0	0	0	1
IS elements	2	10	10	2	24
Phage-related	10	1	1	0	12

Tableau 1.

Les quatre réplicons du génome de la souche modèle d'*Agrobacterium* C58 et leurs caractéristiques. (Wood *et al.*, 2001).

de la bactérie sont réunis sur un unique chromosome circulaire. La situation est différente pour nombre de bactéries, notamment pour les agrobactéries (Jumas-Bilak et al., 1998).

Le nombre et la taille des réplicons des différentes souches d'agrobactéries est variable. Elles possèdent deux chromosomes ou un chromosome et un méga-réplicon pour l'espèce *R. rhizogenes*. Le plus grand est toujours circulaire, de taille allant de 2,8 à 4Mb. Le deuxième est soit linéaire (Biovar1 et *A. rubi*) soit circulaire (*R. rhizogenes* et *A. vitis*), de taille comprise entre 1 et 2,6Mb. Il n'est pas certain que le méga-réplicon des souches de *R. rhizogenes* soit réellement un chromosome. Le contenu plasmidique des souches est lui aussi variable (Jumas-Bilak et al., 1998 ; Urbanczyk et al., 2003).

La souche modèle de biovar1, C58, qui a été complètement séquencée deux fois. (Wood et al., 2001 ; Goodner et al. 2001). Elle possède quatre réplicons. Un chromosome circulaire (CcC58, 2,8Mb) un chromosome linéaire (LcC58, 2Mb), un plasmide cryptique At (pAtC58, 0,55Mb) et un plasmide Ti (pTiC58, 0,2Mb) (Tableau 1). Le chromosome linéaire des souches de biovar1 serait dérivé d'un mégaplasmide (Goodner et al., 2001).

La comparaison, par hybridation, entre le chromosome circulaire du biovar1 et le chromosome de *R. rhizogenes* a montré qu'ils sont très proches. Par contre il n'y a que très peu de similarité entre le chromosome linéaire du biovar1 et le méga-réplicon de *R. rhizogenes*. Le phénomène d'acquisition d'un chromosome secondaire ne serait donc pas un fait unique dans l'histoire évolutive des agrobactéries (Urbanczyk et al., 2003). A l'heure actuelle, deux autres autres agrobactéries sont en cours de séquençage ; une souche d'*Agrobacterium vitis*, et la souche K84 de l'espèce *Rhizobium* (« *Agrobacterium* ») *rhizogenes*.

### **I. A. 2. Taxonomie des agrobactéries**

D'abord décrit par Smith et Townsend en 1907 (Smith and Townsend, 1907), Conn à défini le genre *Agrobacterium* en 1942 (Conn, 1942). Il y a inclus les espèces *A. radiobacter* (i.e. non-pathogène) (Beijerinck and Van Delden, 1902) et *A. rhizogenes* (i.e. induisant la maladie du Chevelu Racinaire)( Riker et al., 1930).

Les premières classifications du genre *Agrobacterium* étaient donc basées sur la pathogénie. Les *Agrobacterium* qui induisent des tumeurs ont été appelés *A. tumefaciens*, ceux induisant des racines surnuméraires *A. rhizogenes* et les



*Agrobacterium* non pathogènes *A. radiobacter*. Une quatrième espèce, infectant les plantes du genre *Rubus*, à été décrite en 1943 par Starr et Weiss, il s'agit d'*Agrobacterium rubi* (Starr and Weiss, 1943 ; Hildebrand, 1940). En fait, les caractéristiques pathogènes reposent sur la présence ou l'absence des plasmides Ti ou Ri, et les plasmides étant accessoires, cette taxonomie n'est pas valide.

Les travaux de Keane *et al.* en 1970, ont montré qu'indifféremment des espèces précédemment décrites, les souches du genre *Agrobacterium* se répartissent en deux biovars. Puis en 1973, Kersters *et al.* ont défini un troisième biovar. Diverses propositions ont été faites pour faire correspondre les biovars et les espèces, mais rien d'officiel n'a été adopté.

Dans l'édition de 1984 du Bergey's manual (Kersters and De Ley, 1984) le genre *Agrobacterium* est divisé en quatre espèces :

*Agrobacterium tumefaciens* (Smith and Townsend, 1907) Conn 1942.

Cette espèce comprend les souches tumorigènes des biovars 1, 2 et 3.

*Agrobacterium radiobacter* (Beijerinck and Van Delden, 1902) Conn 1942.

Cette espèce comprend les souches non pathogènes des biovars 1, 2 et 3.

*Agrobacterium rhizogenes* (Riker *et al.*, 1930) Conn 1942.

Cette espèce comprend les souches rhizogènes (induisant le développement de racines surnuméraires) du biovar2. Les souches rhizogènes du biovar1 peuvent être appelées *A. rhizogenes* biovar1.

*Agrobacterium rubi* (Hildebrand 1940) Starr et Weiss 1943.

Cette espèce regroupe les souches pathogènes des plants de *Rubus sp.*

En 1984, Popoff *et al* ont effectué les hybridations ADN-ADN entre de nombreuses souches représentantes du genre *Agrobacterium*. Ces résultats montrent que les souches du biovar1 se répartissent en 9 espèces génomiques distinctes. Par contre il est très difficile de les séparer selon des critères phénotypiques, sur des galeries Api ou Biolog, par exemple (Marion Le Saux, communication personnelle). De plus, certaines espèces partagent la même séquence du gène *rrs* , c'est notamment le cas pour les espèces G6 et G8 ou encore G2 et G9 (Mougel *et al.*, 2001).

Autres espèces du genre *Agrobacterium*

En 1990 Ophel et Kerr (Ophel and Kerr, 1990) ont décrit l'espèce *Agrobacterium vitis*, qui correspond au biovar3 d'*Agrobacterium*. Bouzar et Jones en 2001 ont décrit l'espèce *Agrobacterium larrymoorei*, qui infecte les plants de *Ficus benjamina*.

La taxonomie du genre *Agrobacterium* est toujours en évolution. En 2001 Young *et al.*, à la suite de Sawada *et al.* (1993), ont proposé de supprimer le genre *Agrobacterium* et d'inclure ces bactéries dans le genre *Rhizobium* car l'analyse du gène *rrs* montrait que les bactéries nommées alors *Agrobacterium* avaient une origine polyphylétique. Farrand *et al.* en 2003 ont répondu que les bactéries du biovar 1 d'*Agrobacterium*, ainsi que les espèces *A. vitis* et *A. rubi*, formaient une entité cohérente dans la famille des Rhizobiacées. Ils proposent donc de conserver le genre *Agrobacterium*. Bien que l'espèce *A. rhizogenes* soit plus proche du genre *Rhizobium*, Farrand *et al.* (2003) proposent pour l'instant de conserver cette espèce dans le genre *Agrobacterium*.

Dans ce travail nous suivrons la proposition de Farrand *et al.* de conserver le genre *Agrobacterium* comme décrit précédemment. Cependant, pour tenir compte du réel problème posé par la polyphylie des dénominations précédentes, et parce que beaucoup d'arguments prèchent pour ce rapprochement, nous appellerons *Rhizobium rhizogenes*, les agrobactéries du biovar 2. Dans ce travail, nous distinguerons les agrobactéries pathogènes hébergeant un plasmide Ti ou Ri des non-pathogènes dépourvues de plasmide Ti ou Ri. Néanmoins, nous pensons qu'une nomenclature raisonnable devrait garder, outre les noms de genre et d'espèce, une indication de la pathogénie en ce référant au concept de pathovar, c'est à dire *Agrobacterium* sp. plus la dénomination pathovar tumefaciens ou pathovar rhizogenes.

Les agrobactéries de biovar 1 seront définies selon le numéro d'espèce génomique (G1 à G9) que leur a donnée Popoff *et al.* en 1984. Il est à noter que la souche type de l'espèce *A. tumefaciens* est B6 (= ATCC23308). Comme B6 appartient à l'espèce G4, G4 est l'unique *Agrobacterium tumefaciens* au sens des règles de la taxonomie.

### **I. A. 3. Mécanismes de Pathogénie**

#### *I. A. 3. a. Conséquences pour la plante*

Les plantes infectées par des agrobactéries pathogènes développent une tumeur au niveau du collet (Galle du Collet, *Crown Gall*), ou des racines adventives surnuméraires (Cheveulu Racinaire, Hary Root). Ces bactéries ont un très large spectre d'hôte puisqu'au moins 634 plantes réparties dans 331 genres et 93 familles y sont sensibles (De Cleene and De Ley, 1976). La seule souche B6 infecte près de

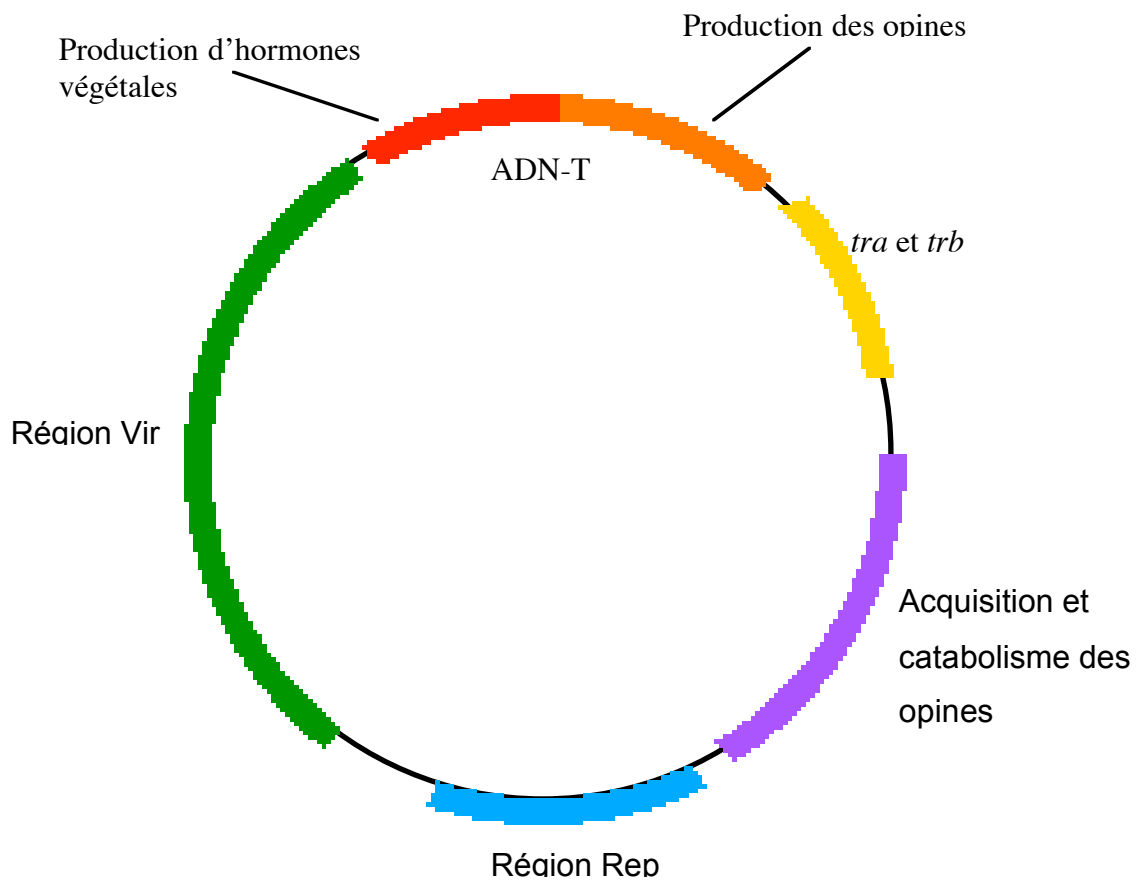


Figure 1  
Schéma du plasmide Ti et de ses différentes régions géniques

280 espèces différentes. La vigne, les fruitiers à noyaux, les fleurs et le peuplier sont très affectés.

La maladie provoque une baisse de la production de la plante qui est moins vigoureuse et dans des cas extrêmes peut provoquer la mort de la plante.

Les dégâts les plus importants sont observés en serre ou en pépinières (Pionnat *et al.*, 1999b ; Beneddra *et al.*, 1996). En serre, les conditions environnementales sont très favorables aux agrobactéries. En pépinière, le greffage des plantes facilite la maladie. Les plants infectés ne peuvent être commercialisés (Directive 93/64/CEE). Le manque à gagner pour les pépiniéristes peut être considérable. En outre, c'est une maladie de quarantaine dans de nombreux pays d'Afrique du nord et d'Amérique du sud.

### *I. A. 3. b. Plasmide Ti et pathogénie*

La galle du collet est liée à la présence dans les bactéries du plasmide Ti (pour *Tumor inducing*, plasmide inducteur de tumeur) (Van Larebeke *et al.*, 1974). Les racines adventices surnuméraires sont liées à la présence du plasmide Ri (pour *Root Inducing*, plasmide inducteur de racines). Ces plasmides se disséminent dans la population bactérienne par transfert conjugatif (Genetello *et al.*, 1977). Dans la suite de cette étude je parlerai essentiellement du plasmide Ti.

Les agrobactéries sont attirées par des composés phénoliques (tel l'acétosyringone) libérés par les cellules végétales lorsqu'elles sont blessées (Kado, 1991 ; Stachel *et al.*, 1985). Au niveau de cette blessure, la bactérie transfère à la plante une portion du plasmide Ti qui va ensuite s'intégrer dans le génome nucléaire de la plante (Chilton *et al.*, 1977). Les gènes transférés à la plantes induisent la biosynthèse par la plante des hormones de croissance, ce qui provoque la tumeur, et d'opines, qui sont des composés azotés. Le plasmide Ti comporte aussi les gènes de catabolisme des opines. Ceux-ci ne sont pas transférés à la plantes. Ils permettent à la bactérie d'utiliser ces composés comme source de nutriments. La plante transformée par la bactérie offre à son hôte « le gîte et le couvert » (Zhu *et al.*, 2000), c'est à dire une niche écologique.

Ce plasmide porte cinq types de gènes (Zhu *et al.*, 2000) (Figure 1) :

#### La région T (ADN-T)

Il s'agit du segment transféré à la plante. Cette région porte les gènes de production d'hormone végétale (auxine et cytokine) qui provoque la tumeur, et les gènes de

production d'opines, qui seront utilisées comme source de carbone par les agrobactéries.

#### La région Vir

Ces gènes permettent le transfert de l'ADN-T vers la plante. Ils sont communs aux différents types de plasmides Ti. Elle comporte 20 gènes essentiels à l'induction de tumeur chez l'hôte.

#### Les locus *tra* et *trb*

Ces gènes sont impliqués dans le transfert conjugatif du plasmide. Ils sont activés par les opines.

#### La région Rep

Les trois gènes de cette région permettent une réplication stable du plasmide et sont impliqués dans les fonctions d'incompatibilité.

#### Acquisition et catabolisme des opines

Cette quarantaine de gènes permet aux agrobactéries d'utiliser les opines produites par la plante comme source de nutriments.

### *I. A. 3. c. Mécanismes de virulence*

La tumeur est provoquée par le transfert de l'ADN-T de la bactérie vers la plante. Ce processus est dérivé d'un transfert conjugatif. Les gènes qui régulent ce processus sont exprimés en réponse à des signaux chimiques libérés par l'hôte. Il se déroule en sept étapes (Hooykaas and Beijersbergen, 1994 ; Zambryski, 1992).

#### Reconnaissance d'une cellule sensible

Les composés phénoliques (l'acétosyringone et l'acide sinapinique en particulier) excrétés par les cellules blessées attirent les agrobactéries (Kado, 1991 ; Stachel *et al.*, 1985)

#### Attachement de la bactérie à la cellule végétale

Quand la bactérie arrive sur la blessure un attachement réversible faible a lieu. Puis la bactérie produit des fibrilles de cellulose qui vont l'ancrer fermement à la surface de l'hôte. Les gènes chromosomiques *chvA*, *chvB*, production et sécrétion de  $\beta$ -1-2-glucane, et *pscA* sont impliqués dans cet attachement essentiel à la pathogénie.

#### Induction de l'expression des gènes vir

En présence de sucre, en pH acide, et sous l'action de l'acétosyringone végétale, la protéine VirA s'autophosphoryle et transfère son groupement phosphate à la protéine VirG. Ceci permet l'activation de tout l'opéron vir.

### Production d'une copie transférable de l'ADN-T

Il y a formation d'un complexe T stable. L'ADN-T est protégé par des protéines SSB (*Single Strain Binding*) VirE2 et porte à l'extrémité 5' une protéine VirD2 qui porte les séquences NLS (*Nuclear Localisation Signal*) d'adressage vers le noyau de la cellule végétale.

### Transfert du complexe T à la cellule végétale

Le système de transfert du complexe T est probablement dérivé d'un système conjugatif. Un pilus est créé entre la bactérie et la cellule végétale. Soit, ce pilus sert de canal et le complexe T passe dedans, soit il sert de crocher à rapprocher les deux cellules. VirD2 guide le complexe T vers le noyau de la cellule.

### Intégration du complexe T dans le génome nucléaire de la plante

L'ADN-T est intégré par recombinaison illégitime. Il semble s'intégrer au hasard dans le génome végétal. Il n'y a pas de prérequis dans la séquence végétale. Cependant, l'intégration est apparemment plus facile dans les zones de transcription active.

### Expression des gènes portés par l'ADN-T

Une fois intégré dans le génome l'ADN-T se comporte comme un locus de plus (transgène) pour la plante transmis de façon mendélienne à la descendance. Le transgène est très exprimé car il comporte les signaux de la transcription eucaryotiques. Les hormones végétales produites par les tissus tumoraux induisent la tumeur. Les opines sécrétées par les tumeurs régulent positivement les gènes de leur catabolisme portés par le plasmide Ti des bactéries. Ces opines stimulent aussi la conjugaison bactérienne et l'expression des gènes de virulence (Petit *et al.*, 1970).

### *I. A. 3. d. Contrôle de la maladie*

Une fois la plante transformée, la tumeur se développe même si la bactérie n'est plus présente. Les techniques pour contrôler la maladie sont donc surtout préventives.

La prévention de la Galle du Collet passe par un contrôle de l'état sanitaire des plantes et du sol. Le sol peut être désinfecté à la vapeur ou par solarisation (en effet *Agrobacterium* ne pousse plus au-delà de 38°C) (Raio *et al.*, 1997). Les plantes, notamment la vigne, peuvent être désinfectées par thermothérapie (traitement à l'eau chaude) (Moore and Allen, 1986).

La rotation des cultures, en alternance avec des plantes insensibles ou résistantes à *Agrobacterium*, permet aussi de limiter les épidémies (New and Kerr, 1972 ; Moore and Allen, 1986 ; Raio *et al.*, 1997).

#### **I. A. 4. Écologie des agrobactéries**

Les bactéries du genre *Agrobacterium* peuvent dans certains cas être isolées des tissus végétaux, notamment de plantes génétiquement modifiées (Mogilner *et al.*, 1993). Certaines études montrent que cette bactérie peut être aussi systémique et même induire des tumeurs aériennes (Pionnat *et al.*, 1999a ; Bouzar *et al.*, 1995). Cependant, l'environnement habituel de ces bactéries est soit les tumeurs, soit le sol, soit les rhizosphères.

##### *I. A. 4. a. Le concept d'opine*

L'écologie de la bactérie est influencée par la présence ou non d'un plasmide Ti. En fait nous sommes en présence d'une association à trois partenaires : la plante, la bactérie et le plasmide Ti. Dans ce ménage à trois le plasmide Ti force la plante à produire les opines qui vont pouvoir être utilisées par la bactérie porteuse de ce même plasmide. En présence d'une tumeur (donc d'opines) les *Agrobacterium* porteurs du plasmide Ti sont favorisés par rapport aux autres car ils exploitent une niche écologique particulière (Dessaux *et al.*, 1998). De plus, les opines favorisent le transfert conjugatif du plasmide Ti. Donc, sur les tumeurs les agrobactéries majoritaires sont celles qui portent le plasmide Ti qui a permis l'induction de la tumeur. Donc, en présence d'opine, et ce malgré le fardeau génétique du plasmide Ti (qui représente 5% du génome), les bactéries pathogènes sont favorisées. Par contre, en l'absence de tumeur et donc d'opines, les agrobactéries non-pathogènes vont être avantagées et la population pathogène va décliner (Guyon *et al.*, 1993).

##### *I. A. 4. b. Survie des agrobactéries dans le sol*

Krimi *et al.* (2002) ont étudié la survie des *Agrobacterium* pathogènes et non pathogènes dans des sols infectés après élimination des plantes.

Cette étude a montré que dans un sol en jachère où ne poussent que des plantes sans symptômes, il est possible de détecter des *Agrobacterium* pathogènes 16 ans après en avoir retiré les plantes infectées. Les *Agrobacterium* pathogènes peuvent donc persister à long terme dans les sols. Ces auteurs ont aussi observé une fluctuation saisonnière des populations d'*Agrobacterium*. Dans certains sols, la population totale des agrobactéries augmente au printemps et en été ( $\approx 10^6$  cfu.g<sup>-1</sup>). Cette augmentation de la population correspond probablement à une augmentation

de la quantité de nutriments relargués dans le sol par les plantes, mais aussi de particularités propres aux sols dits alors « conductifs ». Dans ces sols, les *Agrobacterium* pathogènes ne sont détectables qu'au printemps et en été. Lorsque la population totale revient à des niveaux moins élevés ( $\approx 10^3$  cfu.g<sup>-1</sup>), en automne et en hiver, il n'est plus possible d'isoler des *Agrobacterium* pathogènes. Ces résultats sont confirmés par des amplifications par PCR des gènes *vir* sur l'ADN total extrait du sol. Cette étude suggère que les *Agrobacterium* pathogènes sont avantagés par rapport aux non-pathogènes en période de croissance des plantes (Krimi *et al.*, 2002).

#### *I. A. 4. c. Structure des populations*

L'étude des populations d'*Agrobacterium* pathogènes ou non présentes dans les sols, les rhizosphères et les tumeurs ont amenés des précisions sur le rôle de la plante dans la structure des populations d'*Agrobacterium*.

### **POPULATION DU SOL**

L'étude de la répartition spatiale à microéchelle des agrobactéries naturellement présentes dans un sol (sol de La Côte Saint André) à montré que cette population contient trois espèces du biovar 1 et que la diversité observable est déjà présente dans un cm<sup>3</sup> (Vogel *et al.*, 2003).

#### Épidémie du peuplier

L'étude des *Agrobacterium* pathogènes présent sur les tumeurs de peuplier (*Populus tremula* X *P. alba*) dans deux localités françaises à montré que le matériel végétal pouvait être une source de contamination (Ponsonnet, 1994). Les agrobactéries pathogènes colonisant les tumeurs présentent une grande variabilité révélée par sérotypage et ribotypage (Nesme *et al.*, 1987 ; Beneddra *et al.*, 1996, Ponsonnet and Nesme, 1994). L'analyse AFLP de ces données à montré elle que la plupart de ces agrobactéries, bien que diverses, appartiennent à l'espèce G1 (D'Angelo, 2000).

#### Populations d'agrobactéries non-pathogènes dans les rhizosphères

L'analyse des populations d'*Agrobacterium* associées aux rhizosphères du lotier, de la morelle et du maïs à montré une structuration des populations non-pathogènes en fonction des plantes et des compartiments d'isolement. Des plants axéniques de lotier (*Lotus corniculatus*) et de morelle noire (*Solanum nigrum*) ont été plantés dans le sol de La Côte Saint André (Isère, France) et des grains de maïs (*Zea mais*) y ont



été semés. Puis, après incubation, les agrobactéries ont été isolées du sol rhizosphérique et des tissus racinaires. Parmi les *Agrobacterium* présents dans ce sol (au moins trois espèces du biovar1), seul des agrobactéries appartenant à une seule de ces espèces sont isolées de ces plantes. De plus les génotypes des agrobactéries associées au sol rhizosphérique et aux tissus racinaires de la morelle sont différents (Mougel, 2000).

Ces études ont montré le rôle de la plante dans la structuration des populations d'agrobactéries. Elles apportent des indications sur les pressions de sélections exercées par la plante sur ces populations. La question est de savoir si les plantes ne seraient pas à l'origine de la différenciation des populations, d'abord en écotypes puis en espèces.

## I. B. Genres, espèces et souches bactériens

### I. B. 1. La classification chez les procaryotes

Afin de reconnaître les êtres vivants, il est nécessaire de les nommer (Boyer and L'Hour, 2001). Ceci permet d'avoir une base commune de référence, indispensable pour une communication efficace. La nomenclature des organismes vivants est basée sur l'espèce, qui représente la plus petite unité taxonomique. La définition de l'espèce a évolué avec les techniques et les concepts disponibles.

Aujourd'hui, le concept d'espèces eucaryotes le plus connu, est le concept d'espèces biologiques (*Biological Species Concept* ou BSC) énoncé par Ernst Mayr en 1942. Il définit l'espèce comme une : « population d'individus interféconds, ou potentiellement interféconds, isolés sexuellement des autres populations ».

Cette définition de l'espèce biologique repose sur la notion d'isolement sexuel et permet de différencier de façon pratique et fiable les espèces animales (Claridge *et al.*, 1997). Même si ce concept d'espèce est controversé, car difficilement applicable pour les espèces parthénogénétiques, ou qui ne font pas partie du règne animal comme les algues, les lichens, les plantes ou les microorganismes (Claridge *et al.*, 1997), le BSC a permis de clarifier la classification des organismes eucaryotes à reproduction sexuée.

Chez les procaryotes la définition n'est pas aussi claire car il n'existe pas de concept universel (Claridge *et al.*, 1997). Cependant, depuis la découverte des microorganismes en 1638 par Antonie von Leeuwenhoek, les bactériologistes ont développés différentes approches afin de les classer. L'énorme diversité des procaryotes (Whitman *et al.*, 1998 ; Torsvik *et al.*, 1990) rend leur classification indispensable. Il y aurait sur terre environ  $5 \cdot 10^{30}$  cellules bactériennes sur terre. La longueur de leurs ADN mis bout à bout représenterait  $90 \cdot 10^9$  années lumières, c'est à dire plus que la taille de l'univers (Simonet et Pernodet, communication personnelle). La question est maintenant de savoir comment classer les microorganismes.

Les premières tentatives de classification des microorganismes étaient basées sur la morphologie des bactéries observables au microscope. Cette classification est basée sur le postulat que la morphologie des bactéries est indépendante des conditions du milieu. Les limites de ce critère sont très vite apparues. En effet une même bactérie

peut avoir un cycle de vie avec plusieurs formes, comme par exemple *Ramlibacter tataouinensis* (Heulin *et al.*, 2003), ou avoir une morphologie différente dans différents milieux. Par exemple, les bactéries du genre *Azospirillum* en milieu liquide portent de nombreux flagelles et sont mobiles. Par contre, en milieu solide elles perdent leurs flagelles (Alexandre *et al.*, 1999).

Avec le développement des cultures pures de microorganismes, il est devenu possible d'étudier la physiologie des bactéries, en recueillant des informations d'un seul et même organisme. De nombreux tests furent développés, notamment sous formes de « galeries », afin de déterminer leur physiologie et leur capacités métaboliques. Plus le nombre de caractères testés est grand plus la classification qui en résulte est précise (D'Amato *et al.*, 1991)

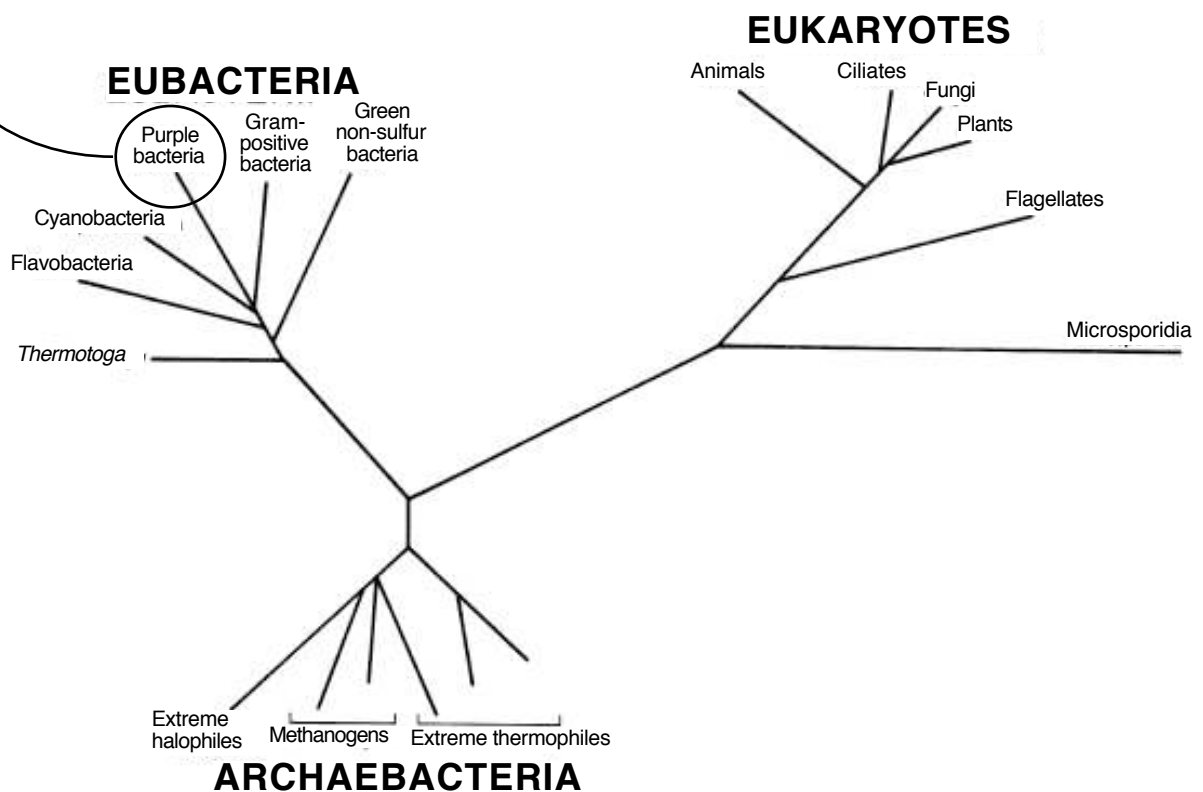
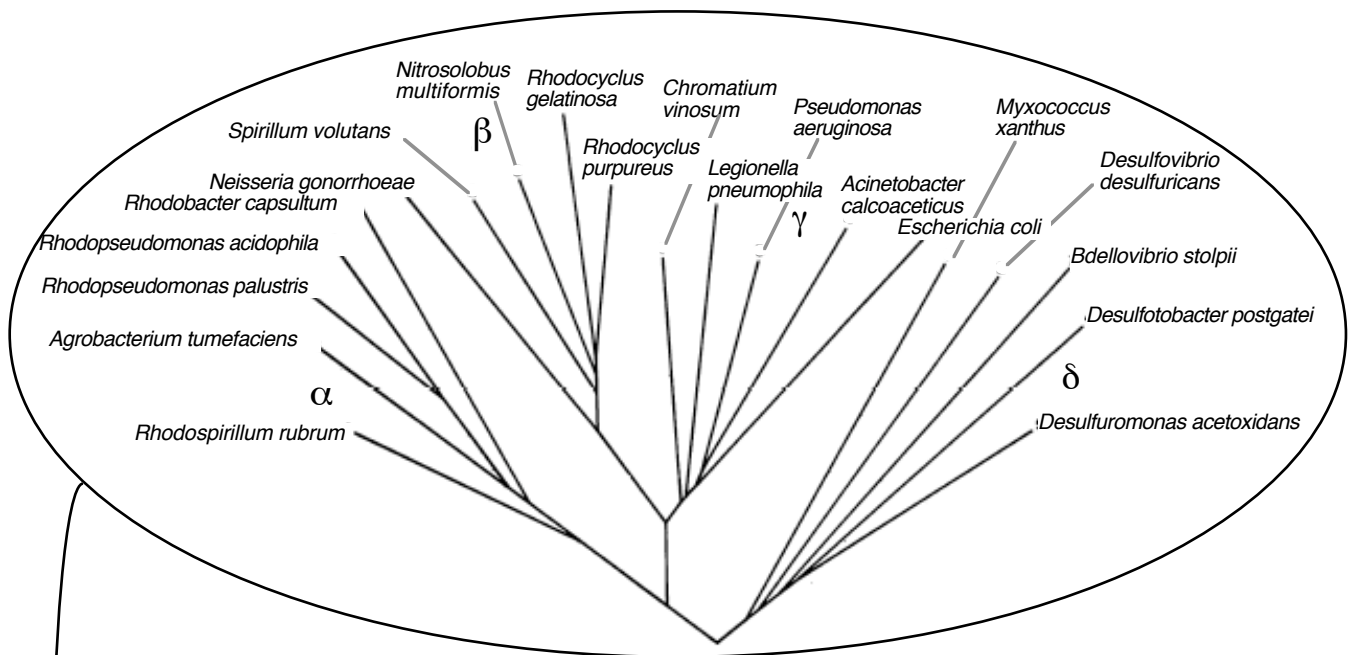
#### *I. B. 1. a. Genre*

Le genre regroupe un ensemble d'espèces apparentées. Par contre, la définition en est très floue car il n'y a pas de consensus pour définir le degré de parenté des espèces d'un genre. Cela reste une question d'opinion plus que de critères scientifiques. C'est la cause de querelles incessantes parmi les taxonomistes. Il est d'ailleurs question à l'heure actuelle d'inclure les bactéries du genre *Agrobacterium* dans le genre *Rhizobium* (Young *et al.*, 2001). Mais cette proposition ne fait pas, loin de là, l'unanimité et le rejet de la proposition par Farrand *et al.* (2003) a reçu le soutien publié de dizaines de chercheurs.

#### *I. B. 1. b. Souches*

Le terme souche bactérienne recouvre deux sens. Le premier est couramment utilisé en laboratoire. Une souche correspond à un isolat en culture pure et possède un identifiant (nom, numéro) qui lui est propre.

En épidémiologie, le terme souche, défini par Tenover (Tenover *et al.*, 1995), représente un ensemble d'isolats indistingables par la méthode de typage utilisée (ribotypage, champs pulsé, serotypage...). Ainsi plusieurs isolats peuvent appartenir à la même souche. Celle-ci est une unité taxonomique, au niveau intra-spécifique. Nous verrons, dans la suite de ce travail de thèse (Partie II) comment l'AFLP nous permet de déterminer ces souches.



Arbres phylogénétiques du vivant et des Protéobactéries (bactéries pourpres) établis à partir des séquences des ARN ribosomaux  
D'après Woese, 1987

### **I. B. 2. Le concept d'espèces évolutives**

Un des buts de la classification, outre de nommer les espèces, est de refléter l'histoire évolutive des organismes. Le séquençage du gène *rrs* permet, à la fois de classer les bactéries, par comparaison avec des bases de données de plus en plus fournies, et de déterminer la distance génétique les séparants (Woese, 1987).

Le gène *rrs* détermine l'ARN 16S de la petite sous-unité du ribosome chez les bactéries. Il est universel chez les procaryotes, et les organites des eucaryotes, et permet de comparer rapidement de nombreux isolats entre eux. Il est suffisamment variable pour apporter des informations taxonomiques et suffisamment conservé pour être facilement analysable (Woese, 1987). A partir de telles données Carl Woese a reconstitué l'arbre général du vivant (Figure 2). Il a été observé que les gènes *rrs* des souches d'une même espèce bactérienne présentent au moins 97% de similarité (Stackebrandt and Gobel, 1994). Ce type d'analyse s'est révélé très pertinent pour classer les bactéries jusqu'au niveau du genre.

Par contre, cette méthode n'est pas suffisamment résolutive pour classer les individus par espèces. En effet, le gène *rrs* étant très conservé il n'y a pas de relation linéaire entre le taux d'hybridation ADN-ADN et la similarité du gène *rrs* pour des organismes phylogénétiquement proches (Stackebrandt and Gobel, 1994).

Certaines espèces génomiques d'*Agrobacterium* partagent des gènes *rrs* identiques (Mougel *et al.*, 2002). A l'inverse, comme *Haloarcula marismortui*, certaines bactéries possèdent plusieurs gènes *rrs* qui ne sont pas identiques (Mylvaganam and Dennis, 1992). De plus, la divergence des génomes ne correspond pas toujours avec celle calculée avec la séquence du gène *rrs* (Figure 3).

Le séquençage du gène *rrs* n'est donc pas suffisant pour permettre une classification des bactéries en espèces. Par contre, vu son côté pratique et universel, il est recommandé d'inclure la séquence du gène *rrs* dans toute nouvelle définition d'espèce procaryote (Ludwig, 1999 ; Stackebrandt *et al.*, 2002).

### **I. B. 3. Le concept de cohésion des espèces**

Ce concept d'espèces bactériennes est principalement empirique. Il a été à la base des classifications numériques (Sneath and Sokal, 1962). Il repose maintenant sur l'observation que les individus d'une même espèce présentent au moins 70% d'homologie génomique (Grimont, 1981). La diversité au sein d'une espèce étant

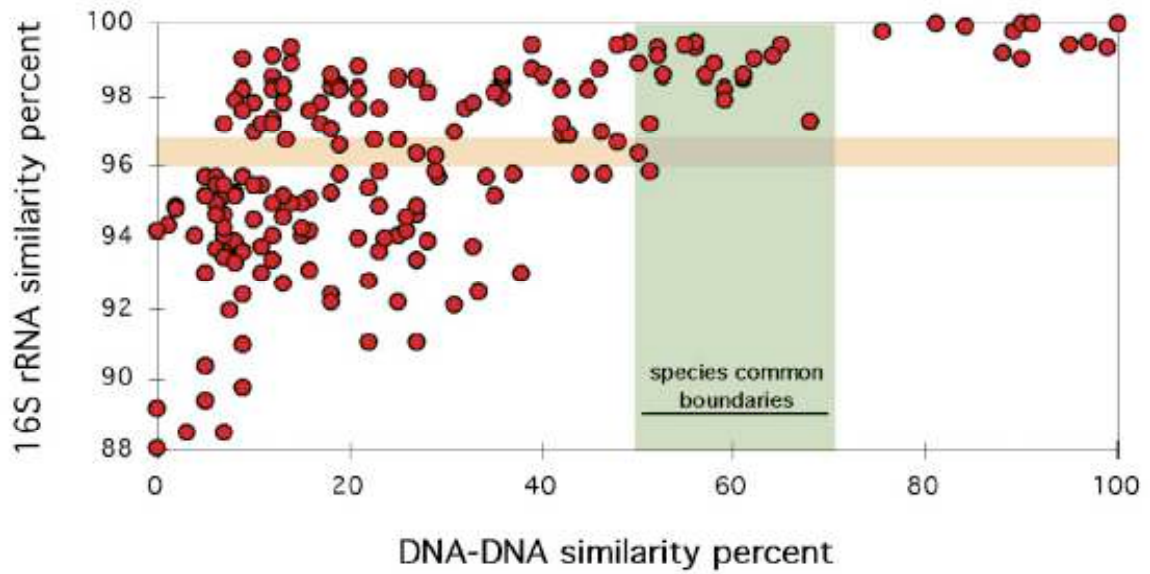


Figure 3

Comparaison entre les valeurs d'hybridation ADN-ADN et la similarité des séquences du gène *rrs*. Cette étude est basée sur 180 valeurs de 27 articles indépendants du International Journal of Systematic Bacteriology volume 49 (1999). Ces données incluent des valeurs obtenues avec des bactéries appartenant à la famille des Protéobactéries, du groupe des Cytophaga-Flavobacterium-Bacteroides, des bactéries à Gram positif et des bactérie à haut GC%.

Extrait de Rosselo-Mora et Amann, 2001

limitée, des forces de cohésions permettent à cette espèce de conserver son intégrité.

Les critères génomiques permettent donc de définir les espèces génomiques. En effet, celles-ci correspondent à des groupes d'individus génomiquement proches.

### *I. B. 3. a. Hybridations ADN-ADN*

La similarité de deux génomes peut-être mesurée par hybridation ADN-ADN. Deux paramètres sont mesurés : le RBR (*Relative Binding Ratio*) et le  $\Delta T_m$ . Le premier correspond au pourcentage d'hybridation d'un hétéroduplexe (confrontation de deux génomes, un à tester et un de référence) rapporté à l'hybridation homoduplexe du génome de référence. Le deuxième paramètre permet de déterminer la stabilité thermique de l'hétéroduplexe (Figure 4).

L'ADN de l'organisme de référence est généralement marqué par radioactivité puis dénaturé par la chaleur. L'ADN de l'organisme étudié est dénaturé par la chaleur. Les ADN sont mélangés puis renaturés. L'ADN de l'organisme étudié (non marqué) est mis en excès afin d'éviter l'autohybridation de l'ADN de l'organisme de référence. Les ADN sont refroidis afin de permettre leur hybridation.

Les ADN hybridés et non hybridés sont séparés puis la radioactivité est mesurée. La quantité de radioactivité est comparée à celle de l'expérience contrôle (l'ADN de référence est hybridé avec lui-même). La stabilité thermique des deux génomes est comparée, permettant le calcul du  $\Delta T_m$  (Figure 3). Ces deux données permettent de définir la similarité des génomes étudiés et donc leur degré de parenté.

Il a été observé que la plupart des espèces bactériennes bien définies comprennent des individus présentant des divergences génomiques supérieures ou égales à 70% (Schleifer and Stackebrandt, 1983 ; Grimont, 1988). Ces observations ont amené la réunion d'un comité *ad-hoc* pour la définition de l'espèce bactérienne (Wayne *et al.*, 1987), et à baser celle-ci sur l'hybridation ADN-ADN. Définie par ce comité, « L'espèce bactérienne inclut les souches ayant à la fois une similitude de séquences des nucléotides de l'ADN supérieure à 70% et 5°C ou moins de  $\Delta T_m$ , les deux devant être considérés car ils ne coïncident pas toujours ».

Cette définition inclut les notions de similarité génomique. Cependant, des caractères phénotypiques doivent permettre de différencier les espèces. Cette définition de l'espèce bactérienne présente l'avantage d'être pratique, opérationnelle et universellement applicable. La similarité génomique a été confirmée comme le

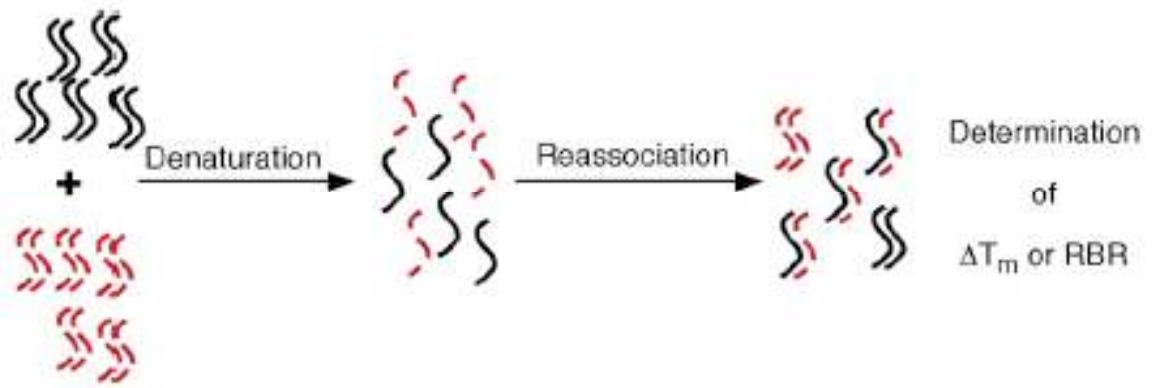


Figure 4  
Hybridation ADN-ADN et calcul du  $\Delta T_m$ .  
Extrait de Rosselo-Mora et Amann, 2001



concept de choix pour distinguer les espèces bactériennes (Stackebrandt *et al.* 2002).

Depuis la réunion du premier comité *ad-hoc* pour la définition de l'espèce bactérienne (Wayne *et al.*, 1987) les techniques ont évoluées, nécessitant la réévaluation de cette définition.

### *I. B. 3. b. Définition actuelle de l'espèce bactérienne*

En 2002, un nouveau comité *ad-hoc* s'est réuni afin de ré-évaluer la définition de l'espèce bactérienne (Stackebrandt *et al.*, 2002).

La définition énoncée par Wayne *et al* en 1987 reste valable car elle est pratique, opérationnelle et universellement applicable. De plus elle est basée sur la divergence de séquences existant entre deux génomes. Mais cette définition est essentiellement empirique, basée sur la confrontation expérimentale des génomes deux à deux. Elle ne permet pas de comparer un isolat à une base de données.

Le nouveau comité encourage donc de proposer une définition des espèces basée sur des méthodes moléculaires. A condition que dans le taxon étudié il y ait suffisamment de congruence entre la technique utilisée et l'hybridation ADN-ADN.

Le comité encourage aussi le développement de nouvelles méthodes pour compléter ou remplacer l'hybridation ADN-ADN. Ces méthodes doivent satisfaire trois critères :

La méthode doit être quantitative et les résultats doivent être analysables statistiquement.

La méthode doit être validée par l'analyse d'une collection de souches déjà analysée par hybridation ADN-ADN et autres méthodes.

Une collection de souches représentatives du taxon doit être étudiée.

Les méthodes prometteuses sont les méthodes de séquençage multiple (MLST), les méthodes qui permettent d'obtenir des profils génomiques (AFLP, ribotypage, Rep-PCR, PCR-RFLP) et les puces à ADN.

Par contre les études sur le transfert de gène ayant montré une grande fréquence de ce phénomène (Ochman *et al.*, 2000), baser la taxonomie sur l'analyse d'un seul gène, comme le gène *rrs*, n'est pas suffisant. Le séquençage de ce gène apporte des informations précieuses, notamment permet de replacer l'espèce étudiée dans un arbre plus général, mais ne suffit pas à lui seul, pour classer les microorganismes (Brown, 2003 ; Stackebrandt *et al.*, 2002).

### *I. B. 3. b. Les méthodes prometteuses pour la taxonomie des bactéries*

#### MLST

Le séquençage multiple ou MLST (*Multilocus Sequence Typing*) consiste à séquencer 5 à 7 gènes, souvent ménager, puis à analyser les séquences. Ces séquences donnent le plus souvent suffisamment d'information pour affilier ou non un isolat à une espèce connue (Maiden *et al.*, 1998 ; Urwin *et al.*, 2003).

#### Puces à ADN

Cette technique consiste à hybrider un génome complet, par exemple, à une série de sondes fixées sur un support (lame de verre le plus souvent). Ceci permet la comparaison à haut débit des individus entre eux, par rapport à une référence (les sondes fixées sur la lame) (Schena *et al.*, 1998 ; Salama *et al.*, 2000). Il s'agit cependant, d'une simple extension de l'hybridation ADN-ADN.

#### AFLP

Le polymorphisme de taille de fragments de restriction ou AFLP (*Amplified Fragment Length Polymorphism*) est à l'heure actuelle une des méthodes les plus prometteuses pour la taxonomie. Il s'agit, après une digestion du génome complet par une ou deux enzymes de restriction différentes, d'amplifier de façon sélective une partie des fragments obtenus (Vos *et al.*, 1995 ; Blears *et al.*, 1998). L'avantage de cette technique est qu'elle permet de calculer les divergences génomiques (Mougel *et al.*, 2002). Il est donc possible, à partir de ces résultats, d'affilier ou non un isolat à une espèce connue. Il est possible ainsi de détecter des espèces potentiellement nouvelles. Cette méthode permet les analyses à haut débits indispensables en analyse de population.

À l'heure actuelle, en écologie microbienne, cette technique est surtout utilisée à des fins épidémiologiques, pour suivre la dissémination de bactéries pathogènes et estimer la diversité des populations (Guan *et al.*, 2002 ; Geornaras *et al.*, 1999 ; Pirnay *et al.*, 2003 ; Owen *et al.*, 2001 ; Jiang *et al.*, 2000 ; Bojesen *et al.*, 2003 ; Arnold *et al.*, 1999). D'autres applications complémentaires commencent à se développer. Notamment pour préciser la taxonomie de certains genres et voir les relations phylogénétiques entre membres d'un taxon (Duim *et al.*, 2001 ; Mougel *et al.*, 2001 ; Vancanneyt *et al.*, 2002 ; Koeleman *et al.*, 1998 ; Clerc *et al.*, 1998). Il est même possible de suivre par AFLP l'évolution de populations (Kuipers *et al.*, 2000 ; Kuhn *et al.*, 1997 ; Jarraud *et al.*, 2002).

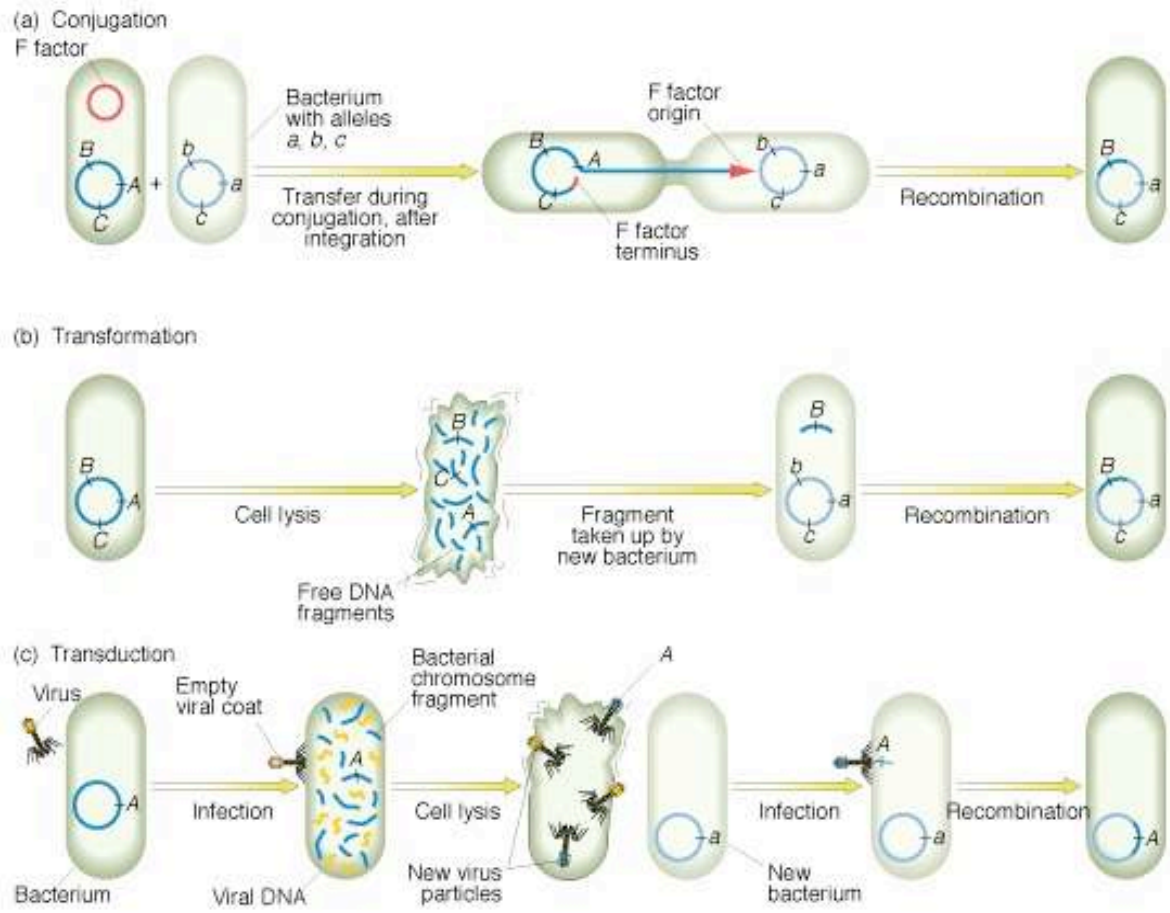


Figure 5

Les trois modes de transfert horizontal de gène : transduction, recombinaison et transformation.

Le séquençage complet de différents génomes a permis le développement de l'AFLP prédictive, par simulation *in silico* des digestions et amplifications sélectives, ce qui permet de comparer les données attendues et expérimentales (Arnold *et al.*, 1999 ; Mougél *et al.*, 2002). Les analyses AFLP prédictives et expérimentales de *Escherichia coli* montrent la fiabilité de l'AFLP pour l'analyse génomique.

L'analyse, par Mougél *et al.* en 2002, du genre *Agrobacterium* par AFLP, montre qu'il est possible de faire correspondre les valeurs d'hybridation ADN-ADN aux données AFLP. En effet, l'AFLP permet de calculer que les individus d'une même espèce présente une distance génomique évolutive de 10,5%. Ceci correspond à un taux de mésappariements moyen entre génomes de 13%.

La possibilité de calculer les distances génomiques évolutives, permet de substituer cette méthode à l'hybridation ADN-ADN et de joindre le concept d'espèce génomique et d'espèce évolutive.

Le fait qu'à l'intérieur d'une même espèce, le taux de mésappariement entre génomes soit limité (13%), suggère qu'à l'intérieur des espèces les gènes sont beaucoup plus facilement échangeables, par recombinaison homologue, qu'entre espèces. La distance (13-14%) entre espèces génomiques pourrait éventuellement contribuer à l'isolement sexuel des espèces bactériennes. Ainsi, le concept d'espèces biologiques (Mayr, 1942) pourrait éventuellement s'appliquer aux bactéries.

#### **I. B. 4. Concilier les définitions d'espèces eucaryotes et procaryotes ?**

##### *I. B. 4. a. Le sexe chez les bactéries*

A priori, la reproduction des procaryotes n'étant pas sexuée, le concept d'espèce biologique paraît donc difficilement applicable aux bactéries. Malgré tout, les bactéries ont une sexualité à la fois permissive (*promiscuous*) et partielle (seule une portion du génome est concernée). Le sexe étant défini ici comme tout échange de matériel génétique entre deux bactéries. Le fait est que les populations de bactéries ne sont pas entièrement clonales (Maynard-Smith *et al.*, 1993). Certaines, comme *Helicobacter pylori*, étant résolument panmixtiques (Israël *et al.*, 2001 ; Falush *et al.*, 2003).

Les bactéries peuvent échanger du matériel génétique selon trois mécanismes : la transduction, la conjugaison et la transformation naturelle (Figure 5) (Majewski,

2001). Ces trois mécanismes permettent à la cellule receveuse d'intégrer de l'ADN étranger dans son génome.

La transduction est l'échange de gène via les bactériophages. Il s'agit du résultat d'une « erreur d'emballage » du génome du phage dans la capsid. De l'ADN bactérien est « embarqué » avec l'ADN viral, qui sera « relargué » dans le nouvel hôte du virus. La conjugaison passe par les plasmides ou les transposons. Après formation d'un pilus qui résulte en contact physique entre la cellule donatrice et la réceptrice, de l'ADN chromosomique de la cellule donatrice peut être co-mobilisé par l'élément conjugatif et transféré dans la cellule réceptrice. La transformation naturelle est un processus par lequel les bactéries intègrent de l'ADN libre présent dans l'environnement. Tous ces mécanismes activent les processus de recombinaison de la cellule receveuse. L'ADN exogène est intégré au génome receveur par recombinaison. La recombinaison est activée par les trois mécanismes cités ci-dessus (Majewski, 2001).

Les échanges horizontaux de gènes sont très communs chez les procaryotes et représentent un mécanisme majeur d'évolution chez les bactéries (Ochman *et al.*, 2000). Ces échanges constituent un mécanisme très performant d'acquisition de nouveaux allèles. C'est même une source majeure de biodiversité chez les procaryotes. Ce mécanisme permet aux bactéries de s'adapter très rapidement à des changements environnementaux qui pourraient sinon leur être fatal (Levin and Bergstrom, 2000).

#### *I. B. 4. b. mécanismes de régulation des transferts de gènes*

Si l'acquisition de matériel génétique exogène peut être bénéfique à la cellule, receveuse, l'intégration de tel matériel peut aussi être délétère (Majewski, 2001). Ainsi la cellule possède des mécanismes qui limitent ces échanges génétiques et ainsi équilibrent les risques et les gains de l'acquisition d'ADN exogène.

La principale barrière à l'échange génétique entre bactéries est le système de réparation des mésappariements de bases (SRM) (Majewski *et al.*, 2000 ; Majewski, 2001 ; Rayssiguier *et al.*, 1989).

La recombinaison est induite par la reconnaissance, par la protéine RecA, d'une séquence minimale parfaitement homologue entre l'ADN receveur et exogène (Majewski and Cohan, 1999 ; Shen and Huang, 1986). Plus les séquences sont éloignées l'une de l'autre, moins la protéine RecA pourra identifier les zones

homologues nécessaires à la recombinaison. De fait, il a été observé que la fréquence de recombinaison homologue décroît quand la divergence de séquence entre les ADN donneur et receveur augmente (Roberts and Cohan, 1993 ; Vulic *et al.*, 1997).

Cette diminution de la fréquence de la recombinaison en fonction de la distance génétique est la conséquence de l'action du système de réparation des mésappariements de base (ou SRM) (Majewski *et al.*, 2000 ; Majewski, 2001). Le système SRM agit via les protéines MutS et MutL. Il reconnaît les séquences qui ne s'hybrident pas parfaitement et met fin à la recombinaison. Plus les séquences sont éloignées moins la protéine RecA pourra agir vite. Ainsi l'ADN exogène est exposé à l'action du système SRM. Ce système permet à la cellule d'éviter l'intégration d'ADN exogène trop différent (Majewski, 2001).

#### *I. B. 4. c. Barrière génétique et isolement sexuel*

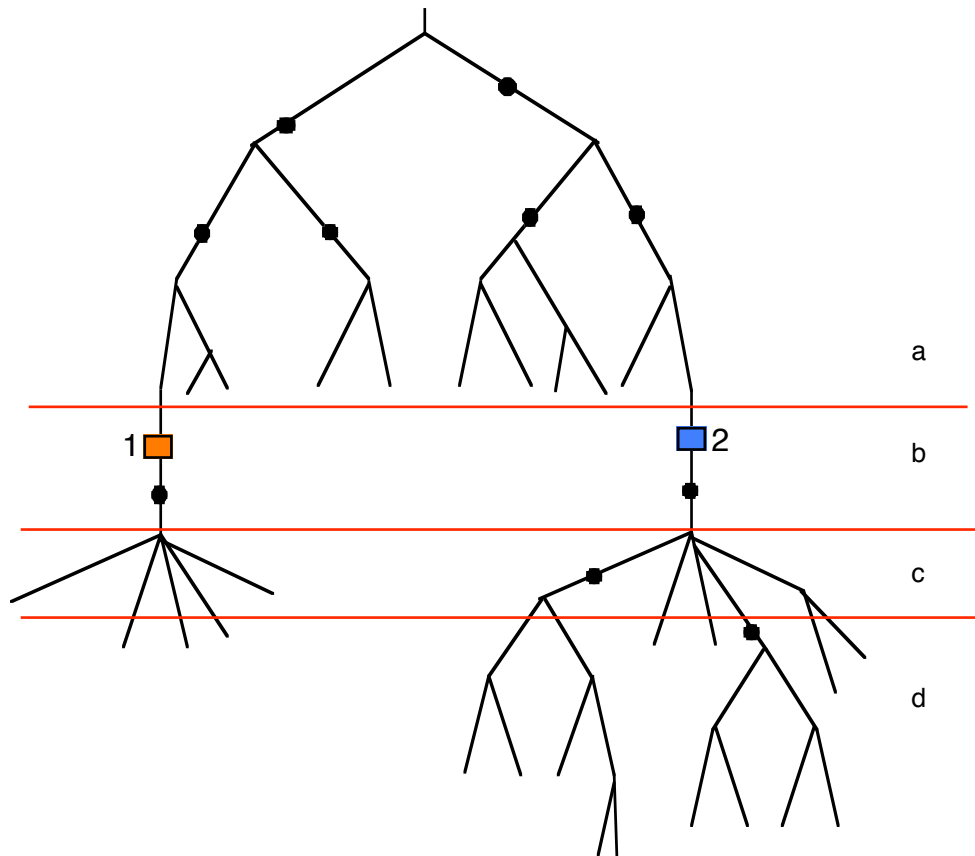
La conséquence de l'action du système SRM est l'établissement d'une barrière génétique (Majewski *et al.*, 2000).

Par exemple entre *Escherichia coli* et *Salmonella typhimurium* on observe une réduction de la fréquence de recombinaison entre ces deux génomes d'un facteur  $10^5$  qui présentent environ 16% de divergence. Par contre la recombinaison intra-génomique est très peu influencée par ces deux systèmes (Vulic *et al.*, 1997).

Ainsi les échanges génétiques, via la recombinaison homologue, sont de plus en plus rares à mesure que la distance génétique entre les organismes augmente.

Il existe donc chez les procaryotes des mécanismes d'isolement sexuels (Majewski *et al.*, 2001). Ceci permet de rapprocher la définition de l'espèce eucaryote, basée sur l'isolement sexuel, à celle de l'espèce bactérienne.

Par contre, si chez les eucaryotes l'isolement sexuel est une force qui permet la cohésion des espèces, chez les procaryotes la sexualité est beaucoup plus permissive, permettant ponctuellement l'échange de matériel génétique entre organismes phylogénétiquement très éloignés. En fait, l'isolement sexuel apparaît, chez les procaryotes, plus une conséquence de la divergence des espèces qu'une des causes de cette spéciation. Néanmoins, cette barrière existant de fait, les échanges génétiques par recombinaison ont plus de chance de se produire au sein de l'espèce, maintenant son intégrité génomique.



- Valeur de bootstrap significative
- Individu 1
- Individu 2

a: Les bactéries occupant une même niche écologique forment un clade, soutenu par des valeurs de bootstraps

b: L'individu 1 acquiert une mutation adaptative, il supprime tous les autres individus du clade. Sauf l'individu 2 qui a acquis une autre mutation qui lui permet d'occuper une autre niche écologique et donc d'échapper à la sélection périodique

c: Les individus 1 et 2 fondent un nouveau clade. À partir de ces individus fondateurs les descendants divergent de façon radiale.

d: Les nouveaux clades continuent d'évoluer. De nouvelles subdivisions se forment à l'intérieur de ces clades, certaines pouvant être même soutenues par des bootstraps.

Figure 6.  
Évolution des écotypes au cours du temps.  
D'après Cohan, 2001.

## **I. B. 5. Concept d'espèces écologiques**

### *I. B. a. Ecotypes*

La notion d'écotypes proposée par Cohan (Cohan, 2001, 2002) reprend le concept d'espèces écologiques (Van Valen, 1976) et le développe.

Selon Cohan, les écotypes sont des groupes d'individus occupant la même niche écologique et susceptibles d'être l'objet de sélection périodique. Les individus d'un écotype différent ne sont pas sensibles aux mêmes événements de sélection périodique (Cohan, 2001).

La sélection périodique a pour conséquences d'augmenter la divergence entre écotypes, et de limiter la diversité intra-écotype. Ce mécanisme, qui purge régulièrement la diversité de l'écotype, est une force de cohésion qui permet à un écotype de conserver son identité génétique. Pour Cohan, les espèces génomiques sont résolument des écotypes dont la cohésion résulte de purge périodique de la diversité.

Avec le temps la diversité inter-écotype est plus importante que la diversité intra-écotype et les écotypes forment des entités discrètes (Figure 6). A terme, les différents écotypes sont suffisamment divergents pour qu'il s'établissent entre eux une barrière génétique. Ainsi, ce concept d'écotype permet à la fois d'appréhender les causes de la spéciation (les conditions écologiques), les forces de cohésions garantant de l'intégrité de l'espèce (la sélection périodique) et les conséquences de la spéciation bactérienne (l'isolement sexuel).

### *I. B. 5. b. Caractérisation des écotypes*

Les écotypes comprennent des individus issus d'un même individu fondateur, et forment des groupes monophylétiques d'individus génétiquement proches (Cohan 2001 ; Cohan, 2002). Le séquençage d'un seul gène peut déjà révéler les écotypes car des séquences différentes d'un même gène peuvent être associées à des groupes écologiques distincts (Palys *et al.*, 1997). Ainsi chacun des clades regroupant des séquences quasi similaires pourra correspondre à un écotype (Cohan, 2001). Les écotypes pourront aussi être caractérisés par des clades en formes d'étoile.

Une technique prometteuse pour la caractérisation des écotypes est le MLST. Cette technique, par le séquençage simultané de plusieurs gènes permet de caractériser plus précisément les écotypes. Les souches analysées sont regroupées en



complexes clonaux. Un complexe clonal est formé des individus identiques, à un individu de référence, pour au moins cinq des locus analysés. Les complexes clonaux définis par MLST correspondent remarquablement bien à des clades écologiquement distincts (Maiden *et al.*, 1998 ; Chan *et al.*, 2001 ; Zhu *et al.*, 2001). De plus, l'étude des gènes qui présentent des différences significatives entre écotypes permettraient de déterminer les niches écologiques auxquelles sont associées ces écotypes et ainsi permettrait de comprendre les causes écologiques en cause dans la spéciation bactérienne.

#### *I. B. 5. c. Espèces et écotypes*

Les analyses MLST des espèces *Neisseria Meningitidis*, *Staphylococcus aureus* et *S. pneumoniae* révèlent que ces espèces contiennent de nombreux complexes clonaux assimilables à des écotypes (Feil *et al.*, 1999 ; Cohan, 2002). Cohan propose d'identifier les écotypes aux espèces bactériennes. La notion d'écotype permet de réconcilier approches techniques de la définition de l'espèce bactérienne et causes biologiques de cette spéciation.

Mais certaines questions restent ouvertes, notamment de savoir si il existe plusieurs niveaux taxonomiques au sein d'une espèce et donc d'un écotype (Cohan, 2001, 2002).

## Conclusion et hypothèses de travail

La classification des bactéries, d'abord intuitive, empirique, est de plus en plus objective. La distance génétique, mesurée d'abord par les hybridations ADN-ADN puis maintenant par des méthodes alternatives, est le fondement biologique de cette classification. L'adaptation à des niches écologiques très précises, sous les pressions de sélection, ont favorisé l'apparition de cette divergence génétique entre bactéries. Il est probable qu'à partir d'écotypes adaptés à des niches écologiques particulières, le temps et la dérive génétique aient débouchés sur l'apparition de l'énorme divergence génétique observable aujourd'hui chez les bactéries.

Parmi elles les espèces du biovar1 du genre *Agrobacterium* sont très proches phylogénétiquement. De plus, les plantes jouent un rôle prépondérant (*cf* Partie II) dans la structuration génomique de ces populations. La différence écologique entre ces espèces n'est pas liée au sol puisque différentes espèces de ce biovar1 peuvent y cohabiter étroitement. Il est possible que ces espèces du biovar1 correspondent plutôt à des écotypes liés à des plantes particulières qui auraient ensuite divergé suffisamment pour former des espèces différentes.

L'étude, par AFLP (méthode génomique à haut débit) des populations d'*Agrobacterium* associées aux plantes, permet de définir et de préciser les associations entre espèces d'*Agrobacterium* et espèces végétales. Ceci permet de définir l'influence de facteurs comme le type de plante, la présence ou non d'un plasmide Ti dans la bactérie ou encore la localisation géographique des plantes.

Par cette approche, nous cherchons à savoir si l'association entre plante et espèce bactérienne observée chez le peuplier se vérifie chez d'autres espèces végétales, et si le statut pathogène de la bactérie influe sur cette association.

Une autre voie pour déterminer les facteurs impliqués dans la spéciation dans le genre *Agrobacterium* est de définir les gènes impliqués dans la différenciation de ces espèces. A partir de la séquence complète du génome de la souche modèle C58, de l'analyse par AFLP des souches de référence de ce genre, d'outils bioinformatiques et statistiques, il est possible de déterminer les gènes impliqués dans la différenciation des espèces d'agrobactéries. Il est probable que les gènes impliqués dans la différenciation des espèces pointent des fonctions écologiques qui seraient à l'origine de la différenciation en espèces des agrobactéries.

Ces études apportent des éléments nouveaux sur les facteurs influençant la structuration des populations bactériennes potentiellement pathogènes dans les sols. Ceci pourrait à terme aider à mieux comprendre l'épidémiologie de cette bactérie et aider au contrôle de la Galle du Collet.

Ces études permettent aussi d'éclairer les mécanismes biologiques qui conduisent à la spéciation des espèces bactériennes et les causes qui en sont à l'origine. Ceci permet de mieux comprendre la dynamique de l'évolution des populations et des espèces bactériennes.

Dans ce travail de thèse nous allons étudier les populations du genre *Agrobacterium*, notamment celles associées aux plantes, sous l'angle des écotypes, afin de déterminer les liens entre écotypes et espèces bactériennes.

## Partie II

### Analyse des populations d'agrobactéries associées aux plantes

## II. Analyse des populations d'agrobactéries associées aux plantes

II. 1. Analyse des populations d'*Agrobacterium* pathogènes isolées de tumeurs de peuplier : intérêt de l'approche génomique.

### II. 1. 1 Introduction

Les agrobactéries impliquées dans une épidémie de Galle du Collet peuvent être très diverses. Au cours de sa thèse, Cécile Ponsonnet (Ponsonnet, 1994), a étudié par sérotypage et ribotypage les agrobactéries pathogènes isolées de tumeurs de peuplier (*Populus tremula* X *P. alba*). Cette étude a révélée la grande diversité des agrobactéries dans une même épidémie. Cependant, le sérotypage reste une méthode limitée car si elle permet de classer convenablement de nombreux isolats, certains ne réagissent avec aucun des sérums utilisés ; ceci nécessite la fabrication de nouveaux sérums pour reconnaître chaque nouvelle souche. Au contraire, La PCR-RFLP sur le gène *rrs* (codant l'ARNr 16S) plus l'ITS (*Intergenic Spacer*) entre 16S et 23S, est une alternative qui a permis à C. Ponsonnet (Ponsonnet *et al.*, 1994) de classer la totalité des isolats. Lorsque deux isolats présentent le même profil RFLP, ils sont considérés comme appartenant à la même souche (Tenover *et al.*, 1995), voire à un même clone. Cette méthode de caractérisation des isolats est couramment utilisée en épidémiologie de la Galle du Collet (Pionnat *et al.*, 1999a ; Nesme, 2001). Cette méthode a également été utilisée par Vogel *et al.* (2003). lors d'une étude de la répartition spatiale de la diversité des agrobactéries. Néanmoins, cette technique cible un seul locus et reste limitée au typage infra-spécifique. Elle ne permet pas, en particulier, de distinguer sans ambiguïté les espèces génomiques car celles-ci présentent fréquemment des séquences du gène 16S identiques (Mougel *et al.*, 2002). En outre, si ces données permettent de calculer des dendrogrammes, la validité des embranchements ainsi révélés ne peut pas être testée statistiquement. Dans le cas du peuplier, les travaux précédents ne permettent pas de définir sans ambiguïté les espèces en présence.

Population	Lieux	Age plants	Type de plants	Nb tumeurs
ORL2A	Orléans	2 ans	Bouture	26
PEY2A	Peyrat	2 ans	Bouture	34
PEYJT	Peyrat	1 an	Boutures hors sol	13
PEYVP1	Peyrat	1 an	Culture <i>In Vitro</i>	5
PEYVP2	Peyrat	2 ans	Culture <i>In Vitro</i>	20

Tableau 2.

Caractéristiques des plants de peuplier portant les tumeurs dont les agrobactéries ont été isolées

Espèce	Nom	Origine biologique et géographique	Référence
G1	TT111 S56	Galle, USA Plante	
G2	CIP127-76		
G3	CFBP 6624	Prélèvement humain	Keridjian, 1979
G4	B6 ATCC19358	Pommier	Kerstens <i>et al</i> , 1973 Skerman <i>et al</i> , 1980
G6	NCPPB 925 F/1 Zutra	<i>Dahlia</i> sp., Afrique du sud <i>Dahlia</i> sp., Israël	Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973
G7	CIP RV3 Zutra 3/1	<i>Malus</i> sp., Israël	Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973
G8	C58 T37	<i>Prunus</i> sp cv Montmorency (cerise), New York USA <i>Juglans</i> sp. , galle, Californie USA	Kerstens <i>et al</i> , 1973
G9	O363	Potting soil, Australie	Kerstens <i>et al</i> , 1973
<i>A. larrymoorei</i>	AF3.44 Fb33	<i>Ficus benjamina</i> <i>Ficus benjamina</i> , Salerno, Italie	Bouzar <i>et al.</i> , 1995
<i>A. rubi</i>	LMG 17935 SCRI551	<i>Rubus ursinus</i> var <i>Loganobaccus</i> , USA	
<i>A. vitis</i>	CFBP 2736 CFBP 2678	<i>Vitis vinifera</i> <i>Vitis vinifera</i>	
<i>Rhizobium rhizogenes</i>	CFBP 2408	Pommier	

Tableau 3.

Liste des souches de références analysées avec les isolats issus de tumeurs de peuplier.

C'est pourquoi, l'analyse génomique, par AFLP, d'une sélection d'isolats représentative de la diversité rencontrée dans ces tumeurs à été effectuée (D'Angelo, 2000).

Dans ce chapitre, nous avons repris ces données afin de déterminer les espèces génomiques en cause et les correspondances entre les analyses monocus (16S plus ITS) et les analyses génomiques (AFLP). En particulier, nous avons voulu déterminer la diversité génomique de groupe d'isolats présentant le même ribotype ou le même sérotype.

## **II. 1. 2. Matériel et méthodes**

### *II. 1. 2. a. Modèle végétal*

Les peupliers étudiés sont des hybrides *Populus tremula* X *P. alba* de la section *Leuce*. Les plants cultivés à Peyrat-le-Château sont soit issus de la pépinière d'Orléans où ils ont été multipliés par bouture de racines (Ponsonnet, 1994), soit issus de culture *in vitro*, acclimatés hors sol et plantés dans le sol contaminé de la pépinière de Peyrat-le-Château.

### *II. 1. 2. b. Provenances des populations bactériennes analysées*

Les populations analysées sont décrites par C. Ponsonnet (1994) (Tableau 2). Au total les agrobactéries analysés proviennent de 98 tumeurs.

Population isolée à Orléans (pépinière amont).

ORL2A : ces agrobactéries ont été isolées en 1985 de 26 tumeurs sur plants de peuplier âgés de 2 ans

Populations isolées à Peyrat-le-Château (pépinière aval).

PEY2A : ces agrobactéries ont été isolées en 1987 de 34 tumeurs de plants de peuplier âgés de 2 ans.

PEYJT : ces agrobactéries ont été isolées de treize tumeurs issues de boutures de racines encore cultivées sous serre dans un substrat fait à partir d'écorce d'arbres, exempt d'agrobactéries pathogènes. Ces plants ont été obtenus par bouture de racines de plants porteurs sains et n'ont pas été en contact avec le sol de la pépinière (essai de mise en évidence de transmission par la plante).

PEYVP1 et PEYVP2 : ces agrobactéries ont été isolées de tumeurs de plants, issus de cultures *in vitro*, et cultivés dans la parcelle ayant hébergé l'essai PEY2A six mois

Espèces	Populations					Total
	ORL2A	PEY2A	PEYJT	PEYVP1	PEYVP2	
G1	14	15	13	0	10	52
<i>Rhizobium rhizogenes</i>	6	5	0	1	2	14
G1+ <i>R. rhizogenes</i>	6	14	0	3	8	31
G1+ G4 + <i>R. rhizogenes</i>	0	0	0	1	0	1
Total	26	34	13	5	20	98

Tableau 4.

Nombre de tumeurs analysées, par populations, et classées en fonction des espèces d'agrobactéries qu'elles hébergent



auparavant (essai de piégeage des agrobactéries pathogènes du sol). La population PEYVP1 est issue de cinq tumeurs récoltées un an après la plantation, et la population PEYVP2 est issue de 20 tumeurs récoltées deux ans après la plantation. SOL : les agrobactéries du sol nu d'une parcelle adjacente, non contaminée, de la pépinière de Peyrat-le-Château ont été isolées et analysées.

### *II. 1. 2. c. Souches bactériennes et analyses AFLP.*

Les agrobactéries présentes dans les tumeurs de peuplier ont été isolées, caractérisées et typées par C. Ponsonnet (1994). Dans la suite de cet exposé, le terme agrobactéries recouvrira les espèces du genre *Agrobacterium* (biovar 1) ainsi que l'espèce *Rhizobium rhizogenes* (ex-biovar 2).

Une sélection d'isolats, représentatifs de la diversité des agrobactéries présente dans les tumeurs de peuplier, ont été analysés par AFLP (D'Angelo, 2000). Les analyses AFLP ont été réalisées sur séquenceur à gel Perkin Elmer 373A. Ces isolats ainsi que leurs caractéristiques sérologiques et ribotypiques, sont indiqués dans le tableau 5. Un ensemble de souches de référence (Tableau 3) du genre *Agrobacterium* représentatives de la plupart des espèces du genre a été analysé (Mougel *et al.*, 2002). Les profils obtenus permettent de déterminer les proximités entre nouveaux isolats et souches de référence.

## **II. 1. 3. Résultats**

### *II. 1. 3. a. Identification des espèces génomiques*

L'analyse AFLP de 28 isolats représentatifs de la diversité révélée par sérotypage et ribotypage, a permis de déterminer à quelles espèces génomiques correspondent les différents ribotypes.

Parmi les 98 des tumeurs analysées, 52 (53%) hébergent uniquement des isolats de l'espèce G1, 14 (14%) hébergent uniquement des isolats de *R. rhizogenes* et 31 (32%) hébergent simultanément des isolats de ces deux espèces. Une seule tumeur (1%), en plus d'isolats de l'espèce G1 et de *R. rhizogenes*, héberge quelques isolats qui appartiennent à l'espèce G4 (Tableau 4).

Nom	Serotype	Ribotype	Type plasmidique	Population d'origine
85-2	1a	2177	2177	ORL2A
85-6	1b	85-3	2516	ORL2A
85-30	nd	85-30	2516	ORL2A
85-49	1a	2177	2177	ORL2A
85-52	1c	2517	2516	ORL2A
85-66	1c	2517	2516	ORL2A
85-104	1c	85-104	2516	ORL2A
85-120	nd	85-110	2516	ORL2A
85-123	nd	85-30	2516	ORL2A
85-186	nd	2520	2516	ORL2A
A134-3	nd	85-46	1903	PEYVP1
A134-6	nd	A134-5	1903	PEYVP1
A68-19	nd	A68-20	M80	PEYVP1
A96-11	1b	2516	M80	PEYVP2
C104-22	nd	C104-12	2516	PEYVP2
M2	1c	2516	1903	PEY2A
M3	nd	85-30	1903	PEY2A
M10	1b	M9	2516	PEY2A
M111	nd	85-30	M80	PEY2A
M143	nd	2177	292	PEY2A
M214	1b	85-3	1903	PEY2A
M292	1b	M292	M80	PEY2A
M32	nd	85-30	M80	PEY2A
M84	nd	85-30	1903	PEY2A
X88-283	1b	85-46	2516	PEYJT
X88-289	1b	85-46	2516	PEYJT
X88-293	1a	2177	292	PEYJT
X88-303	1b	2516	1903	PEYJT

Tableau 5.

Caractéristiques sérologiques, ribotypiques et plasmidiques des isolats issus des tumeurs de peuplier analysés par AFLP.

### II. 1. 3. b. Diversité intra-spécifique

La diversité intra-spécifique révélée par AFLP à été étudiée en relation avec les analyses de sérotypage et de ribotypage.

#### Espèce G1

Certains isolats de peuplier se répartissent en clades soutenus par des valeurs de bootstrap significatives (Figure 7). Ces clades regroupent des isolats de même sérotype et appartenant au même ribotype (Tableau 6).

Les isolats qui ne font pas partie d'un clade sont chacun de ribotype spécifique (Tableau 6). Il y a donc concordance exacte entre clade statistiquement révélé par AFLP et ribotypes.

L'analyse PCR-RFLP à permis de montrer que les isolats de biovar1 du sol de Peyrat-le-Château appartiennent tous à l'espèce G1 mais de ribotypes for différents de ceux des isolats pathogènes (résultats non montrés).

#### *Rhizobium rhizogenes*

Les isolats de cette espèce, ainsi que la souche de référence CFBP 2408, se regroupent un clade soutenu à 100% (Figure 7). À l'intérieur de cette espèce un clade se différencie à 85%. Il regroupe au sein de cette espèce, les deux isolats qui appartiennent au même ribotype (Figure 7, Tableau 6).

De même dans cette espèce il y a concordance entre ribotype et clade AFLP.

#### Espèce G4

L'analyse des profils AFLP Eco+CA/Mse+0, Eco+CC/Mse+0 et Eco+CT/Mse+0, d'un isolat de ribotype A134-5 indique qu'il appartient à l'espèce G4 (Figure 7). Douze isolats, présentant le même ribotype, ont tous été isolés de la même tumeur, qui s'est développée sur un plant issu de culture *in vitro* (population PEYVP1).

### II. 1. 4. Discussion

Dans cette partie du travail nous avons voulu évaluer la capacité de l'AFLP à décrire la structure génétique d'une population d'agrobactéries pathogènes préalablement décrite par des analyses monocus (PCR-RFLP du gène 16S et de l'ITS, sérologie).

À partir d'un échantillon des isolats étudiés par C. Ponsonnet (1994), nous avons pu établir, que la grande majorité des agrobactéries isolées de tumeur de peuplier appartenaient soit à l'espèce génomique G1 soit à *Rhizobium rhizogenes*. La PCR-RFLP sur le gène 16S plus l'ITS, effectuée par C. Ponsonnet, permet d'établir en

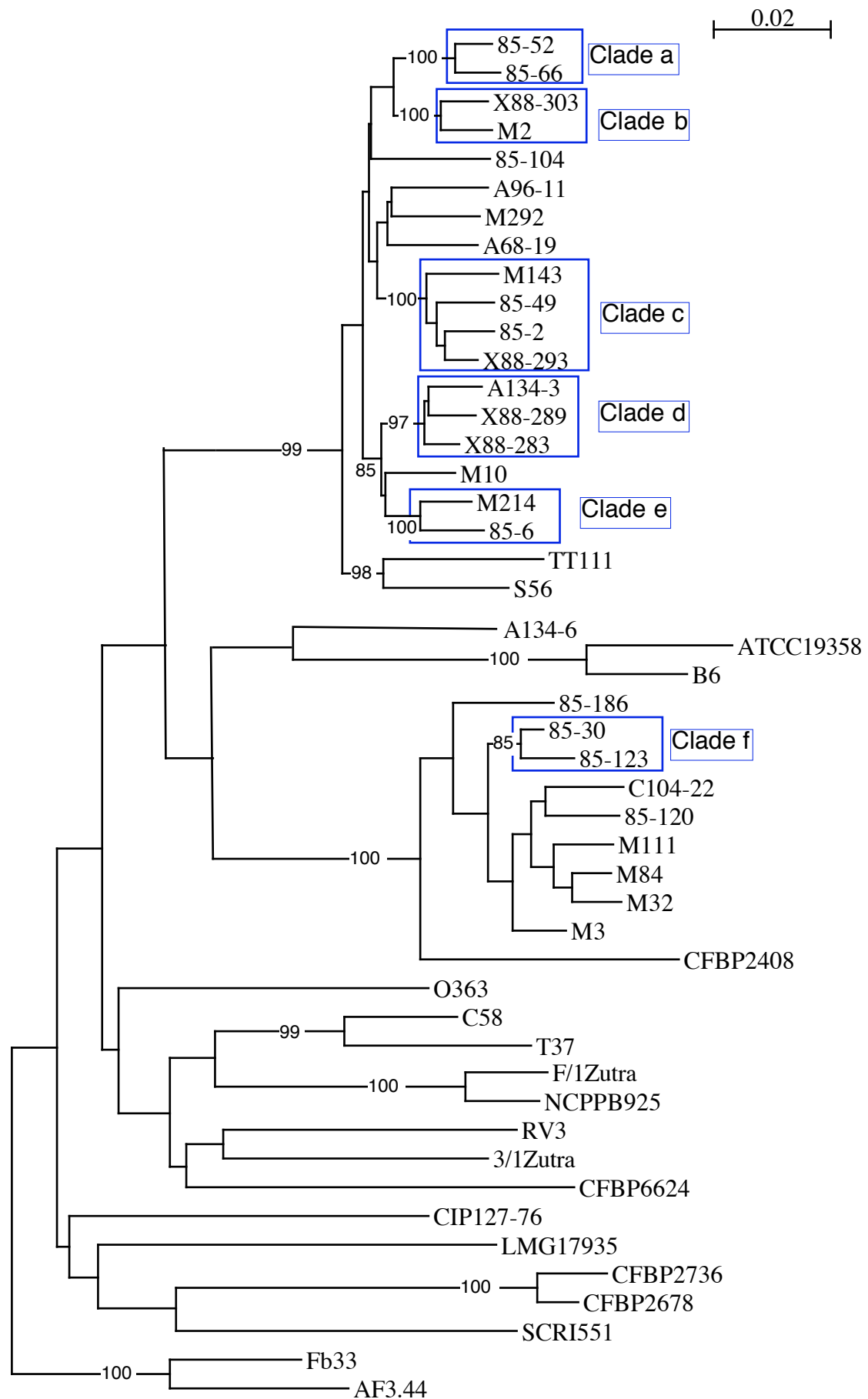


Figure 7: Arbre phylogénomique de 28 isolats, représentatifs de la diversité des agrobactéries pathogènes isolées de tumeurs de peuplier, et des souches de référence, calculé à partir de quatre conditions d'AFLP différentes. Les valeurs significatives de bootstraps, pour 100 rééchantillonnages, sont indiquées.

général l'espèce « 16S » de l'isolat (Vogel *et al.*, 2003 ; Ponsonnet and Nesme, 1994) mais pas l'espèce génomique. En effet, cette méthode est limitée par le pouvoir résolutif du gène 16S lui-même. Dans le genre *Agrobacterium*, les espèces G6 et G8, ou G2 et G9 ou encore G5 et G7 (Partie II, Chapitre 2 ; Mougel *et al.*, 2002) partagent le même gène 16S. Ici, l'AFLP nous indique sans ambiguïté l'espèce génomique.

Au niveau intra spécifique cette étude apporte des informations originales. En effet, les clades significatifs les plus fins, correspondent toujours à des ribotypes ou des sérotypes bien distincts. En reprenant la définition de Tenover (1995) sur les souches dans un contexte épidémiologique, les ribotypes déterminés par PCR-RFLP sur le gène 16S et l'ITS correspondent bien à des souches. C'est à dire à des isolats indistinguables par la méthode utilisée et dont on assume, le plus souvent, qu'ils appartiennent au même clone. Ces souches correspondent à des profils AFLP très similaires, significativement regroupés en clades, que nous nommerons clades-souches. La diversité des profils AFLP au niveau de ces clades-souches peut découler, sans que l'on sache trancher entre les deux causes, soit de l'inévitable irréproductibilité de l'AFLP (Mougel *et al.*, 2002), soit d'une réelle diversité, au niveau génomique, de souches définies comme homogènes mais par l'analyse d'un seul locus. De plus, au niveau supérieur, certains clades regroupent plusieurs clades-souches. Ceci permet de prendre en compte la diversité interne d'une lignée clonale (sensu Maynard-Smith) (Maynard-Smith *et al.*, 1993), c'est à dire la diversité de souches apparentées mais non identiques, dans un suivi épidémiologique.

Cette étude nous apporte également des renseignements sur l'épidémiologie de la Galle du Collet chez le peuplier. Ceux-ci confirment les résultats obtenus par PCR-RFLP sur le gène 16S et l'ITS, et sur les déterminants du plasmide Ti, indiquant que des isolats d'un même clade-souche déterminé par AFLP peuvent héberger différents plasmides Ti. Ceci indique deux choses, 1) qu'il y a de très nombreux transferts de plasmide Ti entre souches différentes au sein d'une même épidémie, voire d'une tumeur, confirmant les observations précédentes (Picard *et al.*, 1992, Ponsonnet, 1994) ; et 2) que la méthode AFLP n'est pas significativement influencée par le contenu plasmidique de la cellule. La méthode permet donc de faire un suivi épidémiologique de souches essentiellement définies au niveau chromosomique. Par conséquence, elle doit donc permettre de regrouper dans un même clade des isolats pathogènes et non-pathogènes ; et donc de mettre en évidence un lien

Espèce	Clade	Nom	Serotype	Ribotype	Plasmide
G1	a	85-52	1c	2517	2516
		85-66	1c	2517	2516
	b	X88-303	1b	2516	1903
		M2	1c	2516	1903
	c	M143	nd	2177	292
		85-49	1a	2177	2177
		85-2	1a	2177	2177
		X88-293	1a	2177	292
	d	A134-3	nd	85-46	1903
		X88-289	1b	85-46	2516
		X88-283	1b	85-46	2516
	e	M214	1b	85-3	1903
		85-6	1b	85-3	2516
	Hors clade	85-104	1c	85-104	2516
A68-19		nd	A68-20	M80	
A96-11		1b	2516	M80	
M292		1b	M292	M80	
M10		1b	M9	2516	
G4	Hors clade	A134-6	nd	A134-5	1903
<i>R. rhizogenes</i> f		85-123	nd	85-30	2516
		85-30	nd	85-30	2516
	Hors clade	85-186	nd	2520	2516
		C104-22	nd	C104-12	2516
		85-120	nd	85-110	2516
		M111	nd	85-30	M80
		M84	nd	85-30	1903
		M32	nd	85-30	M80
		M3	nd	85-30	1903

Tableau 6.

Répartition en clades des isolats issus de tumeurs de peuplier, en relation avec leurs caractéristiques sérologiques et ribotypiques.

épidémiologique consistant en l'acquisition d'un plasmide Ti pour une souche non-pathogène.

Enfin, cette étude montre, qu'à l'exception de quelques cas, les bactéries pathogènes du peuplier appartiennent à une seule espèce génomique dans chacun des genres concernés : *Agrobacterium* (espèce G1) et *Rhizobium* (espèce *rhizogenes*). Cette faible diversité au niveau des espèces génomiques suggère que les peupliers opèrent une sélection des agrobactéries au niveau de l'espèce. Néanmoins, il n'est pas à exclure qu'il s'agisse d'un résultat contingent à l'épidémie considérée et la preuve de l'existence d'une telle sélection par la plante requiert l'étude d'autres épidémies et d'autres plantes hôtes.

En reprenant l'hypothèse de Cohan (Cohan, 2001, 2002) sur la relation entre espèce génomique et écotypes, s'il y a effectivement sélection d'une espèce génomique dans les tumeurs de peuplier, ces dernières offriraient une niche écologique favorable aux agrobactéries de l'espèce G1. Cohan indique que la structure génétique d'un écotype, se caractérise par une diversification radiale des descendants à partir de l'ancêtre commun. C'est peut-être ce que montre l'arbre de la divergence génomique des isolats de peuplier. En effet, les clades-souches regroupant ces isolats sont regroupés en un « râteau », donc une « structure radiale », mais ne montrent pas de structure hiérarchisée soutenue par des bootstraps. Ainsi, cet arbre des isolats pathogènes du peuplier montre toutes les caractéristiques attendues des isolats d'un même écotype.

Si l'hypothèse d'une sélection par la plante au niveau de l'espèce génomique est vraie, il faut néanmoins expliquer l'existence de deux espèces génomiques sélectionnées, G1 et *Rhizobium rhizogenes*. Nous n'avons pas d'explications à fournir sinon que les genres *Rhizobium* et *Agrobacterium* (tels que définis ici) sont suffisamment éloignés pour exploiter des niches écologiques différentes, et donc ils exploitent probablement des ressources différentes au sein des tumeurs se développant sur une espèce végétale.

## **II. 1. 5. Conclusion**

En conclusion, grâce à l'AFLP nous pouvons réaliser une épidémiologie génomique plus complète que l'épidémiologie monolocus car elle permet de tester statistiquement la significativité des clades obtenus. Cette épidémiologie révèle essentiellement la diversité chromosomique et n'est pas affectée par la présence

d'un plasmide Ti. Enfin, nos résultats suggèrent que ces peupliers pourraient opérer une sélection des agrobactéries au niveau des espèces.

Ceci pourrait avoir d'importantes conséquences sur l'épidémiologie de la Galle du Collet et pourrait être un moyen pour étudier la relation entre espèce génomique et espèce écologique.



## II. 2. Génétique des populations d'*Agrobacterium* associées à *Prunus persica* X *P. amygdalus* (GF677) : épidémiologie génomique.

### II. 2. 1. Introduction

Nous avons montré, au chapitre précédent, que les agrobactéries pathogènes présentes dans les tumeurs de peuplier, appartiennent presque exclusivement à une seule espèce génomique dans chacun des genres concernés, *Rhizobium* et *Agrobacterium*. Ceci suggère qu'il y aurait dans ces tumeurs une sélection des agrobactéries au niveau de l'espèce. Afin de déterminer si une telle sélection des *Agrobacterium*, dirigée vers l'espèce G1 chez le peuplier, s'exerce également chez d'autres plantes hôtes, nous avons étudié la structure des populations d'agrobactéries isolées d'une autre plante, un clone de pêcher-amandier.

Le cultivar GF677 est un hybride de pêcher-amandier (*Prunus persica* X *P. amygdalus*) sélectionné par l'INRA. Il est utilisé comme porte-greffe de pêcher et d'abricotier. Il est adapté aux sols calcaires, et pour cela, est largement cultivé dans le bassin méditerranéen. Cet hybride est reproduit par bouturage, il s'agit donc du même clone. Les boutures sont obtenues soit à partir de cultures *in vitro*, soit à partir de rameaux aériens *a priori* indemnes de maladie. Néanmoins, ce clone est fréquemment atteint de galle du collet suite à des contaminations intervenant lors de la préparation des plants dans les pépinières. Pour cette raison, ce clone a fait l'objet d'une étude particulière dans le cadre du projet européen INCO « Integrated Control of Crown Gall in Mediterranean Countries » (CG-Med, ERBIC18CT97198, (Nesme, 2001)), impliquant la France, l'Espagne, l'Italie, le Maroc, l'Algérie, la Tunisie et la Jordanie.

Outre la relation plante-pathogène, les agrobactéries sont également impliquées dans des relations associatives commensales notamment dans les rhizosphères. Il est donc également intéressant de vérifier s'il existe une sélection par la plante des agrobactéries colonisant la rhizosphère. Comme le plasmide Ti est conjugatif et que le lieu privilégié de la conjugaison est la tumeur (Petit and Tempe, 1978), il est raisonnable de supposer que les agrobactéries non-pathogènes associées au GF677 sont les plus abondantes et donc seraient les récepteurs privilégiés de plasmide Ti. Afin de vérifier cette hypothèse, nous avons simultanément étudié les agrobactéries

pathogènes isolées de tumeurs et des agrobactéries non-pathogènes isolées de rhizosphères saines du GF677.

Il est possible de mener une analyse épidémiologique dans le genre *Agrobacterium* en se basant sur l'analyse des plasmides TI (Pionnat *et al.*, 1999a). Mais celle-ci est limitée car on observe une faible diversité chez les plasmides Ti hébergés par les agrobactéries associées aux fruitiers à noyaux (Nesme, communication personnelle). Notre but est donc d'étudier la diversité des agrobactéries au niveau spécifique et intra-spécifique, d'identifier les espèces génomiques et les souches, au sens de Tenover, (Tenover *et al.*, 1995). Nous utilisons pour cela l'AFLP, une méthode d'analyse génomique très résolutive (Mougel *et al.*, 2002), permettant de déterminer simultanément l'espèce et de réaliser un typage infra-spécifique.

L'objectif de cette étude est d'analyser et de comparer la diversité génomique des agrobactéries isolées du clone GF677 aux niveaux spécifique et intra-spécifique afin d'en étudier l'épidémiologie en relation avec une éventuelle sélection des agrobactéries. Cette analyse doit nous permettre en outre, de vérifier si la sélection observée s'exerce sur les mêmes espèces ou écotypes d'agrobactéries pathogènes et non-pathogènes.

## **II. 2. 2. Matériel et méthodes**

Les *Agrobacterium* et *Rhizobium* pathogènes analysés dans cette étude ont été isolés de tumeurs de GF677 provenant de cinq pays du pourtour méditerranéen, la France, l'Italie, la Jordanie, le Maroc et l'Espagne, dans le cadre du programme de recherche européen INCO (Nesme, 2001). Les agrobactéries non-pathogènes associées aux rhizosphères de plants sains de GF677 ont été isolées de plantes propagées dans les conditions normales de culture, dans une pépinière ayant joué un rôle majeur dans la diffusion mondiale du cultivar GF677.

### *II. 2. 2. a. Isolement, purification, tests biochimiques et de pathogénie, culture et conservation des agrobactéries*

#### Isolement

Un broyat de matériel végétal ou de sol est préparé dans de l'eau stérile. Celui-ci est ensuite étalé, soit par dilutions successive, soit par isolement trois secteurs, sur les milieux 1A-Te et 2E-Te (Mougel *et al.*, 2001) spécifiques respectivement

Espèce	Nom	Origine biologique et géographique	Référence
G1	ATCC 4720 TT111 NCPBP 396 S-377 S-56 S-4 CFBP 5622 CFBP 2517 CFBP 5771 CP85.6	Framboisier, USA Galle, USA <i>Dahlia</i> sp., Allemagne Plante plante plante <i>Solanum nigrum</i> , tissus racinaires, LCSA, France <i>Populus leuce</i> Sol nu, LCSA France <i>P. leuce</i> , galle, France	Kerstens <i>et al</i> , 1973  Popoff <i>et al</i> , 1983 Popoff <i>et al</i> , 1983 Popoff <i>et al</i> , 1983 Mougel, 2000 Ponsonnet, 1994 Mougel, 2000 Ponsonnet, 1994
G2	M 2/1 CIP 497-74 CIP 28-75 CIP 43-76	Eau de fossé, Belgique Prélèvement humain Prélèvement humain Prélèvement humain	Kerstens <i>et al</i> , 1973 Keridjian, 1979 Keridjian, 1979 Keridjian, 1979
G3	CFBP 6623 CFBP 6624	Flacon d'éosine Prélèvement humain	Popoff <i>et al</i> , 1983 Keridjian, 1979
G4	B6 ATCC 4452 ATCC 4718 Kerr 14 O322 ATCC 19358 CFBP 5621 CFBP 5769 ICPB TR6 CIP 67-1 CFBP 5627 360-1	Pommier <i>Rubus idaeus</i> USA <i>Prunus dulcis</i> , Sud de l'Australie <i>Prunus persica</i> , galle sur tronc, Australie <i>Lotus corniculata</i> , tissus racinaires, LCSA France <i>Librocedrus</i> sp., galle  Autre nom pour B6 Sol nu, LCSA France <i>Vitis vinifera</i> , Espagne	Kerstens <i>et al</i> , 1973    Skerman <i>et al</i> , 1980 Mougel, 2000  Kerstens <i>et al</i> , 1973  Mougel, 2000
G5	CFBP 6625 CFBP 6626	Aliment pour malade Prélèvement humain	Popoff <i>et al</i> , 1983 Kerstens <i>et al</i> , 1973
G6	NCPBP 925 F/1 Zutra	<i>Dahlia</i> sp., Afrique du sud <i>Dahlia</i> sp., Israël	Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973
G7	CIP RV3 NCPBP 1641 Zutra 3/1	<i>Flacourtia indica</i> , Royaume Uni <i>Malus</i> sp., Israël	Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973
G8	ICPB TT9 T37 J-07 6-Mushin LMG 75 LMG 46 AW137 C58	<i>Juglans</i> sp., galle, Californie USA 1cm <sup>3</sup> du sol, LCSA France <i>Humulus lupulus</i> , galle, Victoria Australie <i>Euonymus alata</i> cv <i>Compacta</i> , galle, USA <i>Rubus macropetalus</i> , Oregon USA  <i>Prunus</i> sp cv Montmorency (cerise), New York USA	Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973 Vogel <i>et al</i> , 2003 Kerstens <i>et al</i> , 1973
G9	o362 o363	Potting soil, Australie Potting soil, Australie	Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973
<i>Agrobacterium</i> sp.	NCPBP 1650	<i>Rosa</i> sp., Afrique du sud	
<i>A. larrymoorei</i>	AF 3.10	<i>Ficus benjamina</i>	Bouzar <i>et al</i> , 1995
<i>A. rubi</i> (G11)	LMG 17935	<i>Rubus ursinus</i> var <i>Loganobaccus</i> , USA	
<i>A. vitis</i>	CFBP 5523		
<i>Rhizobium rhizogenes</i> (G10)	CFBP 2408 K-84 CP85.120	Pommier <i>Prunus persicae</i> , sol autour d'une galle, Australie <i>P. leuce</i> , galle, France	Ponsonnet, 1994

Tableau 7.  
Souches d'*Agrobacterium* de référence analysées parallèlement aux isolats de GF677.  
LCSA: La Côte Saint André

d'*Agrobacterium spp.* et de *Rhizobium rhizogenes* et placés à 28°C (protocoles d'isolement et composition des milieux en Annexe).

#### Purification des isolats

Après six à huit jours de culture les colonies typiques d'*Agrobacterium* (noires avec la surface présentant un reflet métallique cuivré, rondes et bombées sur milieu 1A-Te ; muqueuses, rondes et vertes à reflets métallique sur milieu 2E-Te), sont transférées sur milieu MG + Tellurite (Mougel *et al.*, 2001) et incubées à 28°C pendant 24 à 48 heures. Les colonies noires au centre, avec un liseré blanc, rondes, bombées et qui ne rendent pas le milieu fluorescent, sont placées en agitation une heure dans de l'eau stérile, puis transférées à nouveau sur milieu MG + Tellurite. La pureté est considérée comme suffisamment assurée après trois isolements successifs sur MG + Tellurite après passage d'une heure dans l'eau stérile, conformément aux recommandations du projet INCO (protocole et milieux en Annexe)

#### Tests biochimiques

En accord avec les recommandations du projet INCO, l'appartenance des isolats au genre *Agrobacterium* (e.g. biovar1 et biovar2) a été confirmée par des tests positifs pour la présence d'uréase et de  $\beta$ -glucosidase. L'appartenance au biovar 1 est déterminée par la capacité à produire du 3-cétolactose (protocoles et milieux en Annexe). Les isolats ne produisant pas de 3-cétolactose appartiennent généralement au biovar 2, qui correspond à l'espèce *Rhizobium rhizogenes* comme le confirme l'analyse du gène *rrs* (Sawada *et al.*, 1993). Dans la suite du travail nous nous référerons aux *Agrobacterium* de biovar 1 comme à des bactéries appartenant au genre *Agrobacterium* et aux bactéries de biovar 2 comme appartenant à l'espèce *R. rhizogenes*. Le terme « agrobactéries » recouvre indifféremment les isolats d'*Agrobacterium* et de *R. rhizogenes*.

#### Tests de pathogénie

La pathogénie des isolats est déterminée par des tests *in planta* et est confirmée par détection PCR d'une portion du plasmide Ti (Ponsonnet and Nesme, 1994).

Le développement d'une tumeur sur kalanchoë après inoculation par la souche à tester confirme son pouvoir pathogène.

L'amplification par PCR, d'une zone du plasmide Ti située entre les gènes *virG* et *virB11*, avec les amorces F14 (GAACGTGTTTCAACGGTTCA) et F749 (GCTAGCTTGGAAGATCGCAC) (Mougel *et al.*, 2001), indique que la souche testée

Pays	Nom	Localité	greffon
Jordanie	J1045	Vallée du Jourdain	Pêche
	J1046	Vallée du Jourdain	Pêche
	J1152	Al Shubak	Pêche
	J1240	Al Mafrak	Abricot
	J1242	Al Mafrak	Abricot
	J1244	Al Mafrak	Abricot
	J1245	Vallée du Jourdain	Abricot
	J1246	Vallée du Jourdain	Abricot
	J1248	Vallée du Jourdain	Abricot
	J1249	Vallée du Jourdain	Abricot
	J1250	Vallée du Jourdain	Abricot
	J1251	Vallée du Jourdain	Abricot
	J1252	Vallée du Jourdain	Abricot
	J1253	Vallée du Jourdain	Abricot
Italie	63n6	Battipaglia (Campanie)	
	64n6	Battipaglia (Campanie)	
	64s10	Battipaglia (Campanie)	
	66n7	Battipaglia (Campanie)	
	67s11	Battipaglia (Campanie)	
	67s12	Battipaglia (Campanie)	
	68n2	Battipaglia (Campanie)	
	69n9	Battipaglia (Campanie)	
	69n11	Battipaglia (Campanie)	
	70b2	Battipaglia (Campanie)	
	72n6	Battipaglia (Campanie)	
	76b1	Battipaglia (Campanie)	
	76n8	Battipaglia (Campanie)	
	84s9	Battipaglia (Campanie)	
	85n6	Battipaglia (Campanie)	
	85s9	Battipaglia (Campanie)	
	115b1	Carinola (Campanie)	
	115b2	Carinola (Campanie)	
	115n6	Carinola (Campanie)	
	115s12	Carinola (Campanie)	
	116n7	Carinola (Campanie)	
	116s10	Carinola (Campanie)	
	134b3	Villasor (Sardaigne)	
134b4	Villasor (Sardaigne)		
137n7	Villasor (Sardaigne)		
137n8	Villasor (Sardaigne)		
137s9	Villasor (Sardaigne)		
137s10	Villasor (Sardaigne)		
Espagne	E678-1	N. I.	
	E678-13	N. I.	
Maroc	M513.11	Marrakech	
	M513.12	Marrakech	
	M513.14	Marrakech	
	M514.04	Marrakech	
	M514.05	Marrakech	
	M514.06	Marrakech	
	M514.09	Marrakech	
	M516.11	Marrakech	
France	CFBP 2712	Vaucluse (Ridé, 1984)	

Tableau 8.

Souches d'agrobactéries pathogènes isolées de tumeurs de GF677.

N. I. : Non Indiqué

porte des gènes de virulence du plasmide Ti, et, en pratique, hébergent un plasmide Ti (protocole PCR en Annexe). Dans la suite du rapport, les souches hébergeant un plasmide Ti (pTi) sont dites « pathogènes », et celles qui n'en possèdent pas sont dites « non-pathogènes ».

#### Culture des agrobactéries

En routine, les agrobactéries sont cultivées sur boîtes MG à 28°C pendant 24 à 48 heures.

Pour l'extraction de l'ADN, elles sont cultivées à 28°C, 24 à 48 heures en milieu LB sans sel liquide (composition des milieux en Annexe).

#### Conservation des isolats

En routine, les *Agrobacterium* sont conservés à 4°C dans de l'eau stérile ou sur pentes de gélose de LPGA ou MG, à température ambiante.

Les isolats sont conservés en souchier à -80°C en suspension dans une solution de glycérol à 50% dans de l'eau stérile (protocoles en Annexe).

### II. 2. 2. b. Souches bactériennes

Lors de cette étude nous avons analysé 91 souches isolées de tumeurs ou de rhizosphères de GF677.

#### Souches isolées de tumeurs

Les souches pathogènes d'*Agrobacterium* analysées lors de ce travail ont été isolées au cours d'un projet européen INCO, Integrated Control of Crown Gall in Mediterranean Countries (CG-Med, ERBIC18CT97198) et fournis par les participants à ce projet. Ces isolats proviennent d'Espagne (2 isolats fournis par M. Lopez), d'Italie (28 isolats fournis par A. Zoina), de Jordanie (14 isolats fournis par H. Klhaif) et du Maroc (8 isolats fournis par A. Benjama). A cette collection s'ajoute un isolat de la pépinière productrice de GF677 (Vaucluse, France) (Tableau 8).

#### Souches isolées de rhizosphères saines

Les agrobactéries ont été isolées de rhizosphères de plants sains à deux stades de croissance de GF677 produit dans la pépinière modèle de cette étude située dans le Vaucluse (France).

Le premier stade de croissance des GF677 correspond à de jeunes plants issus de culture *in vitro* placés dans des godets de tourbe. Ces godets sont placés sur une dalle de béton sans contact avec le sol et l'eau utilisée pour l'irrigation est au préalable traitée aux ultra violets bactéricides.

Isolats piégés dans le sol du Vaucluse		
Nom	Plant	Compartiment
P03.11	C	TR
P03.68	B	TR
P03.69	B	TR
P03.76	D	SR
P03.91	D	TR
P03.98	A	SR
P03.99	A	SR
P03.101	A	SR
P03.102	A	SR
P03.103	A	SR
P03.104	A	SR
P03.105	A	SR
P03.107	A	SR
P03.108	A	SR
P03.109	A	SR
P03.225	C	SR
P03.230	C	TR
P03.236	D	SR
P03.257	B	SR
P03.258	B	SR
P03.259	B	SR

Isolats piégés dans la tourbe		
Nom	Plant	Compartiment
PGF01.13	2	TR
PGF01.14	2	TR
PGF01.16	1	TR
PGF01.17	1	TR
PGF01.18	1	TR
PGF01.20	1	TR
PGF01.22	2	SR
PGF01.26	2	SR
PGF01.57	2	SR
PGF01.59	2	SR
PGF01.76	2	SP
PGF01.81	1	TR
PGF01.84	1	TR
PGF01.100	1	SR
PGF01.101	1	SR
PGF01.113	2	TR
PGF01.114	2	SP

Tableau 9.

Agrobactéries isolées dans les rhizosphères de GF677 cultivés dans le sol du Vaucluse (France) ou en godets de tourbe

SR: Sol Rhizosphérique

TR: Tissus Racinaires

SP: Portion de tourbe libre de racines

Ultérieurement, nous nous référerons aux agrobactéries isolées de deux plants à ce stade comme les « bactéries de la tourbe » (Tableau 9).

Dans un deuxième stade de croissance, les plants acclimatés en tourbe sont transférés dans le sol de la pépinière pour suivre une culture conventionnelle. Les quatre plants analysés ont passé environ un an dans ce sol avant arrachage. Dans la suite de l'étude nous nous référerons aux agrobactéries isolées de ces plants comme les « bactéries du sol du Vaucluse » (Tableau 9).

De ces deux types de plants, nous avons isolés les agrobactéries présentes dans deux compartiments de la rhizosphère, le sol rhizosphérique (SR) et les tissus racinaires (TR), comme définis par Mougel (2000). De plus, nous avons isolé les agrobactéries colonisant une zone de tourbe sans racines des pots des GF677 au premier stade de croissance (SP).

#### Souches de référence

Afin d'identifier les espèces génomiques nous avons inclus dans notre étude un ensemble de souches de références représentant toutes les espèces génomiques d'*Agrobacterium* spp. connues à ce jour (Tableau 7).

### *II. 2. 2. c. Analyses moléculaires*

#### Extraction d'ADN

L'ADN des agrobactéries est extrait avec le kit DNeasy TissueKit de Qiagen (Hilden, Allemagne), selon les recommandations du fournisseur (protocole en Annexe)

#### Séquençage du gène *rrs*

Les espèces génomiques mises en évidence par AFLP, ont été confirmées par séquençage du gène *rrs* de quelques souches représentatives. Le gène *rrs* est amplifié par PCR avec les amorces F667pA (AGAGTTTGATCCTGGCTCAG) et F668pH (AAGGAGGTGATCCAGCCGCA) (Bruce et al., 1992). Si nécessaire les produits PCR sont purifiés, par élution à partir du gel d'agarose, avec le « Gel extraction kit » (Qiagen Hilden, Allemagne), selon les recommandations du fournisseur. Les produits PCR ont été séquençés par la société GenomeExpress (Meylan, France).

#### Séquençage du gène *chvA*

Nous avons séquençé le gène *chvA* des souches de l'espèce G4 de notre modèle de référence. Les amorces utilisées pour l'amplification PCR sont F2044chvA1dgT3 (ATTAACCCTCACTAAAGGGATTCGGCCGWATCATYGACGC) et F2047chvA2T7



(TAATACGACTCACTATAGGGCGATGATGAAGGTCGTCC). Les produits PCR ont été séquencés par la société GenomeExpress (Meylan, France).

#### AFLP

L'AFLP utilisé dans cette étude est adapté du protocole de Vos *et al.* (Vos *et al.*, 1995), en utilisant *EcoRI* et *MseI* pour digérer l'ADN génomique. Les analyses sont menées sur une quantité de 55ng d'ADN génomique. Aux fragments obtenus, des adaptateurs spécifiques des deux sites de restriction sont liés (adaptateur *EcoRI* : F1363AdEco+ CTCGTAGACTGCGTACC et F1931AdEco- AATTGGTACGCAGTCTAC, adaptateur *MseI* : F1365AdMse+ GACGATGAGTCCTGAG et F1931AdMse- TACTCAGGACTCAT). Les amorces F1247coreEco (GACTGCGTACCAATTC) et F1248coreMse (GATGAGTCCTGAGTAA) dont les séquences correspondent aux adaptateurs sont utilisées pour une première PCR non-sélective. La PCR sélective est effectuée avec l'amorce F1247coreMse et des amorces contenant CoreEco plus deux nucléotides discriminants (CA, amorce F1598EcoCA-FAM ; CC, amorce F1599EcoCC-HEX ; CT, amorce F1915EcoCT-FAM ; ou CG, amorce F1601EcoCG-HEX) sont ajoutés en position 3'. Ces trois amorces sélectives différentes sont marquées en fluorescence par les fluorochromes FAM (pour F1598EcoCA-FAM et F1915EcoCT-FAM) et HEX (pour F1599EcoCC-HEX et F1601EcoCG-HEX). L'ADN polymérase, les nucléotides et les tampons nécessaires à la réaction de PCR sont fournis par « l'AFLP amplification CoreMix » (Perkin-Elmer Applied Biosystems, Foster City, Californie). La taille des fragments marqués est déterminée par passage sur séquenceur automatique à capillaires, Megabace 1000 (Amersham Pharmacia Biotech Europe, Orsay, France) par comparaison au marqueur de taille « MegaBACE ET-900-R size standard » (Amersham biosciences, Orsay, France). Le logiciel Genetic Profiler V. 1.5. (Molecular Dynamics Inc., Sunnyvale, Californie) permet de visualiser les résultats, de fixer les seuils d'analyse et d'exporter les tailles de fragments sous forme d'un tableau compatible avec les logiciels d'analyse phylogénomique (protocole en Annexe).

#### *II. 2. 2. d. Analyses phylogénomiques*

Les logiciels Align2, Lis et LecPCR permettant la préparation des données, pour, respectivement, comparer les duplications d'AFLP, lisser les différences de taille de fragments entre répétitions et mettre les données en forme binaire, ainsi que le

Souches isolées de :	G13	G1	G4	<i>Rhizobium rhizogenes</i>	<i>Agrobacterium albertimagni</i>
Tumeurs	0	4	28	21	0
Sol du Vaucluse	0	9	5	3	0
Tourbe	15	1	0	0	1

Tableau 10.

Répartition en espèces génomiques des agrobactéries isolées de GF677

logiciel DistAFLP permettant de calculer les distances génomiques évolutives sont décrits par Mougel *et al.* (2002). Ils sont accessibles à travers le site internet du Pôle Bioinformatique Lyonnais à l'adresse suivante : <http://pbil.univ-lyon1.fr/ADE-4/microb/>.

Les cladogrammes phylogénomiques sont calculés avec le module Neighbor/UPGMA du package PHYLIP (Felsenstein, 2004) et visualisés avec le logiciel NJ-Plot (<http://pbil.univ-lyon1.fr/>).

Les bootstraps sont calculés avec DistAFLP et le module Consens du package PHYLIP (<http://evolution.genetics.washington.edu/phylip.html>).

Les analyses obtenues avec des conditions d'AFLP indépendantes (trois amorces sélectives différentes, F1598EcoCA-FAM, F1915EcoCT-FAM et F1601EcoCG-HEX) sont concaténées afin d'obtenir la meilleure précision possible dans les analyses phylogénomiques (procédures d'analyse en Annexe).

## **II. 2. 3. Résultats**

Un total de 53 agrobactéries pathogènes et 31 non-pathogènes ont été isolées respectivement de tumeurs, de rhizosphères saines de GF677 et de portions de tourbe sans racines.

### *II. 2. 3. a. Identification des espèces génomiques*

#### Isolats pathogènes issus de tumeurs de GF677

En accord avec la procédure d'identification des espèces génomiques décrite par Mougel *et al.* en 2002 et validée par Stackebrandt *et al.* en 2002, l'analyse phylogénomique à partir des données AFLP montre que les 53 souches isolées de tumeurs de GF677 se répartissent en trois espèces génomiques. Les clades correspondant aux espèces génomiques sont soutenus par des valeurs de bootstraps comprises entre 91 et 100% et la distance génomique (clades-espèces) évolutive minimale entre deux espèces génomiques est de 12,1% (données non montrées).

Quatre de ces isolats (7,5%) appartiennent à l'espèce G1, 28 à l'espèce G4 (53%) et 21 à *Rhizobium rhizogenes* (39,5%) (Tableau 10).

La répartition en espèces génomiques de ces isolats n'est pas dépendante de leur origine géographique.

#### Agrobactéries non-pathogènes piégées dans le sol du Vaucluse

Espèce	Clade	Nom	Pays	Localité	origine biologique	Pathogène
<i>R. rhizogenes</i>	a	63n6	Italie	Battipaglia (Campanie)		+
		69n9	Italie	Battipaglia (Campanie)		+
		69n11	Italie	Battipaglia (Campanie)		+
		70b2	Italie	Battipaglia (Campanie)		+
		72n6	Italie	Battipaglia (Campanie)		+
		76b1	Italie	Battipaglia (Campanie)		+
		76n8	Italie	Battipaglia (Campanie)		+
		84s9	Italie	Battipaglia (Campanie)		+
		85n6	Italie	Battipaglia (Campanie)		+
		115s12	Italie	Carinola (Campanie)		+
		116n7	Italie	Carinola (Campanie)		+
	116s10	Italie	Carinola (Campanie)		+	
	b	134b3	Italie	Villasor (Sardaigne)		+
		137n7	Italie	Villasor (Sardaigne)		+
	c	P03.11	France	Sol du Vaucluse	Plant C, TR	-
P03.76		France	Sol du Vaucluse	Plant D, SR	-	
P03.101		France	Sol du Vaucluse	Plant A, SR	-	
P03.103		France	Sol du Vaucluse	Plant A, SR	-	
G1	d	P03.68	France	Sol du Vaucluse	Plant B, TR	-
		P03.69	France	Sol du Vaucluse	Plant B, TR	-
	d'	P03.68	France	Sol du Vaucluse	Plant B, TR	-
		P03.69	France	Sol du Vaucluse	Plant B, TR	-
		NCPB396 CFBP 5771	Allemagne France	<i>Dhalia</i> sp. Sol LCSA		+ -
	e	P03.98	France	Sol du Vaucluse	Plant A, SR	-
		P03.99	France	Sol du Vaucluse	Plant A, SR	-
		P03.236	France	Sol du Vaucluse	Plant D, SR	-
	f	E678-1	Espagne	N. I.		+
		E678-13	Espagne	N. I.		+
G4	g	J1244	Jordanie	Al Mafrak		+
		J1245	Jordanie	Vallée du Jourdain		+
		J1246	Jordanie	Vallée du Jourdain		+
		J1248	Jordanie	Vallée du Jourdain		+
		J1249	Jordanie	Vallée du Jourdain		+
		J1250	Jordanie	Vallée du Jourdain		+
		J1251	Jordanie	Vallée du Jourdain		+
		J1252	Jordanie	Vallée du Jourdain		+
		J1253	Jordanie	Vallée du Jourdain		+
		Kerr 14	Australie	N. I.	<i>Prunus dulcis</i>	+
		h	P03.107	France	Sol du Vaucluse	Plant A, SR
	P03.108 85s9		France Italie	Sol du Vaucluse Battipaglia (Campanie)	Plant A, SR	- +
	i	66n7	Italie	Battipaglia (Campanie)		+
		67s11	Italie	Battipaglia (Campanie)		+
		67s12	Italie	Battipaglia (Campanie)		+
	j	J1152	Jordanie	Al Shubak		+
		J1240	Jordanie	Al Mafrak		+
		J1242	Jordanie	Al Mafrak		+
		360-1	Espagne	N. I.	<i>Vitis vinifera</i>	+
	k	P03.225	France	Sol du Vaucluse	Plant C, SR	-
		P03.230	France	Sol du Vaucluse	Plant C, TR	-
	l	P03.91	France	Sol du Vaucluse	Plant D, TR	-
		P03.109	France	Sol du Vaucluse	Plant A, SR	-
	m	513.11	Maroc	Marrakech		+
		514.04	Maroc	Marrakech		+
514.09		Maroc	Marrakech		+	
516.11		Maroc	Marrakech		+	
115b1		Italie	Carinola (Campanie)		+	
137s9		Italie	Villasor (Sardaigne)		+	
n	M513.12	Maroc	Marrakech		+	
	M513.14	Maroc	Marrakech		+	
o	M514.06	Maroc	Marrakech		+	
	137n8	Italie	Villasor (Sardaigne)		+	

Tableau 11.

Clades statistiquement différenciés et détails des isolats les composants

SR: Sol Rhizosphérique, TR: Tissus Racinaires, N. I.: Non Indiqué, LCSA: La Côte Saint André (Isère, France)

Les dix-sept isolats analysés se répartissent dans trois espèces. Neuf appartiennent à l'espèce G1 (53%), cinq à l'espèce G4 (30%) et trois à *Rhizobium rhizogenes* (18%) (Tableau 10). La répartition en espèce de ces isolats n'est pas dépendante du compartiment (SR ou TR) d'isolement.

#### Agrobactéries non-pathogènes piégées dans la tourbe

Quatorze des seize isolats issus de la tourbe appartiennent à une espèce génomique qui ne correspond à aucune des souches de référence disponible. Il s'agit vraisemblablement d'une nouvelle espèce, que nous nommerons G13 dans la suite logique des espèces définies par Popoff *et al.* en 1984 (Tableau 10). La séquence du gène *rrs* s'est révélée être identique à celle du gène *rrs* des souches des espèces G7 et G5 (n° d'accès GeneBank : AJ389903).

L'espèce G13 regroupe cinq isolats issus du sol rhizosphérique, neuf isolats issus des tissus racinaires et deux isolats issus de la zone de tourbe sans racines.

Les deux derniers isolats appartiennent respectivement à l'espèce G1 et à une autre espèce non présente dans nos références. Le gène *rrs* de cette dernière souche est identique à celui d'*Agrobacterium albertimagni* (d'après le n° d'accès GeneBank : AF316615).

#### II. 2. 3. b. Identification des clades dans les espèce génomique

À partir des données AFLP, nous avons pu déterminer la structure fine des populations, espèce par espèce. Nous avons recherché les clades soutenus par des valeurs significatives ( $\geq 80\%$ ) de bootstraps, révélés par au moins deux analyses AFLP indépendantes.

#### *Rhizobium rhizogenes*

Cette espèce regroupe un isolat pathogène du Maroc, 20 isolats pathogènes provenant des trois localités italiennes et quatre isolats non-pathogènes du sol du Vaucluse.

À l'intérieur de cette espèce, les isolats de GF677 se différencient significativement en trois clades, les clades a, b et c (Figure 8). Le clade c, soutenu à 89%, réunit les quatre isolats du sol du Vaucluse (Tableau 11). Les clades a et b, soutenus respectivement à 81 et 98%, regroupent des isolats d'Italie. Le clade a regroupe des isolats de Campanie (Battipaglia et Carinola) et le clade b des isolats de Sardaigne (Tableau 11). Ces deux clades ne sont pas significativement apparentés.

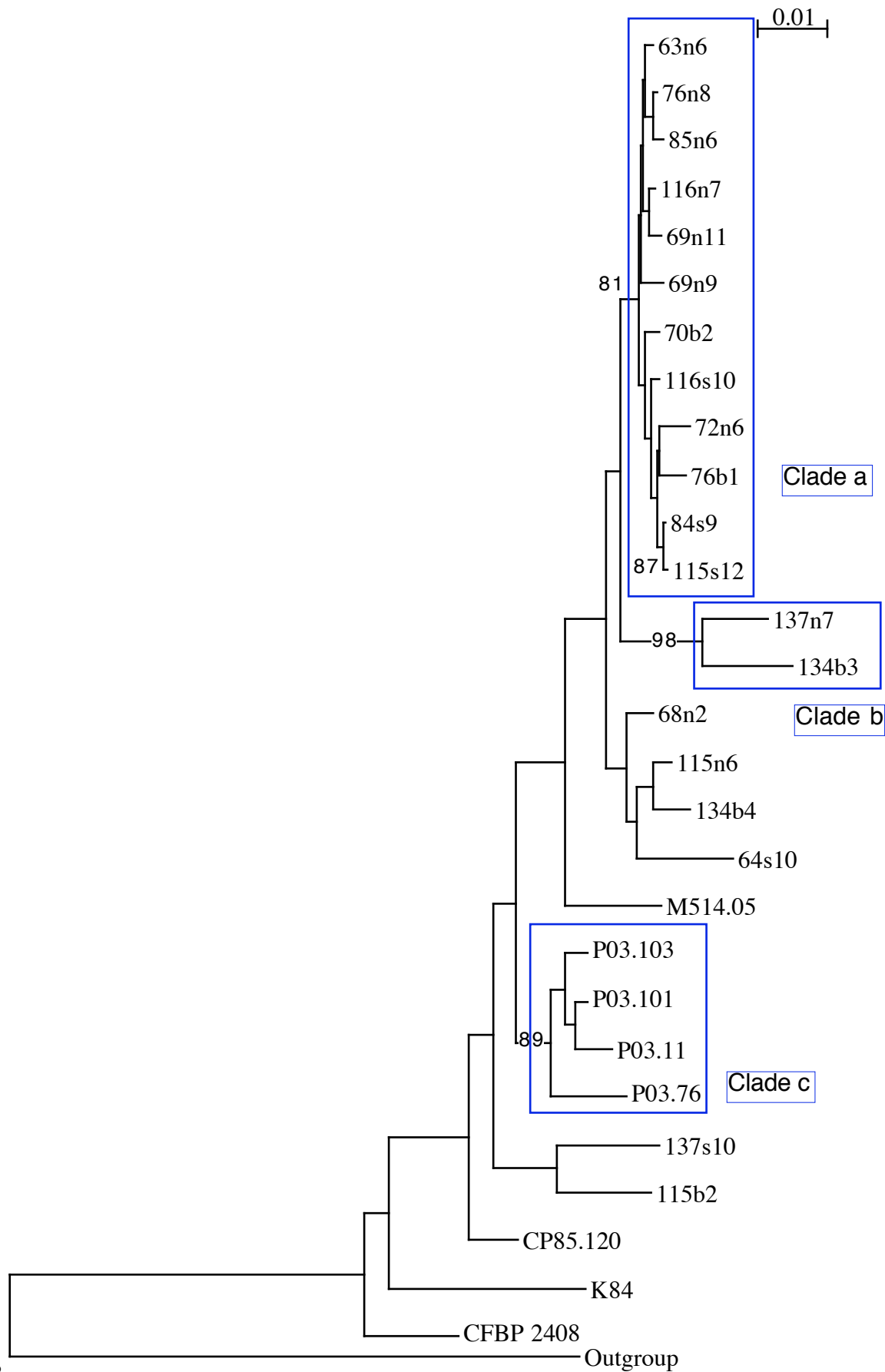


Figure 8. Arbre phylogénomique, des souches de référence et des agrobactéries isolées du cultivar GF677 appartenant à l'espèce *Rhizobium rhizogenes*. Condition AFLP Eco+CA/Mse+0. Les valeurs significatives de bootstrap pour 100 rééchantillonnages sont indiquées.

Les sept derniers isolats de *R. rhizogenes* ne peuvent être regroupés significativement soit entre eux, soit avec des souches de référence.

#### Espèce G1

Cette espèce regroupe onze isolats du sol du Vaucluse, un isolat de la tourbe, un isolat pathogène d'Italie, la souche pathogène de France et deux isolats pathogènes d'Espagne (Figure 9). À l'intérieur de cette espèce, les isolats de GF677 sont très divers. Trois clades se distinguent, les clades d, e et f (Figure 9). Le clade d regroupe deux isolats du sol du Vaucluse, et le clade e en regroupe 3 (Tableau 11). Le clade d apparaît significativement apparenté à deux souches de référence, NCPPB396 et CFBP5771, formant le clade d'. Le clade f regroupe les deux isolats espagnols. Les informations sont résumées dans le Tableau 11. Les neuf autres isolats de GF677 appartenant à cette espèce ne sont pas significativement apparentés ni à d'autres agrobactéries isolées de GF677 ni à des souches de référence.

#### Espèce G4

Cette espèce regroupe les quatorze isolats de Jordanie, neuf isolats d'Italie, sept isolats du Maroc et six isolats non-pathogènes du sol du Vaucluse (Figure 10). Les isolats de GF677 à l'intérieur de cette espèce sont très divers. En effet ils présentent plus de diversité entre eux que celle déjà connue par les souches de référence (Groupe A, Figure 9). Les isolats se regroupent en dix clades (clades g, h, i, j, k, l, m, n et o) soutenus par des valeurs de bootstraps supérieures à 80%.

Le clade g regroupe neuf isolats de Jordanie plus les souche Kerr 14 et O322 isolées respectivement de *Prunus dulcis* et de *P. persica* en Australie.

Le clade h regroupe deux souches non-pathogènes du sol du Vaucluse et une souche pathogène de Battipaglia en Italie.

Le clade i regroupe trois souches pathogènes d'Italie.

Le clade j regroupe trois souches de Jordanie et la souche de référence 360-1 isolée de *Vitis vinifera* en Espagne.

Les clade k et l regroupent chacun deux souches non-pathogènes isolées du sol du Vaucluse.

Le clade m regroupe quatre souches du Maroc et deux souches d'Italie venant l'une de Carinola en Campanie et l'autre de Villasor en Sardaigne.

Les clades n et o sont confirmés, respectivement à 99 et 86% lorsque les données de trois analyses AFLP indépendantes sont concaténées. Le clade n regroupe deux isolats du Maroc et le clade o un isolat du Maroc et un d'Italie (Tableau 11).

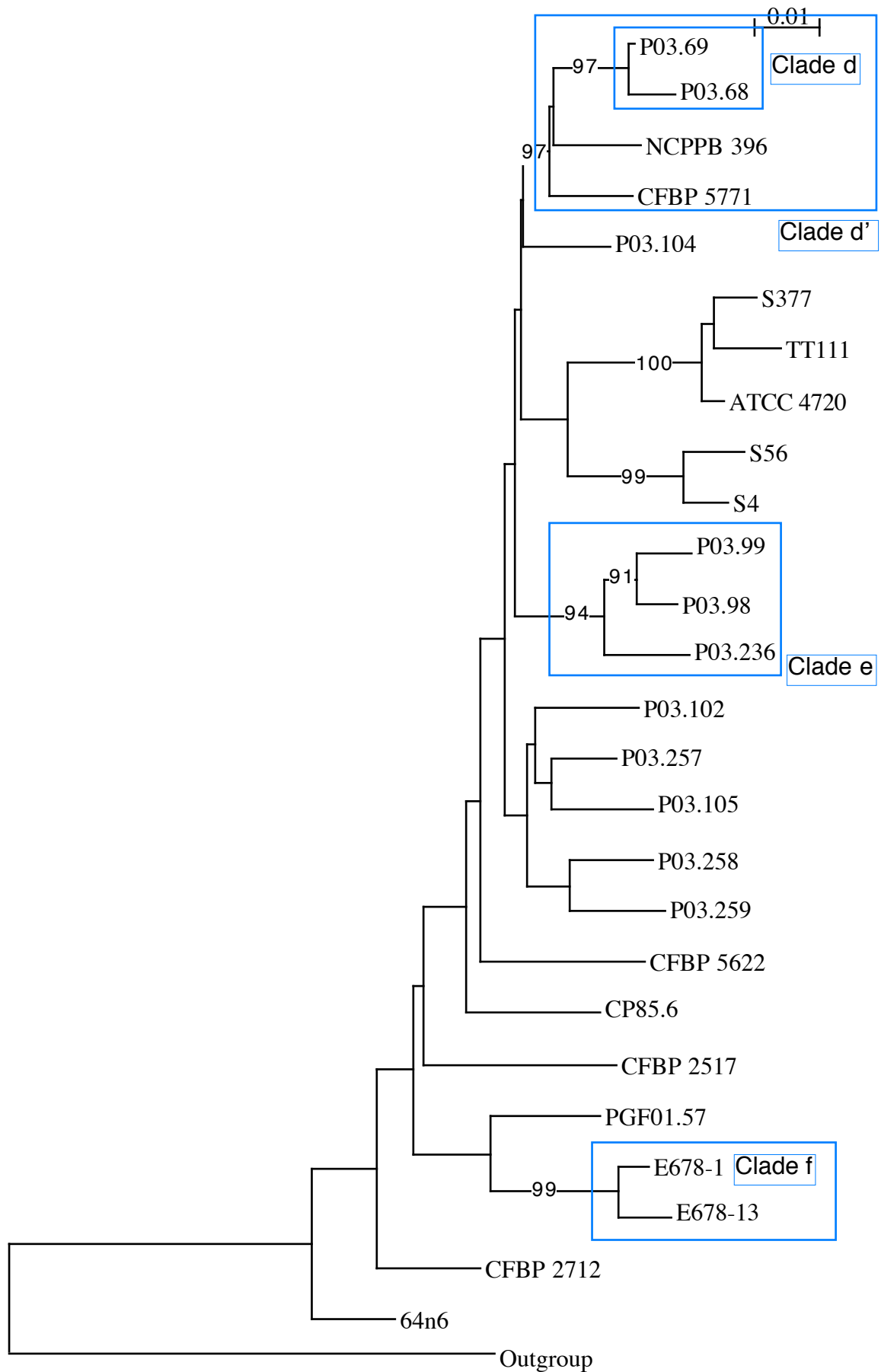


Figure 9. Arbre phylogénomique, des souches de référence et des agrobactéries isolées du cultivar GF677 appartenant à l'espèce G1. Conditions AFLP Eco+CA/Mse+0. Les valeurs significatives de bootstrap pour 100 rééchantillonnages sont indiquées.



Enfin, deux isolats issus de tumeurs de GF677 en Jordanie ne peuvent être regroupés significativement avec d'autres isolats ou souches de références.

À titre indicatif, la souche de référence CIP67-1 et la souche de référence B6 sont deux clones d'un même isolat conservé à l'Institut Pasteur (CIP) et dans la Collection Française de Bactéries Pthyopathogènes (CFBP) via l'ATCC (*American Type Culture Collection*), respectivement. Ces deux souches sont significativement regroupées (à 96%) avec la souche ATCC4452 et présentent entre elles une distance génomique évolutionnaire de 1,5% en moyenne (Figure 10).

Les souches de référence de cette espèce se répartissent en trois groupe suivant la séquence du gène *chvA* (Figure 12). Les analyses AFLP des souches de référence montrent que les regroupements basés sur la séquence du gène *chvA* sont consistants avec ceux basés sur l'analyse génomique par AFLP. Le premier groupe *chvA* correspond à un clade AFLP soutenu à 89%. Les deux groupes suivants correspondent à des clades AFLP soutenus à 100% (résultats non montrés).

#### Isolats de l'espèce génomique G13

Aucune souche de référence n'est proche de ces isolats qui appartiennent à une nouvelle espèce génomique. Toutes sont isolées de la tourbe. Elles forment un seul clade soutenu à 99%, correspondant à l'espèce G13 (Figure 11). La distance génomique moyenne entre les souches de cette espèce est de 2,8%. À l'intérieur de cette espèce, un clade se différencie, soutenu à 96%. La distance génomique moyenne entre les isolats de ce sous-clade est de 1,3%.

## II. 2. 4. Discussion

L'AFLP permet d'analyser simultanément, au niveau spécifique et intra-spécifique, un grand nombre de souches. Cette technique nous a permis d'analyser 91 isolats issus de tumeurs ou de rhizosphères saines de GF677, et de les comparer à 52 souches de référence représentant toutes les espèces génomiques d'*Agrobacterium* connues à ce jour. Ceci représente une avancée méthodologique remarquable, une telle étude génomique était en effet difficilement envisageable par la technique standard d'hybridation ADN-ADN.

### II. 2. 4. a. Structures génétiques intra-spécifiques

Tenover (1995) définit une souche comme des individus dont les profils PFGE (*Pulse Field Gel Electrophoresis*) ne permettent pas de les différencier. Par contre, l'AFLP

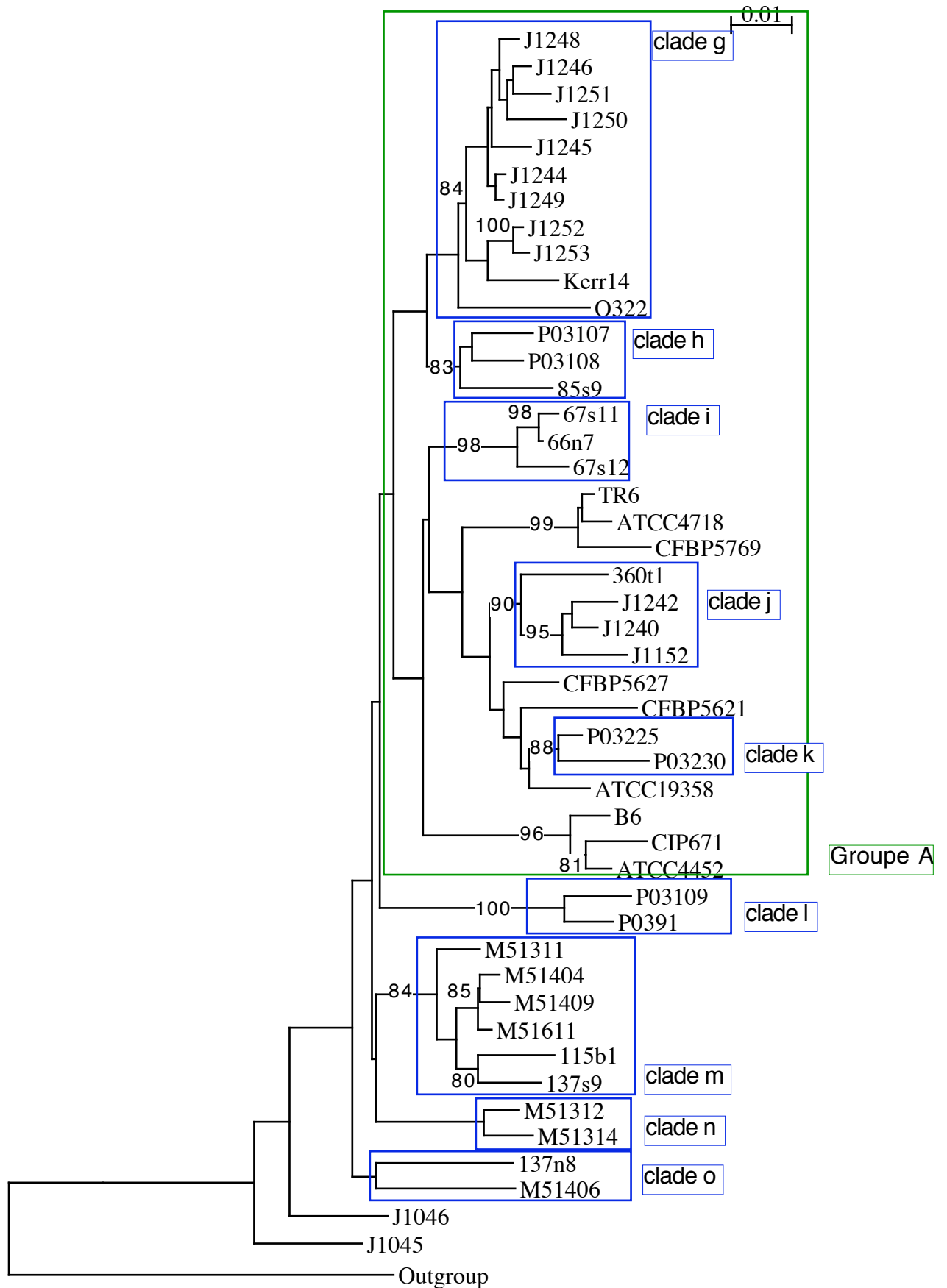


Figure 10. Arbre phylogénomique, des souches de référence et des agrobactéries isolées du cultivar GF677 appartenant à l'espèce G4. Condonation AFLP Eco+CA/Mse+0. Les valeurs significatives de bootstrap pour 100 rééchantillonnages sont indiquées. Le groupe A représente la diversité génomique de cette espèce connue dans les collections.

est une méthode beaucoup plus résolutive que la PFGE, permettant même de distinguer deux clones identiques comme CIP67-1 et B6. Cette diversité entre clones de la même souche résulte probablement de la méthode AFLP elle-même et correspond à un bruit de fond. Néanmoins, la distance génomique moyenne, 1,5%, de substitution nucléotidique par site, donne une mesure vraisemblable du bruit de fond de la mesure de la divergence pour deux isolats de la même souche. À l'exception des isolats du clade o, qui présente une divergence génomique de 4,6% (il est d'ailleurs à noter que ce clade n'est soutenu que si les données de trois analyses indépendantes sont concaténées), dans la présente étude les clades mis en évidence présentent des divergences génomiques de 0,6 à 2,4% (Tableau 11). Chacun de ces clades regroupe vraisemblablement des isolats de la même souche. Nous appellerons ces clades des clades-souches.

Nous avons choisi le gène *chvA*, gène chromosomique impliqué dans la virulence des agrobactéries (Puvanesarajah, 1985), pour commencer une étude par MLST (*Multi Locus Sequence Typing*) du genre *Agrobacterium*. Les premiers résultats montrent que ce gène est variable au sein d'une même espèce. Par contre, les souches de référence de l'espèce G4, regroupées au sein d'un même clade-souche soutenu par des valeurs significatives de bootstrap, partagent la même séquence du gène *chvA* (Figure 12). Ce fait renforce l'hypothèse selon laquelle les clades correspondraient à des souches au sens de Tenover.

#### *II. 2. 4. b. Epidémiologie moléculaire*

Malgré la très grande diversité des agrobactéries associées à GF677 (4 espèces, 15 souches et 18 isolats non-apparentés à d'autres), le regroupement des isolats en clades a permis de mettre en évidence des liens épidémiologiques remarquables.

La plupart des souches déterminées dans cette étude regroupent des isolats de même origine géographique ou du même compartiment rhizosphérique. Cette information est triviale dans la mesure où il est vraisemblable que des isolats appartenant à la même souche puissent être isolés dans la même localité.

Par contre, l'existence de souches regroupant des agrobactéries isolées dans des contextes différents révèle de probables liens épidémiologiques. Nous ne discuterons en détail que des clades-souches les plus pertinents, c'est à dire ceux regroupant des isolats provenant d'origines différentes.

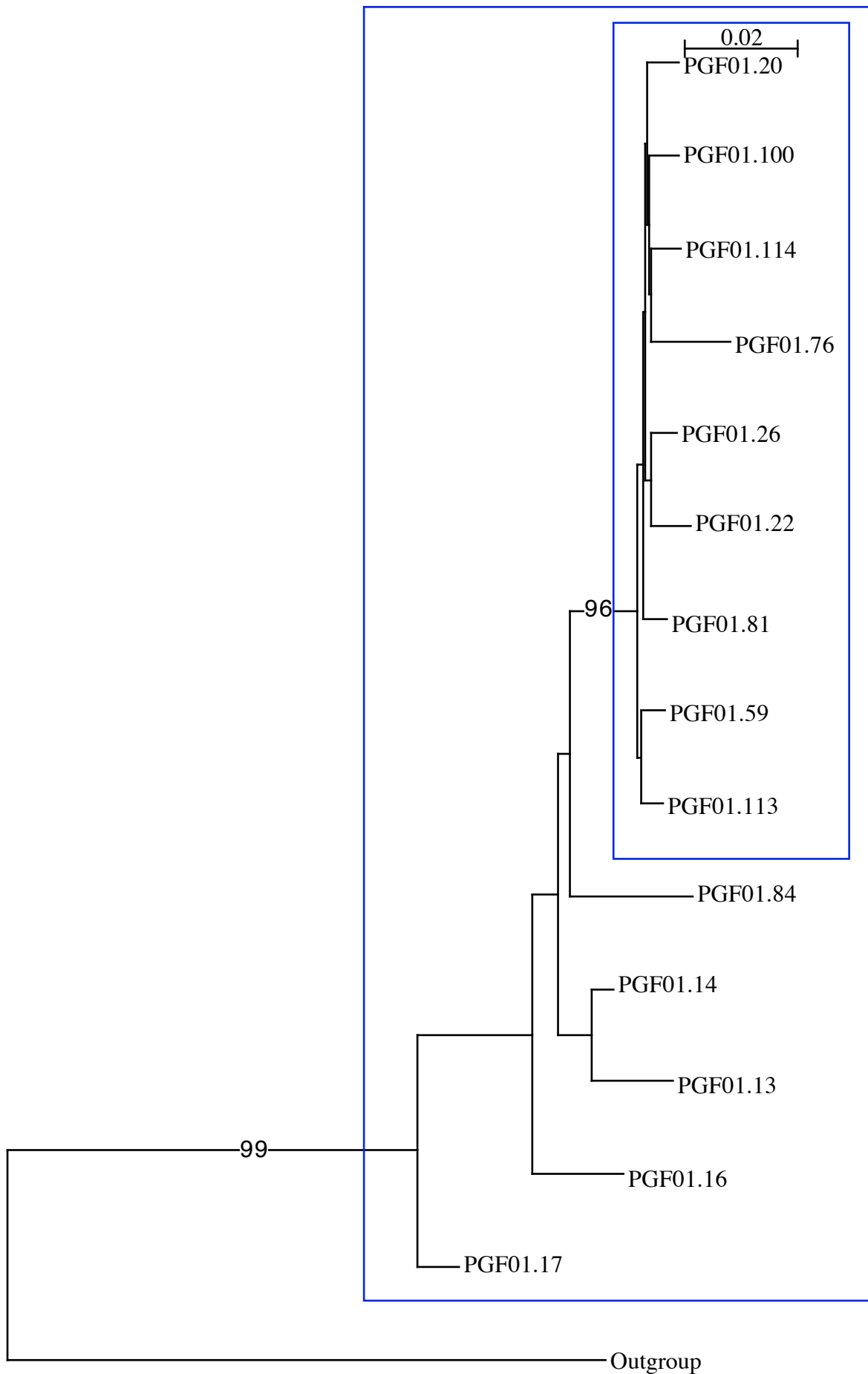


Figure 11. Arbre phylogénomique, des agrobactéries isolées du cultivar GF677 appartenant à l'espèce G13, piégées dans la tourbe. Les données de trois conditions AFLP (Eco+CA/Mse+0, Eco+CT/Mse+0 et Eco+CG/Mse+0) sont concaténées. Les valeurs significatives de bootstrap pour 100 rééchantillonnages sont indiquées.

Concernant les isolats pathogènes, nos résultats montrent qu'il y a indiscutablement des liens épidémiologiques entre les localités d'un même pays. Ceci est vérifié pour l'Italie (clades a et m, Tableau 11) et pour la Jordanie (clade g et j, Tableau 11), indiquant qu'il y a probablement eu des échanges de matériel contaminé entre les pépinières à l'intérieur de ces pays. Cette hypothèse est très vraisemblable car il n'y a généralement pas de limitation à la diffusion de matériel végétal à l'intérieur des frontières. En outre, dans chaque pays, la propagation du matériel végétal sélectionné est assurée par un nombre très limité de pépinières. Par contre, les isolats pathogènes diffèrent généralement d'un pays à l'autre suggérant que chaque pays fait face à une épidémie différente.

Néanmoins, des isolats pathogènes provenant de pays différents peuvent appartenir à un même clade-souche. Ces liens épidémiologiques concernent tout d'abord des isolats provenant d'Italie et du Maroc (clades m et o, Tableau 11). De tels liens concernent aussi des agrobactéries isolées de *Prunus dulcis* et *P. persica* en Australie et de GF677 en Jordanie (clade g, Tableau 11), ou de *Vitis vinifera* en Espagne et de GF677 en Jordanie (clade j, Tableau 11). L'infection de plantes diverses par de la même souche confirme les observations de De Cleene et De Ley (1976) montrant que le spectre d'hôte de ce pathogène est très large. Des travaux ultérieurs (Loper and Kado, 1979 ; Thomashow *et al.*, 1980) ont montré que la gamme d'hôte était déterminée par le plasmide Ti.

La mise en évidence de liens épidémiologiques entre ces pays ne permet pas, cependant, de définir avec certitude le sens dans lequel s'est fait la propagation du matériel contaminé. Au contraire, dans le cas de la parenté entre les isolats non-pathogènes de France et les isolats pathogènes d'Italie (clade h, Tableau 11) il est possible de déterminer le sens de migration des agrobactéries. En effet, il est probable que les agrobactéries soient initialement non-pathogènes puis qu'elles aient reçu un plasmide Ti et l'ait conservé par la suite, le plasmide Ti étant très stable chez *Agrobacterium* (Gallie *et al.*, 1985 ; Koekman *et al.*, 1982).

Le fait que les isolats du clade h de France soient non-pathogènes et que ceux d'Italie soient pathogènes suggère une dissémination des agrobactéries de la France vers l'Italie. L'acquisition du plasmide Ti, probablement par conjugaison (Genetello *et al.*, 1977), a cependant pu avoir lieu lors d'une étape du voyage (soit encore dans le sol français, soit après transfert dans le sol italien).

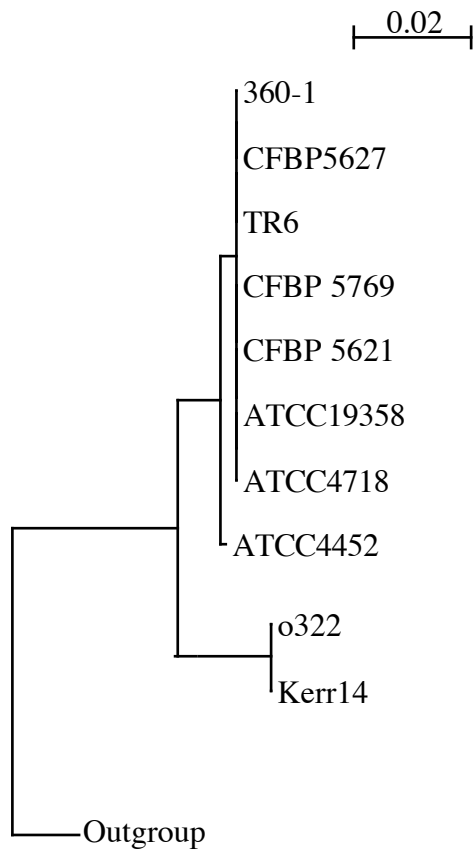


Figure 12.  
Arbre phylogénétique des souches de référence de l'espèce G4 calculé à partir des séquences du gène *chvA*

#### II. 2. 4. c. Découverte d'une nouvelle espèce d'*Agrobacterium*

Un fait marquant et inattendu de la présente étude, est la découverte d'une nouvelle espèce génomique dans le genre *Agrobacterium*. Ceci confirme, si nécessaire la puissance exploratoire de l'AFLP en matière de taxinomie génomique.

Les représentants de G13 ont été isolés uniquement dans les rhizosphères de plants de GF677 cultivés dans la tourbe (Tableau 10). Cette espèce appartient bien au genre *Agrobacterium* (e.g. au biovar1) comme l'avait montré la biochimie (présence d'uréase,  $\beta$ -glucosidase et production de 3-cétolactose (données non montrées), mais surtout la séquence du gène *rrs*. L'identité de cette dernière avec celle de l'espèce génomique G7 confirme deux faits : 1), que cette nouvelle espèce G13 appartient bien au taxon *Agrobacterium*, et 2), que différentes espèces bactériennes peuvent avoir des gènes *rrs* identiques. Ceci avait été montré pour les espèces G6 et G8 ainsi que pour les espèces G2 et G9 (Mougel *et al.*, 2002). Nous confirmons cet état de fait pour les espèces G13 et G7 (ce travail) d'une part, et pour les espèces G13 et G7 avec G5 (résultat non montré) d'autre part.

Parmi les isolats de cette espèce, neuf se différencient en un clade soutenu à 96%. Les individus de ce clade présentent une distance génomique évolutive moyenne de 1,4% indiquant qu'ils forment une unique souche. Les cinq autres isolats de cette espèce appartiennent à des souches différentes. Par contre, la faible distance génomique évolutive moyenne entre tous les individus de l'espèce (2,8%) montre que les différentes souches qui la composent sont très étroitement apparentées.

Le fait que des isolats de l'espèce G13 n'aient été piégés que dans la tourbe est remarquable. Notre hypothèse est que ces *Agrobacterium* sont quasiment les seuls disponibles dans ce substrat. En effet, la tourbe est un substrat très acide beaucoup plus favorable aux champignons qu'aux bactéries (Gobat *et al.*, 2003) (Le sol vivant). Il est donc probable que ces *Agrobacterium* soient quasiment les seuls disponibles dans ce substrat, et ceci expliquerait ce piégeage quasi unique d'isolats de l'espèce G13 dans la tourbe.

Nous avons obtenus quelques isolats provenant de la tourbe dans une zone dépourvue de racines, qui confirme la présence de G13 dans ce substrat. Néanmoins, de plus amples études des agrobactéries présentes dans la tourbe

Souches isolées de	G13	G1	G4	<i>Rhizobium rhizogenes</i>
Tumeurs	0	4	28	21
Sol du Vaucluse	0	11	?	4
Tourbe	15	1	0	0

Tableau 12.

Succession entre les différentes populations colonisants les rhizosphères de GF677



avant plantation sont nécessaires pour confirmer l'écologie de cette nouvelle espèce génomique.

A notre avis, la prévalence exclusive de G13 parmi les isolats d'*Agrobacterium* issus de rhizosphères de GF677 cultivés en tourbe ne préjuge pas d'une affinité particulière de ces *Agrobacterium* pour le GF677. Dans ce cas la population des agrobactéries sur ces racines jeunes serait essentiellement dépendante des *Agrobacterium* présents dans le substrat de culture.

#### *II. 2. 4. d. Succession des populations colonisant les rhizosphères*

Les agrobactéries colonisant les rhizosphères de GF677 cultivés dans la tourbe et le sol du Vaucluse appartiennent à des espèces différentes, montrant qu'il y a une succession des espèces dans ces rhizosphères.

D'une population relativement simple, composée quasi exclusivement d'agrobactéries de l'espèce G13 dans la tourbe, il y a passage à une population plus complexe, composée d'agrobactéries des espèces G1, G4 et *R. rhizogenes* dans le sol du Vaucluse. La disparition totale des agrobactéries de l'espèce G13 après un an de culture dans le sol du Vaucluse renforce la présomption que les agrobactéries de l'espèce G13 sont plus adaptées à la tourbe qu'aux rhizosphères de GF677.

Alors que la succession de populations depuis la tourbe à celle du sol du Vaucluse est expérimentalement avérée sur des plants de même origine, nous n'avons pas rigoureusement observé, sur ces mêmes plantes, le passage à une population pathogène. Cependant, la composition spécifique de la population du sol du Vaucluse est la même que celle de la population pathogène. Seules les proportions relatives des espèces génomiques sont différentes. Mais surtout, certains isolats non-pathogènes du sol du Vaucluse appartiennent au même clade-souche qu'un isolat pathogène d'Italie, suggérant fortement que la population des isolats de l'espèce G4 non-pathogènes du sol du Vaucluse à pu acquérir un plasmide Ti et devenir pathogène à son tour (Tableau 12).

#### *II. 2. 4. e. Association entre agrobactéries et GF677*

Le premier but de ce travail était de déterminer si, comme pour le peuplier, il y a sélection par la rhizosphère de GF677 d'une espèce d'*Agrobacterium* particulière. Nous voulions aussi déterminer si la sélection opérée par les plantes sur les

bactéries pathogènes étaient les mêmes que celle opérée sur les bactéries non-pathogènes.

La présence simultanée, dans ces rhizosphères, de plusieurs espèces génomiques, indique que la situation n'est pas aussi claire que dans le cas du peuplier (Partie II, Chapitre I). L'épidémie touchant le peuplier est récurrente et transmise par les plantes (via les boutures de racines). À l'inverse, le GF677 est concerné par plusieurs épidémies indépendantes, résultants d'épisodes de piégages indépendants. Il n'y a pas d'association exclusive entre une espèce d'*Agrobacterium* et le GF677.

Par contre, nos analyses indiquent que certains isolats non-pathogènes piégés par le GF677 ont probablement reçu un plasmide Ti par conjugaison, apportant une confirmation partielle de notre hypothèse de départ.

La majorité des isolats pathogènes appartiennent à l'espèce G4. De plus, c'est cette espèce qui présente la diversité la plus grande. Dans cette espèce, à partir d'un nœud initial partent onze branches (la plupart soutenues par des bootstraps) et ces onze branches forment un « râteau » (Figure 11). Ceci indique trois faits : 1) que la diversité des isolats de GF677 dans cette espèce est large (et même plus large que la diversité présente dans la collection de référence) ; que 2), cette diversité est difficilement hiérarchisable, suggérant une équidistance génomique avec l'ancêtre commun ; et que 3), la sélection opérée par la plante est globale sur toute l'espèce G4 et pas seulement dirigée vers une souche en particulier.

Ces faits suggèrent qu'il existe une relation privilégiée entre le cultivar GF677 et l'espèce G4 d'*Agrobacterium*. Cette relation n'aboutit pas à une sélection exclusive de l'espèce G4 par la plante, mais tend simplement à favoriser la croissance de ces bactéries et donc leur *fitness*.

A partir de ces données nous pouvons maintenant imaginer de nouvelles expérimentations qui permettraient de vérifier cette hypothèse. Par exemple il serait intéressant de cultiver des plants de GF677 axéniques dans le sol de La Côte Saint André afin de déterminer quelles espèces d'agrobactéries ces plants piègent dans leurs rhizosphères. Il serait intéressant aussi de cultiver de tels plants axéniques dans un sol stérile, recolonisé par une population complexe mais connue d'agrobactéries afin de déterminer si ces plants sont préférentiellement associés aux agrobactéries de l'espèce G4.

#### *II. 2. 4. f. Souches et écotypes*

La diversité au sein de l'espèce G4 révélée par le GF677 est difficilement hiérarchisable, formant un râteau plutôt que des structures dichotomiques. Ceci indiquant que cette espèce pourrait correspondre à un écotype dans le sens de Cohan (Cohan, 2001, 2002).

Selon Cohan, un écotype est un groupe de souches, occupant la même niche écologique, et qui est régulièrement soumis à des épisodes de sélection périodique. Ceci a pour conséquences : 1), le maintien d'une cohésion génétique à l'intérieur de ce groupe ; et 2), la séparation, du plus en plus marquée, d'avec les autres écotypes. Après l'événement de sélection, un écotype recommence à se diversifier mais avec la particularité que chaque nouveau variant dérive de l'ancêtre sélectionné. Les cladogrammes se présentent alors sous une forme caractéristique en étoile ou en râteau, où tous les descendant du même individu fondateur sont tous autant divergents les uns des autres.

L'association préférentielle entre l'écotype correspondant à l'espèce G4 et le cultivar GF677 montre que les plantes, favorisant certaines espèces d'agrobactéries dans leurs rhizosphères, pourraient être à l'origine de la différenciation des populations en écotypes, et à terme responsable de la spéciation au sein du genre *Agrobacterium*.

#### **II. 2. 5. Conclusion**

Dans cette étude, nous avons voulu vérifier l'hypothèse de la sélection préférentielle de certaines agrobactéries par un clone végétal, en analysant comparativement la structure génétique de différentes populations naturelles associées à ce clone.

Cette étude nous a tout d'abord permis de confirmer la puissance de l'AFLP utilisé en génomique des populations.

Une autre conséquence remarquable de travail est la découverte d'une nouvelle espèce d'*Agrobacterium*, probablement spécifique de la tourbe.

Le premier but de ce travail était de déterminer si, comme pour le peuplier avec l'espèce G1, il y a sélection par la rhizosphère de GF677 d'*Agrobacterium* particuliers. Nous voulions aussi déterminer si les interactions entre plantes et bactéries pathogènes étaient les mêmes qu'entre plantes et bactéries non-pathogènes.

Nos analyses indiquent que le cultivar GF677 tend effectivement à favoriser les agrobactéries de l'espèce G4. Les membres de cette espèce sont les plus à même de devenir les récepteurs du plasmide Ti devenant en retour pathogènes.

Les espèces bactériennes étant des écotypes, les niches écologiques différenciant les espèces d'*Agrobacterium* pourraient être les plantes avec lesquelles elles établissent des relations privilégiées : les plantes pourraient être en partie responsables de la spéciation au sein du genre *Agrobacterium*.

## II.3. Analyse des populations d'*Agrobacterium* non pathogènes piégées dans le sol de La Côte Saint André par différentes plantes modèles.

### II. 3. 1. Introduction

À la suite des études menées sur le peuplier et GF677 (Partie II, chapitres I et II) nous supposons que les plantes pourraient sélectionner certaines espèces au sein du genre *Agrobacterium* (Partie II, Chapitre II). Mais ces observations sont probablement dépendantes des agrobactéries disponibles. Les épidémies étudiées ne sont pas directement comparables car elles proviennent de plantes croissantes dans des substrats différents et donc en présence de microflores différentes. Travailler avec des plantes cultivées dans un même sol permet de s'affranchir de la variabilité des microflores présentes dans les sols.

Les relations entre agrobactéries et plantes ne se limitent pas aux interactions plante – pathogène, comme nous l'avons vu au chapitre précédent. Les rhizosphères saines abritent de nombreuses agrobactéries commensales non-pathogènes (qui n'hébergent pas de plasmide Ti).

L'objectif de ce travail était de vérifier si la sélection de certaines espèces d'agrobactéries par les plantes observée pour les agrobactéries pathogènes, s'observe aussi pour les agrobactéries non-pathogènes.

Nous avons au cours de ce travail complété des analyses, effectuées par Christophe Mougel (2000) et par Cathy D'Angelo (2000), effectuées sur le sol nu modèle (sol de La Côte Saint André) et les rhizosphères de maïs, lotier et morelle cultivées dans ce même sol. Les populations naturellement présentes dans le sol de La Côte Saint André ont été isolées et analysées. De plus, des plants sains multipliés *in vitro* ou des semis, des trois plantes modèles ont été cultivés dans ce sol. Après incubation, les agrobactéries piégées par ces plantes dans les rhizosphères ont été isolées et analysées.

Nous avons repris les données précédentes afin de les vérifier (notamment par le séquençage du gène *rrs*) et de les compléter par de nouvelles analyses conduites sur le maïs ainsi que sur la luzerne et le peuplier.

Espèce	Nom	Origine biologique et géographique	Référence
G1	ATCC 4720 TT111 NCPPB 396 S-377 S-56 S-4 CFBP 5622 CFBP 2517 CFBP 5771 CP85.6 CFBP2712	Framboisier, USA Galle, USA <i>Dahlia</i> sp., Allemagne Plante plante plante <i>Solanum nigrum</i> , tissus racinaires, LCSA, France <i>Populus leuce</i> Sol nu, LCSA France <i>P. leuce</i> , galle, France <i>Prunus persica</i> X <i>prunus amygdalus</i> (GF677)	Kerstens <i>et al</i> , 1973  Popoff <i>et al</i> , 1983 Popoff <i>et al</i> , 1983 Popoff <i>et al</i> , 1983 Mougel, 2000 Ponsonnet, 1994 Mougel, 2000 Ponsonnet, 1994 Ridé 1984
G2	M 2/1 CIP 497-74 CIP127-76 CIP 28-75 CIP 43-76	Eau de fossé, Belgique Prélèvement humain Prélèvement humain Prélèvement humain Prélèvement humain	Kerstens <i>et al</i> , 1973 Keridjian, 1979 Keridjian, 1979 Keridjian, 1979 Keridjian, 1979
G3	CFBP 6623 CFBP 6624	Flacon d'éosine Prélèvement humain	Popoff <i>et al</i> , 1983 Keridjian, 1979
G4	B6 ATCC 4452 ATCC 4718 Kerr 14 O322 ATCC 19358 CFBP 5621 CFBP 5769 ICPB TR6 CIP 67-1 CFBP 5627 360-1	Pommier <i>Rubus idaeus</i> USA <i>Prunus dulcis</i> , Sud de l'Australie <i>Prunus persica</i> , galle sur tronc, Australie <i>Lotus corniculata</i> , tissus racinaires, LCSA France <i>Librocedrus</i> sp., galle Autre nom pour B6 Sol nu, LCSA France <i>Vitis vinifera</i> , Espagne	Kerstens <i>et al</i> , 1973  Skerman <i>et al</i> , 1980 Mougel, 2000  Kerstens <i>et al</i> , 1973  Mougel, 2000
G5	CFBP 6625 CFBP 6626	Aliment pour malade Prélèvement humain	Popoff <i>et al</i> , 1983 Kerstens <i>et al</i> , 1973
G6	NCPPB 925 F/1 Zutra	<i>Dahlia</i> sp., Afrique du sud <i>Dahlia</i> sp., Israël	Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973
G7	CIP RV3 NCPPB 1641 Zutra 3/1	<i>Flacourtia indica</i> , Royaume Uni <i>Malus</i> sp., Israël	Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973
G8	ICPB TT9 T37 J-07 6-Mushin LMG 75 LMG 46 AW137 C58	<i>Juglans</i> sp., galle, Californie USA 1cm3 du sol, LCSA France <i>Humulus lupulus</i> , galle, Victoria Australie <i>Euonymus alata</i> cv <i>Compacta</i> , galle, USA <i>Rubus macropetalus</i> , Oregon USA <i>Prunus</i> sp cv Montmorency (cerise), New York USA	Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973 Vogel <i>et al</i> , 2003 Kerstens <i>et al</i> , 1973
G9	o362 o363	Potting soil, Australie Potting soil, Australie	Kerstens <i>et al</i> , 1973 Kerstens <i>et al</i> , 1973
<i>Agrobacterium</i> sp.	NCPPB 1650	<i>Rosa</i> sp., Afrique du sud	
<i>A. larrymoorei</i>	AF 3.10 AF3.44 Fb33	<i>Ficus benjamina</i> <i>Ficus benjamina</i> <i>Ficus benjamina</i> , Salerno, Italie	Bouzar <i>et al</i> , 1995 Bouzar <i>et al</i> , 1995
<i>A. rubi</i> (G11)	LMG 17935 SCRI551	<i>Rubus ursinus</i> var <i>Loganobaccus</i> , USA	
<i>A. vitis</i>	CFBP 5523 CFBP 2736 CFBP 2678	<i>Vitis vinifera</i> <i>Vitis vinifera</i> <i>Vitis vinifera</i>	
<i>Rhizobium rhizogenes</i> (G10)	CFBP 2408 K-84 CP85.120	Pommier <i>Prunus persicae</i> , sol autour d'une galle, Australie <i>P. leuce</i> , galle, France	Ponsonnet, 1994

Tableau 13.

Souches d'*Agrobacterium* de référence utilisées lors des différentes expériences de piégeage des agrobactéries non-pathogènes. LCSA : La Côte Saint André.

## II. 3. 2. Matériel et méthodes

### II. 3. 2. a. Sol modèle

Le sol modèle de La Côte Saint André (LCSA), sous culture de maïs, a été utilisé pour ces expériences de piégeage.

Expériences « campagne1 » et « campagne 2 »

Deux campagnes d'isolement, que nous nommerons dans la suite de ce travail campagne1 et campagne2, ont été réalisées, en mai et en juillet 1998 respectivement, à partir de 300 grammes de sol tamisé et carté. Ces isolement ont été conduits par Christophe Mougel (Mougel, 2000).

Expérience « cm<sup>3</sup> »

La diversité des agrobactéries présente dans un cm<sup>3</sup> de sol a été analysé. Après prélèvement, le cube de sol a été disséqué et les agrobactéries isolés des micro-aggrégats ainsi obtenus. Ce travail a été réalisé par Julia Vogel (Vogel *et al.*, 2003) et par Cathy D'Angelo en 2000.

### II. 3. 2. b. Plantes modèles et conditions de culture

Les expériences de piégeage des agrobacéries par les rhizosphères des plantes modèles ont été conduites en trois essais.

Premier essai : expériences « maïs1 », « lotier » et « morelle»

Cette première série d'expérience a été réalisée par C. Mougel en juillet 1998.

Les graines de maïs (*Zea maïs*, variété DEA) sont directement plantées en microcosme de sol (expérience « maïs1 »). Les plants de morelle (*Solanum nigrum*) et de lotier (*Lotus corniculatus*) ont été fournis aimablement par A. Petit de l'équipe d'Y. Dessaux (ISV, CNRS, Gif-sur-Yvette) et multipliés *in vitro* par micropropagation (expériences « morelle » et « lotier »). Pour chacune des plantes modèles, quatre microcosmes de sol ont été analysés. Les plantes ont été cultivées trente jours avant arrachage.

Deuxième essai : expériences « Luzerne » et « maïs2 »

Ces deux expériences ont été réalisées en Juillet 2003 en collaboration avec C. Mougel.

Les graines de luzerne (*Medicago truncatula*, variété Jemalong J5) désinfectées et scarifiées, ont été semées directement en microcosmes de sol. Les graines de maïs

Expérience	Espèce			
	G1	G4	G8	Total
Campagne 1	17	46	0	63
Campagne 2	68	3	18	89
1cm <sup>3</sup>	47	4	2	53
Total sol	132	53	20	205

Tableau 14.

Répartition en espèce des *Agrobacterium* isolés du sol de LCSA lors des trois campagnes d'isolement.

LCSA: La Côte Saint André

En grisé sont indiqués, par campagne d'isolement, les espèces d'agrobactéries majoritaires.

Espèce	Clade	Campagne	Nombre d'isolats
G1	1	Campagne 1	2
		Campagne 2	1
G4	2	Campagne 1	3
		cm <sup>3</sup>	1

Tableau 15.

Clade-souches regroupant des agrobactéries du sol nu de La Côte Saint André isolées lors de campagne d'isolement différentes.



ont elles aussi été directement plantées en microcosmes de sol. Deux microcosmes pour la luzerne et quatre pour le maïs ont été étudiés.

Lors de ces expériences, deux agrobactéries du sol nu ont été isolées et analysées simultanément aux isolats des rhizosphères.

Troisième essai : expérience « peuplier »

Les plants de peuplier de la section *Leuce* (*Populus alba* X *P. tremula*, clone 717-1b4), multipliés in vitro nous ont été fournis par Gilles Pilate d'Orléans. Cette expérience a concerné trois microcosmes cultivés avec les peupliers pendant huit mois. L'arrachage des plants pour analyse à été fait en Mars 2004.

### *II. 2. 3. c. Souches bactériennes*

Isolats bactériens

Nous avons analysé les isolats issus des sols rhizosphériques (SR) et des tissus racinaires (TR) de tous les microcosmes étudiés. Au total nous avons analysé 205 isolats du sol et 341 isolats de rhizosphères.

Souches de références

Les analyses AFLP ont été menées simultanément sur l'ensemble de souches de référence (Tableau 13).

### *II. 2. 3. d. Traitement des isolats*

Isolement

Les souches bactériennes ont été isolées selon la procédure décrite précédemment (Partie II, Chapitre 2).

Les *Agrobacterium* ont été isolés du sol de LCSA lors de deux campagnes (campagne1 et campagne2), respectivement en mai et en juillet 1998 à partir de 300 grammes de sol tamisé.

L'analyse de la répartition des *Agrobacterium* dans ce sol à microéchelle à été étudiée à partir d'1 cm<sup>3</sup> de sol (expérience 1cm<sup>3</sup>) par Vogel *et al.* (2003).

Les *Agrobacterium* ont été isolés de deux compartiments, sol rhizosphérique (SR) et tissus racinaires (TR), des rhizosphères des plantes modèles comme définit par C. Mougel (2000).

Purification, culture et conservation des isolats, tests biochimiques

Expérience	Espèce	G4	G8	Total
	G1			
Lotier	59	10	0	69
Morelle	70	0	0	70
Maïs1	72	0	0	72
Maïs2	51	22	0	73
Luzerne	4	0	20	24
Peuplier	12	21	0	33
Total	268	53	20	341

Tableau 16.

Effectifs des *Agrobacterium* isolés de rhizosphères de différentes plantes modèles cultivées dans le sol de LCSA.

LCSA: La Côte Saint André

En grisé sont indiqués, par campagne d'isolement, les espèces d'agrobactéries majoritaires.

Plante	G1		G4		G8		Nb microcosmes
	SR	TR	SR	TR	SR	TR	
Lotier	A+B+C+D	A+B+C+D	A+B	0	0	0	4
Morelle	A+B+C+D	A+B+C+D	0	0	0	0	4
Maïs1	A+B+C+D	A+B+C+D	0	0	0	0	4
Maïs2	2+3+4	1+2+3+4	1+3	1+3+4	0	0	4
Luzerne	2	2	0	0	1	1+2	2
Peuplier	1+3	1+3	2+3	1+2+3	0	0	3

Tableau 17.

Noms des microcosmes en fonctions des espèces bactériennes isolées des tissus racinaires (TR) et du sol rhizosphérique (SR) de diverses plantes cultivées dans le sol de La Côte Saint André.

Les isolats ont été purifiés cultivés et conservés selon les procédures décrites au chapitre précédent. De même la procédure utilisée pour les tests biochimiques est décrite au chapitre précédent.

#### Recherche de formes pathogènes

La recherche de formes pathogènes (possédant le plasmide Ti) parmi les isolats de sol et de rhizosphères a été menée par hybridation sur colonies. Les colonies ont été hybridées avec les sondes *tmr*, *nos* et l'intergène *virB-virG*, spécifiques du plasmide Ti selon le protocole décrit par Mougel *et al.* (2001).

#### *II. 3. 2. e. Analyses moléculaires et analyses phylogénomiques*

Les analyses moléculaires et phylogénomiques ont été menées comme décrites précédemment. Les profils AFLP des isolats du sol et des expériences de l'essai 1 ont été révélés sur séquenceur à gel 373A (Perkin Elmer). Les profils AFLP des isolats des essais 2 et 3 ont été révélés sur séquenceur automatique à capillaire Megabace 1000 (Amersham Pharmacia Biotech Europe, Orsay, France).

Dans la mesure du possible, les données AFLP obtenues par différents amorçages indépendants ont été compilées.

#### Analyses statistiques

La signification statistique de la structure des populations a été analysée par Analyse Moléculaire de Variance (AMOVA) avec le logiciel Arlequin (Schneider *et al.*, 2000 (Arlequin)).

#### *II. 3. 2. f. Séquençage du gène rrs*

L'espèce génomique de quelques isolats a été confirmée par séquençage du gène *rrs* de quelques souches représentatives. Le gène *rrs* est amplifié par PCR avec les amorces F667pA (AGAGTTTGATCCTGGCTCAG) et F668pH (AAGGAGGTGATCCAGCCGCA) (Bruce *et al.*, 1992). Si nécessaire, les produits PCR ont été purifiés, par élution à partir du gel d'agarose, avec le « Gel extraction kit » (Qiagen Hilden, Allemagne), selon les recommandations du fournisseur. Les produits PCR ont été séquencés par la société GenomeExpress (Meylan, France).



Figure 13: Arbre phylogénomique des souches de références et des isolats du sol nu de La Côte Saint André appartenant à l'espèce *Agrobacterium* G1. Les valeurs de bootstrap (calculées pour 100 rééchantillonnages) significatives sont indiquées. LCSAB002 à LCSAB079: isolats de la campagne 1  
 LCSAB080 à LCSAB167: isolats de la campagne 2  
 J2 à J90: isolats du cm3 de sol (les chiffres et lettres suivant le nom de l'isolat correspondent à leur coordonnées dans le cm3 de sol).

## II. 3. 3. Résultats

### II. 3. 3. a. Analyse des populations du sol

Les trois campagnes d'isolement effectuées dans le sol de La Côte Saint André n'ont pas permis d'isoler des agrobactéries pathogènes ni des agrobactéries de l'espèce *Rhizobium rhizogenes*.

### IDENTIFICATION DES ESPECES GENOMIQUES

L'analyse des 205 isolats d'*Agrobacterium* du sol a montré qu'ils se répartissent entre les espèces G1, G4 et G8 (Tableau 14). Les hybridations sur colonies avec les sondes spécifiques du pTi n'ont donné aucun résultat positif.

Les 63 *Agrobacterium* isolés lors de la première campagne appartiennent aux espèces génomiques G1 (27%) et G4 (77%).

Les 89 agrobactéries isolées lors de la deuxième campagne appartiennent aux espèces G1 (77%), G4 (3%) et G8 (20%).

Les 53 agrobactéries isolées du cm<sup>3</sup> de sol appartiennent aux espèces G1 (88%), G4 (8%) et G8 (4%) (Tableau 14).

### ANALYSES INTRA-SPECIFIQUES

Au niveau intra spécifique, les isolats se regroupent en clades-souches soutenus par des bootstraps significatifs (Figures 13, 14 et 15). La plupart de ces clades-souches regroupent des isolats provenant de la même campagne d'isolement. Toutefois, certains clades regroupent des individus isolés lors de campagnes différentes (clades 1 et 2, Figures 13 et 14).

Dans l'espèce G1 le clade 1, soutenu à 87%, regroupe deux isolats de la campagne 1 avec un isolat de la campagne 2. Dans l'espèce G4, le clade-souche 2, soutenu à 100%, regroupe un isolat du cm<sup>3</sup> de sol et deux isolats de la campagne 1 (Tableau 15).

### II. 3. 3. b. Analyse des populations piégées par les rhizosphères

Aucun isolat obtenu sur le milieu 2E+Te, à partir des rhizosphères cultivées dans le sol de La Côte Saint André, n'a été confirmé comme appartenant à l'espèce *Rhizobium rhizogenes* par séquençage du gène *rrs*.

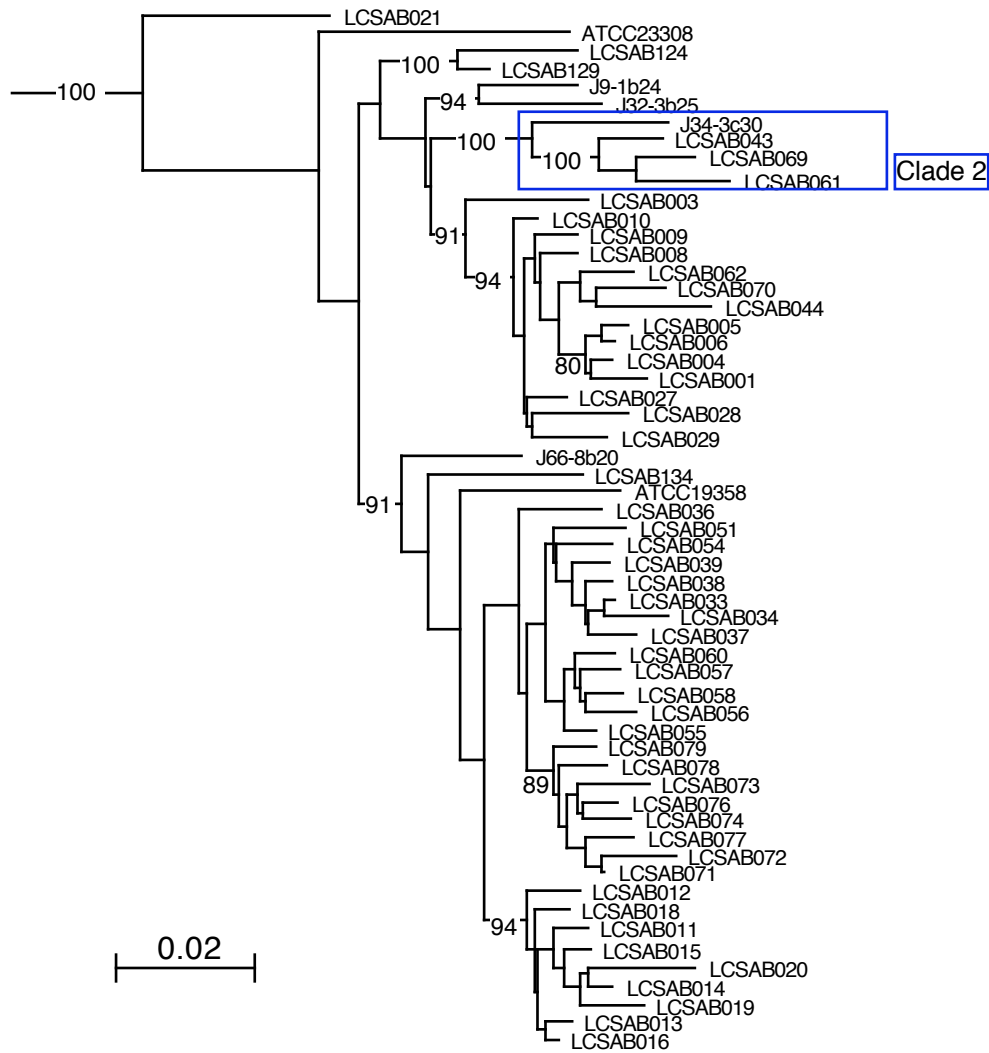


Figure 14:  
 Arbre phylogénomique des souches de références et des isolats du sol nu de La Côte Saint André appartenant à l'espèce *Agrobacterium* G4.  
 Les valeurs de bootstrap (calculées pour 100 rééchantillonnages) significatives sont indiquées.  
 LCSAB001 à LCSAB079: isolats de la campagne 1  
 LCSAB080 à LCSAB167: isolats de la campagne 2  
 J9 à J66: isolats du cm<sup>3</sup> de sol (les chiffres et lettres suivant le nom de l'isolat correspondent à leur coordonnées dans le cm<sup>3</sup> de sol).

## IDENTIFICATION DES ESPECES GENOMIQUES (TABLEAUX 16 ET 17)

### ✦ Essai 1

Suite aux analyses réalisées sur le séquenceur Perkin-Elmer, un grand nombre d'isolats apparaissaient proches de l'espèce G1 mais sans que ce regroupement soit validé par des bootstraps significatifs (voir flèche, figure 16). Nous avons refait des analyses AFLP sur séquenceur Amersham avec quelques isolats. De plus le gène *rrs* de certains de ces isolats a été séquencé. Ces deux approches confirment que ce grand groupe appartient bien à l'espèce G1 (Figure 17).

Les isolats issus de l'essai 1 appartiennent majoritairement à l'espèce G1. Seul 10 isolats du sol rhizosphérique de lotier, issus de deux microcosmes différents appartiennent à l'espèce G4 (Figure 16, Tableaux 16 et 17).

### ✦ Essai 2

51 (70%) isolats issus de l'expérience maïs2 appartiennent à l'espèce G1. 22 (30%) autres isolats appartiennent à l'espèce G4 (Figure 18, Tableau 16).

Les isolats de l'espèce G4 proviennent de trois microcosmes,. Ceux de l'espèce G1 proviennent eux des quatre microcosmes (Tableau 17).

Cinq isolats (20%) (tous du microcosme 2) issus des rhizosphères de luzerne appartiennent à l'espèce G1 et les 20 derniers isolats (80%) appartiennent à l'espèce G8 (Figure 18, Tableaux 16 et 17).

### ✦ Essai 3

Douze (36%) de ces isolats appartiennent à l'espèce G1 et les 21 derniers (64%) appartiennent à l'espèce G4 (Figure 19, Tableaux 16 et 17). L'espèce G1 regroupe des isolats issus de deux microcosmes et l'espèce G4 regroupe des isolats provenant des trois microcosmes (Tableau 17).

## ANALYSES INTRA-SPECIFIQUES

### ✦ 1) Analyses phylogénomiques

Au niveau intra-spécifiques les isolats se regroupent en clades. Nous présenterons ici les clades pertinents, qui regroupent des isolats d'origine (plante, compartiment ou microcosme différents) (Tableau 18).

### ★ Essai 1

Parmi les isolats de l'espèce G1, ceux-ci se différencient en fonction des plantes et des microcosmes (Figure 16, Tableau 18).

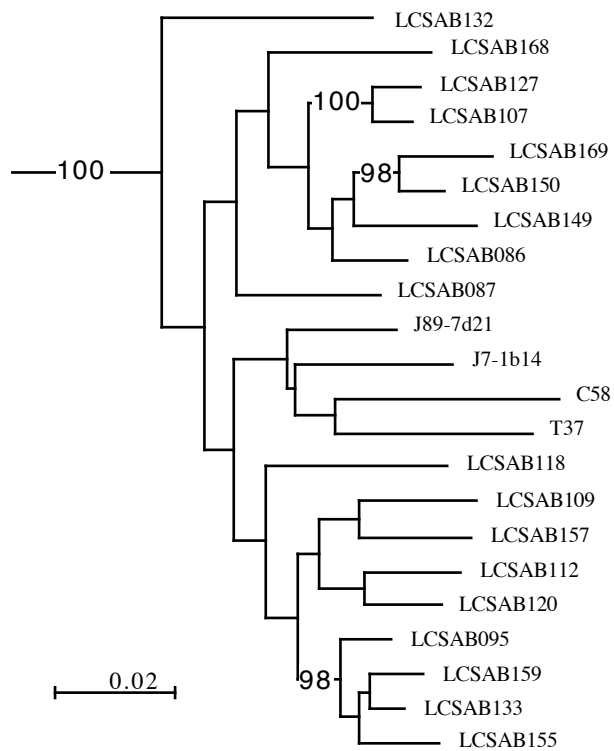


Figure 15:  
 Arbre phylogénomique des souches de références et des isolats du sol nu de La Côte Saint André appartenant à l'espèce *Agrobacterium G8*.  
 Les valeurs de bootstrap (calculées pour 100 rééchantillonnages) significatives sont indiquées.  
 LCSAB080 à LCSAB167: isolats de la campagne 2  
 J7 et J89: isolats du cm<sup>3</sup> de sol (les chiffres et lettres suivant le nom de l'isolat correspondent à leur coordonnées dans le cm<sup>3</sup> de sol).



Les isolats du sol rhizosphérique de lotier se répartissent entre les clades 4, 6 et 8. Le clade 6 regroupe en outre un isolat du sol rhizosphérique de morelle. Les isolats des tissus racinaires de lotier se regroupent avec les isolats issus des quatre microcosmes et des deux compartiments (SR et TR) du maïs (Tableau 18).

Les isolats de morelle forment deux clades significatifs. L'un, le clade 5, regroupant les isolats des tissus racinaires, et l'autre, le clade 7, regroupant les isolats du sol rhizosphérique.

Les isolats de l'espèce G4 piégés par le sol rhizosphérique du lotier sont divers. 7 se regroupent en un clade, les trois autres étant plus distant génétiquement.

La nouvelle analyse AFLP (conditions Eco+CA/Mse+0) montre que des isolats provenant de différents microcosmes se regroupent significativement en clades (Figure 17). Ces clades sont vérifiés par une deuxième analyse AFLP indépendante (conditions Eco+CG/Mse+0).

Le clade 10 regroupe des isolats de maïs, des deux compartiments (SR et TR), issus des microcosmes A et C. Le clade 11 regroupe des isolats des tissus racinaires issus du microcosme D. L'analyse en conditions Eco+Cg/Mse+0, montre qu'un isolat du microcosme B se regroupe avec ceux-ci. Enfin le clade 12 regroupe des isolats des tissus racinaires des microcosmes A et B (résultats non présentés).

Les isolats des rhizosphères de maïs et de morelle apparaissent génétiquement éloignés, dans cette espèce, des souches de référence.

#### ★ Essai 2

Dans cet essai les isolats ne se différencient pas en fonction du compartiment d'origine.

Le clade 13 regroupe tous les isolats issus des rhizosphères de luzerne et appartenant à l'espèce G8 (Tableau 18).

Dans l'espèce G1, le clade 14, regroupe un isolat du sol rhizosphérique du maïs et un isolat issus des tissus racinaires de morelle isolé lors de l'essai 1. Le clade 15, en plus de regrouper des isolats issus de trois des microcosmes du maïs, regroupe un des deux isolat du sol nu. Le clade 16 regroupe lui des isolats de trois microcosmes de maïs et les cinq derniers isolats piégés par la luzerne.

Les isolats de l'espèce G4 piégés par le maïs se regroupent dans le clade 17 et proviennent de trois microcosmes différents (Tableau 18).

#### ★ Essai 3

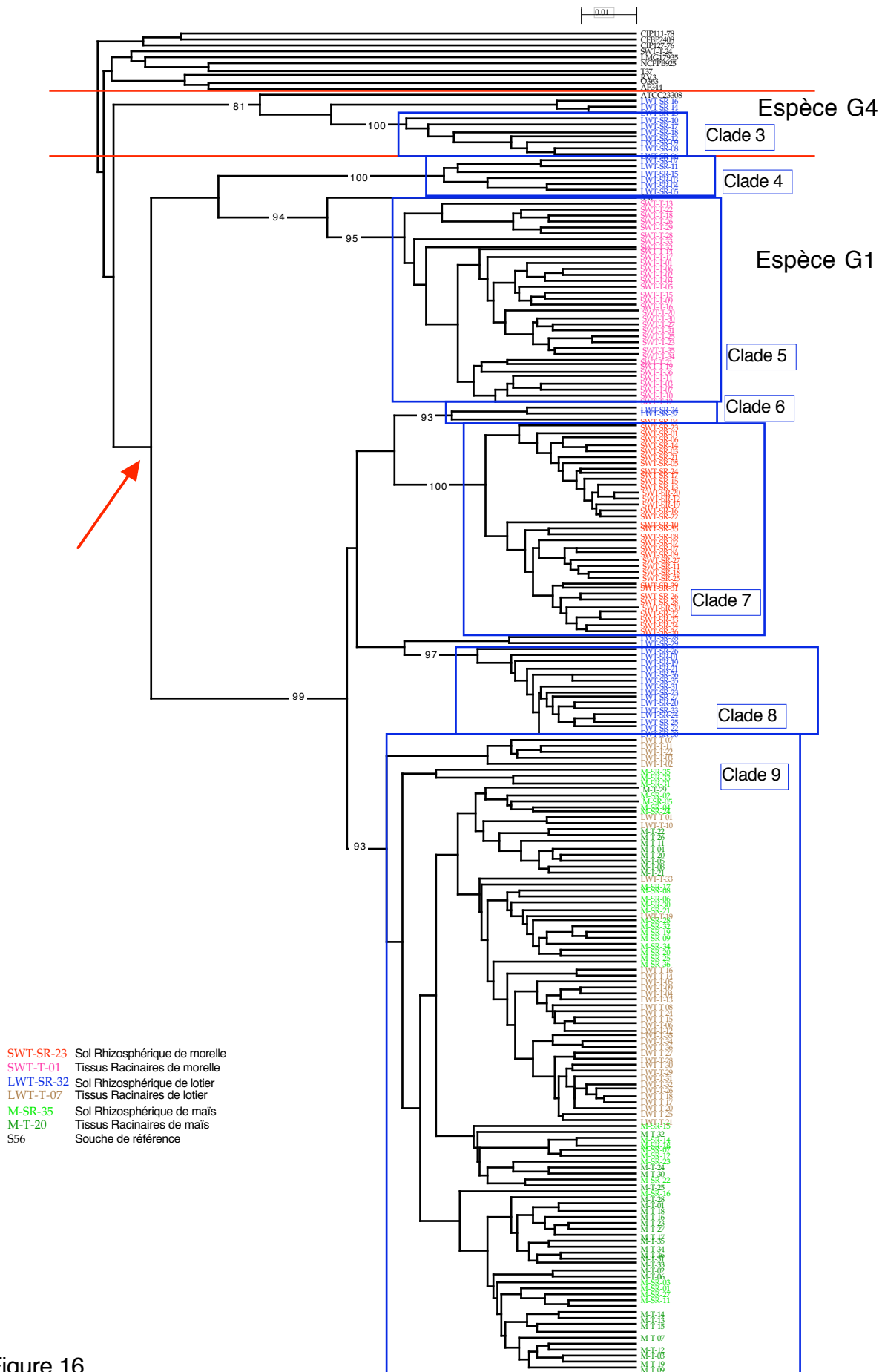


Figure 16 .  
 Arbre phylogénomique des isolats piégés par les rhizosphères de maïs, morelle et lotier cultivés dans le sol de La Côte Saint André (essai 1). Calculé à partir des données de deux conditions d’AFLP indépendantes.  
 Les valeurs significatives de bootstrap, pour 100 rééchantillonnages, sont données.

Dans l'espèce G4, les clades 18 et 19 regroupent chacun des isolats de compartiments et de microcosmes différents (Tableau 18).

#### ✦ 2) AMOVA

Les données AFLP traitées par analyse phylogénomiques ont également été traitées par Analyse MOléculaire de VAriance (au niveau infra-spécifique) afin de vérifier si la séparation en fonction des microcosmes, compartiments et plantes-hôtes modifiait significativement la structure génétique de la population. L'AMOVA n'a pu être utilisée que sur l'espèce G1, et sans pouvoir comparer les essais 1 et 2 réalisés sur des séquenceurs différents.

Les analyses par AMOVA montrent qu'il n'y a pas, pour chaque plantes, de différences significatives entre les microcosmes analysés. Par contre, il existe des différences significatives entre les compartiments (SR ou TR) sauf pour le maïs. Pour cette dernière plante-hôte les populations provenant des deux compartiments ne se différencient pas.

### **II. 3. 4. Discussion**

Lors de ce travail nous voulions déterminer si les différentes plantes modèles étudiées exercent une sélection différentielle dirigée vers les agrobactéries du sol. Afin de s'affranchir des contingences liées à la disponibilité des agrobactéries dans ce sol, nous avons mené les différents essais dans un sol de culture modèle situé à La Côte Saint André. Ceci permet de travailler avec des populations naturelles, qui présentant la diversité microbienne réelle d'un sol. En effet, les études menées sur les épidémies touchant le peuplier et le GF677 (Partie II, Chapitres I et II) montrent que la sélection observée des agrobactéries par la plante est nécessairement contrainte par les agrobactéries disponibles dans le sol.

Un résultat remarquable est l'absence de *R. rhizogenes* tant dans le sol que dans les rhizosphères. Ceci confirme que la population des agrobactéries présente dans les rhizosphères des plantes est dépendante des agrobactéries disponibles dans le sol. Nous n'avons pas non plus mis en évidence de formes pathogènes. Ceci peut s'expliquer par le fait que cette parcelle de terrain n'a jamais été cultivé avec des plantes contaminées qui auraient éventuellement pu favoriser l'installation d'une population pathogène.

Le fait de travailler avec une population naturelle, nécessite d'en connaître la diversité. Bien que les méthodes d'échantillonnage (tamisage, cartage) assurant une

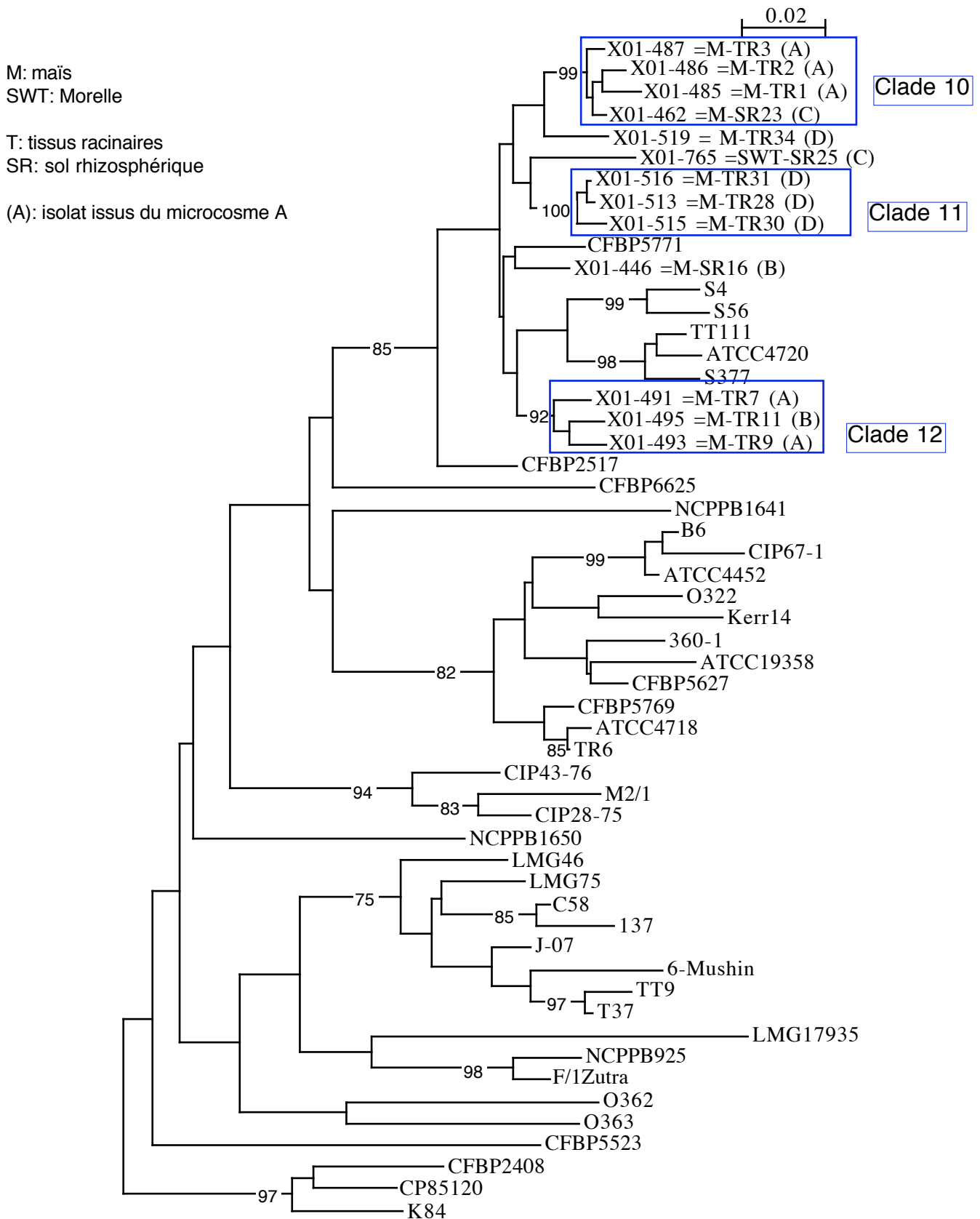


Figure 17.  
Arbre phylogénomique d'agrobactéries piégées par les rhizosphères de maïs et morelle cultivés dans le sol de La Côte Saint André (essai 1), confirmant leur appartenance à l'espèce G1.  
Condition AFLP: Eco+CA/Mse+0.  
Les valeurs de bootstrap significatives, et confirmées par une deuxième analyse AFLP, (conditions Eco+CG/Mse+0), sont indiquées.

reproductibilité maximale des analyses minérales aient été utilisées, il y a des différences significatives dans la diversité des agrobactéries du sol de La Côte Saint André en fonction des essais. La diversité est telle qu'il est très rare d'isoler des agrobactéries appartenant à la même souche lors de deux campagnes d'isolement différentes. De plus, certaines espèces ne sont pas toujours isolées.

La cause de cette variabilité pourrait être imputable aux sites de prélèvements qui pourraient être légèrement différents dans la parcelle expérimentale. Cependant, l'étude de la répartition à microéchelle des agrobactéries (Vogel et al., 2003) montre qu'à l'échelle du cm<sup>3</sup> toute la diversité spécifique observable des agrobactéries de La Côte Saint André est déjà présente. Une autre explication reposerait sur les dates de prélèvement des échantillons au champs. En effet la première campagne a été réalisée en mai et la deuxième en juillet. Krimi *et al* (2003), ont rapporté une variabilité significative des populations d'agrobactéries dans les sols en fonction des saisons. On peut se demander si ces fluctuations n'affecteraient pas également la composition spécifique de la communauté des agrobactéries.

La variabilité de la population des *Agrobacterium* du sol de La Côte Saint André, d'une campagne à l'autre montre que travailler à partir d'un même substrat de culture ne suffit pas pour être assuré de travailler avec une population bactérienne homogène. Nous ne pouvons vraiment comparer entre que les populations isolées lors d'un même essai, donc *a priori* les agrobactéries piégées à partir d'une population présentant une structure identique.

Les expériences morelle, lotier et maïs<sup>1</sup>, qui font partie du même essai, sont donc comparables. Les populations maïs<sup>2</sup> et luzerne sont comparables pour les mêmes raisons.

Les populations d'agrobactéries isolées lors de l'essai 1 appartiennent majoritairement à l'espèce G1, à l'exception de quelques agrobactéries provenant du sol rhizosphérique de lotier et isolées dans deux microcosmes différents. L'effet de piégeage d'agrobactéries de cette espèce dans deux microcosmes indépendants montre que le lotier est significativement associé à l'espèce G4.

Par contre, les isolats de ce compartiments se répartissent en deux espèces et quatre clades différents, alors que tous les isolats des tissus racinaires appartiennent au même clade. Ceci montre que le sol rhizosphérique est probablement moins sélectif que les tissus racinaires où doivent s'établir des associations plus étroites entre plantes et bactéries.



Les isolats du maïs appartiennent tous au même clade. Or, lors de l'arrachage les racines de maïs se sont révélées quasi nues sans sol adhérent. Il est donc possible que le sol rhizosphérique pour cette espèce de plante monocotylédone soit quasi inexistant, ou laissé en place lors de la collecte des échantillons. Ceci expliquerait l'homogénéité entre les deux compartiments alors artificiellement distingués.

La morelle est la plante montrant l'association la plus spécifique avec les agrobactéries. Les isolats se regroupent dans des clades bien différenciés selon leur compartiment d'origine.

L'analyse de ces population sur séquenceur Perkin-Elmer a montré, que bien qu'appartenant à l'espèce G1, les isolats des rhizosphères ne se regroupaient pas de façon significative avec les souches de référence de cette espèce. Plusieurs faits peuvent expliquer ce phénomène : 1), le nombre d'isolats analysés. En effet, l'analyse d'un aussi grand nombre d'isolats homogènes entre eux peut artificiellement augmenter la distance génétique entre d'un côté les isolats et de l'autre les souches de référence. D'autant plus, que les nouvelles analyses AFLP sur certains de ces isolats montrent qu'ils sont génétiquement éloignés des souches de référence de l'espèce G1, c'est l'hypothèse la plus probable (Figure 7) ; 2), la méthode phylogénétique utilisée. En effet la méthode de calcul utilisée est l'UPGMA. La méthode Neighbor-Joining est plus indiquée dans ce cas. Mais elle prend aussi nettement plus de temps de calcul. Il est donc fastidieux de se livrer aux calculs de bootstraps sur un aussi grand nombre de souches ; 3), les analyses AFLP. Celles-ci n'ont pas été dupliquées pour ces isolats. De fait, si des variations dues à l'analyse au séquenceur (un gel qui présente un petit défaut, qui migre moins bien par exemple) sont présentes, la méthode d'analyse est suffisamment sensible pour traduire de petites variations dans l'analyse d'un gel.

L'essai 1 nous a permis de montrer un effet sélectif de la plante dirigé vers certaines agrobactéries. Nous avons voulu confirmer ces effets par de nouvelles expériences de piégeage avec d'autres plantes modèles cultivées dans le sol modèle. Dans un deuxième essai nous avons donc étudié le piégeage des agrobactéries par le maïs et la luzerne.

Parmi les populations piégées par le maïs lors de cet essai 2, il n'y a toujours aucune différence entre les compartiments analysés. Par contre nous observons la présence de nombreuses agrobactéries appartenant à l'espèce G4 dans ces rhizosphères.

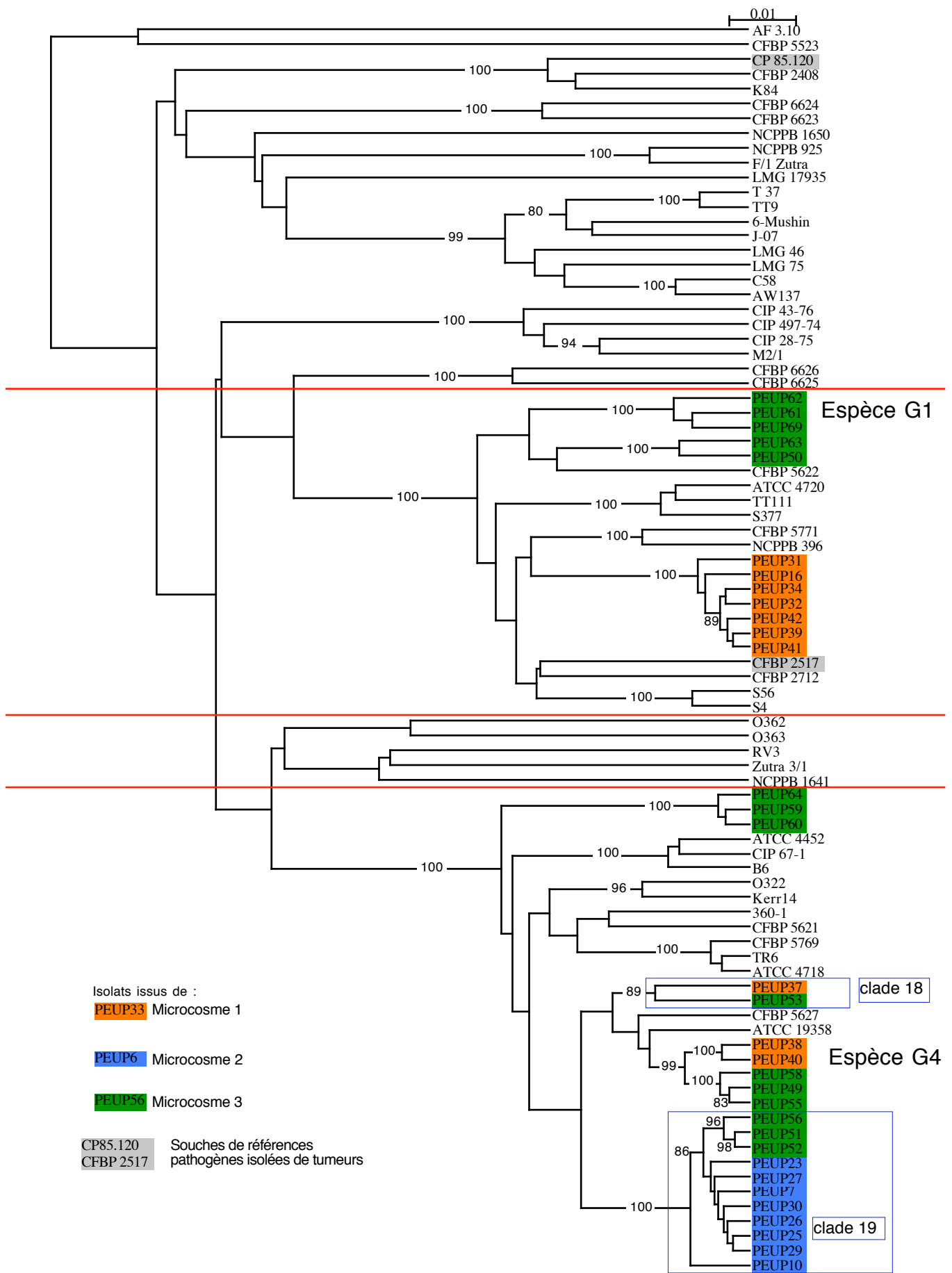


Figure 19: Arbre phylogénomique des isolats d'*Agrobacterium* piégés par les rhizosphères de peuplier (*Populus alba* X *P. tremula*) cultivés dans le sol de La Côte Saint André (essai 3), calculé à partir de deux analyses AFLP indépendantes. les valeurs de bootstrap significatives (calculées pour 100 rééchantillonnages) sont indiquées.



Parallèlement au maïs, nous avons étudié l'effet des rhizosphères de luzerne. Dans ces rhizosphères nous avons observé un piégeage différentiel. En effet, les rhizosphères de luzerne piègent spécifiquement des agrobactéries de l'espèce G8. ces isolats proviennent des deux microcosmes analysés, montrant qu'il ne s'agit probablement pas d'une rencontre au hasard entre les bactéries et les racines. Ceci est remarquable car c'est la première fois que nous piégeons des agrobactéries de l'espèce G8. Or, c'est la recherche de telles bactéries qui nous avait, entre autre, motivé pour l'étude du cultivar GF677. En effet la souche C58, souche type de l'espèce G8, qui est entièrement séquencée, à été isolée d'un plant de cerisier, une plante ligneuse du genre *Prunus*, tout comme le cultivar GF677. Il est apparu que les agrobactéries associées au GF677 ne sont pas particulièrement associées aux agrobactéries de l'espèce G8. La découverte d'une plante piégeant les agrobactéries de cette espèce nous permettraient une analyse plus complète, car c'est à cette espèce qu'appartient C58, la souche d'*Agrobacterium* entièrement séquencée pour laquelle des analyses génomiques poussées peuvent être réalisées.

Les résultats de cet essai ne nous ont pas permis de démontrer la sélection par le maïs des agrobactéries de l'espèce G1. Par contre, ils montrent la sélection, par les rhizosphères de luzerne, d'agrobactéries de l'espèce G8.

Malgré l'abondance des agrobactéries de l'espèce G1 dans le sol, ce sont les agrobactéries de l'espèce G4 qui sont majoritairement isolées des rhizosphères de peuplier. Ce résultat ne va pas dans le sens de ce qui a été observé dans les tumeurs (Partie II, Chapitre I). Ces résultats réfutent donc notre hypothèse selon laquelle le peuplier sélectionne essentiellement des agrobactéries de l'espèce G1. Néanmoins, ces plantes sélectionnent significativement des agrobactéries de l'espèce G4.

### **II. 3. 5. Conclusion**

Ces travaux montrent la diversité des agrobactéries présentes dans le sol modèle de La Côte Saint André. Ils montrent, une variabilité des populations en fonction des prélèvements de sol. La conséquence majeure de ce fait est qu'il n'est possible que de comparer des expériences réalisées simultanément. Un moyen pour s'affranchir de cette variabilité serait de travailler en milieu gnotobiotique, en microcosmes de sol stériles recolonisés par une population d'agrobactéries contrôlée.

espèce	clade	essai	plante	compartiment	microcosmes	Nb isolats
G1	4	1	Lotier	SR	A + B	6
	5	1	Morelle	TR	A + B + C + D	34
	6	1	Lotier	SR	D	2
			Morelle	SR	A	1
	7	1	Morelle	SR	A + B + C + D	35
	8	1	Lotier	SR	C + D	15
	9	1	Lotier	TR	A + B + C + D	35
			Maïs 1	SR + TR	A + B + C + D	70
	14	2	Maïs 2	SR	4	1
CFBP5771			Morelle TR		1	
15	2	Maïs 2	SR + TR	2 + 3 + 4	22	
		Sol Nu			1	
16	2	Luzerne	SR + TR	2	5	
		Maïs 2	SR + TR	1 + 2 + 4	20	
G4	3	1	Lotier	SR	A + B	7
	17	2	Maïs 2	SR + TR	1 + 2 + 4	22
	18	3	Peuplier	SR + TR	1 + 3	2
	19	3	Peuplier	SR + TR	2 + 3	11
G8	13	2	Luzerne	SR + TR	1 + 2	20

Tableau 18.

Clades regroupant des isolats provenant d'expériences de piégeage différentes dans le sol de LCSA

TR : Tissus Racinaire

SR : Sol Rhizosphérique

LCSA : La Côte Saint André

Ces travaux nous ont permis d'observer la sélection de certaines agrobactéries par les plantes. Par contre, ils montrent que cette sélection n'est pas exclusive, il s'agit plus probablement d'une association préférentielle. En outre, l'hypothèse selon laquelle il y a un lien entre agrobactéries sélectionnées dans les tumeurs et agrobactéries sélectionnées par les rhizosphères n'a pas été confirmée.

## II. 4. Conclusion

Ce travail de thèse confirme la puissance de la méthode AFLP pour la détermination des espèces génomiques et la recherche de clades assimilables aux des souches bactériennes au sens de Tenover (Tenover *et al.*, 1995) ; ce qui en fait un outil d'analyse épidémiologique puissant dans le cas d'une maladie comme la Galle du Collet caractérisée par une très grande diversité spécifique et infra-spécifique des agents pathogènes.

Les différentes analyses présentées dans cette partie suggèrent qu'il pourrait avoir une association préférentielle entre une espèce d'agrobactéries d'un coté et une espèce végétale de l'autre.

Cette association dépendrait à la fois des agrobactéries présentes dans le substrat de culture et de la plante.

Il ne s'agit pas d'une sélection exclusive, telle que celle existant entre les bactéries du genre *Rhizobium* et les légumineuses, mais une association préférentielle.

Cette association est souvent dirigée vers des souches présentant une diversification radiale (non hiérarchisée) représentatives de la diversité globale de cette espèce. Les plantes sont donc associées à des espèces et non pas des souches particulières de ces espèces. La diversité des espèces les plus fréquemment observées ici (G1, G4 et *R. rhizogenes*) correspond à celle d'écotypes tels que décrits par Cohan (Cohan, 2001, 2002).

Les écotypes se différencient sur la base de critères écologiques. Ainsi, les espèces d'agrobactéries, étant des écotypes associés à des plantes particulières, il est possible que les plantes aient joué un rôle déterminant pour la spéciation dans ce genre bactérien.

Une telle étude, par approche de biologie des populations, permet d'obtenir des informations précieuses sur l'évolution des populations étudiées. Mais elle reste malgré tout lourde à mettre en œuvre. De plus l'interprétation des résultats n'en est pas toujours aisée.

Ce type d'étude nous montre que la plante pourrait être un facteur prépondérant dans l'évolution du genre *Agrobacterium*.

Afin de confirmer ceci nous avons utilisé une autre approche, basée sur les données des analyses AFLP obtenues sur les souches de référence combinées à des analyses génomiques et bioinformatiques, rendues possibles par l'achèvement en

2001 du séquençage de la souche de référence C58 (Wood *et al.*, 2001 ; Goodner *et al.*, 2001).

C'est cette approche novatrice, qui nous permet de déterminer les gènes et fonctions écologiques impliquées dans la différenciation en espèces dans le genre *Agrobacterium* que nous exposerons dans la troisième partie de ce travail de thèse.

## Partie III

Demasking genome regions involved in *Agrobacterium*  
species delineation using experimental and predicted  
AFLP fragments

# **III. Demasking genome regions involved in *Agrobacterium* species delineation using experimental and predicted AFLP fragments**

## **III. 1. Introduction**

The current consensus for a bacterial species based on the genomic cohesion species concept (Wayne *et al.*, 1987) has been recently confirmed by the *ad hoc* committee for the definition of a bacterial species (Stackebrandt *et al.*, 2002). In addition, the committee indicated that the delineation of genomic species can now be obtained with alternative molecular methods that sample parts of the genome instead DNA-DNA hybridization of whole genomes alone. Among the alternative genome sampling methods, the AFLP (Vos *et al.*, 1995) allows one to very easily analyse and compare AFLP profiles of any unknown isolates to data banks of AFLP profile of reference strains. This allows in turn the accurate assignation of all isolates to genomic species (Mougel *et al.*, 2002), or, at least, it reveals that isolates which do not cluster with reference strains likely belong to novel species not yet included in the data bank (Part II). The genomic species identification ability of AFLP is due to the fact that some AFLP fragments are common to all members of a given genomic species. Those common AFLP fragments are genomic markers of the species.

In addition to the confirmation of the genomic species concept, Stackebrandt *et al.* (2002) strongly encourage researches dealing with the elucidation of what is actually and biologically a genomic species. In this idea, the ecological species concept, proposed by Cohan in 2001, explains why there is gaps in the distribution of the genomic diversity of bacteria. The occurrence of such a gap in the range 50-70% RBR values is the phenomenon which permits the unambiguous recognition of genomic species as clusters of strains with high genome similarities (Grimont, 1988). For Cohan, bacterial species are ecotypes defined with respect to the fate of adaptative mutants. An ecotype is a set of strains using the same or very similar ecological niches, such that an adaptative mutant from within the ecotype outcompete to extinction all other strains of the same ecotype. This will purge the diversity within and not between the ecotypes. For Cohan, a bacterial species is

Species	Strain name	Biological and geographical origin	Référence
G1	ATCC 4720	Rapsberry, USA	Kerstens <i>et al</i> , 1978  Popoff <i>et al</i> , 1983 Popoff <i>et al</i> , 1983 Popoff <i>et al</i> , 1983 Mougel, 2000 Ponsonnet, 1994 Mougel, 2000 Ridé Ponsonnet, 1994
	TT111	Gall, USA	
	NCPPB 396	<i>Dahlia</i> sp., Germany	
	S-377	Plant	
	S-56	plant	
	S-4	plant	
	CFBP 5622	<i>Solanum nigrum</i> , root tissus, LCSA, France	
	CFBP 2517	<i>Populus leuce</i>	
	CFBP 5771	Bulk soil, LCSA France	
	CFBP 2712	GF677, France	
CP85.6	<i>P. leuce</i> , gall, France		
G2	M 2/1	Drain water, Belgium	Kerstens <i>et al</i> , 1978 Keridjian, 1979 Keridjian, 1979 Keridjian, 1979
	CIP 497-74	Human	
	CIP 28-75	Human	
	CIP 43-76	Human	
G3	CFBP 6623	Antiseptic flask	Popoff <i>et al</i> , 1983 Keridjian, 1979
	CFBP 6624	Human	
G4	B6	Apple tree	Kerstens <i>et al</i> , 1978   Skerman <i>et al</i> , 1980 Mougel, 2000  Kerstens <i>et al</i> , 1978  Mougel, 2000
	ATCC 4452	<i>Rubus idaeus</i>	
	ATCC 4718	USA	
	Kerr 14	<i>Prunus dulcis</i> , south Australia	
	O322	<i>Prunus persica</i> , gall on stem, Australia	
	ATCC 19358		
	CFBP 5621	<i>Lotus corniculata</i> , root tissus, LCSA France	
	CFBP 5769	<i>Librocedrus</i> sp., gall	
	ICPB TR6		
CIP 67-1	Other name for B6 strain		
CFBP 5627	Sol nu, LCSA France		
360-1	<i>Vitis vinifera</i> , Spain		
G5	CFBP 6625	Patient food	Popoff <i>et al</i> , 1983 Kerstens <i>et al</i> , 1978
	CFBP 6626	Human	
G6	NCPPB 925	<i>Dahlia</i> sp., South Africa	Kerstens <i>et al</i> , 1978 Kerstens <i>et al</i> , 1978
	F/1 Zutra	<i>Dahlia</i> sp., Israël	
G7	CIP RV3		Kerstens <i>et al</i> , 1978 Kerstens <i>et al</i> , 1978 Kerstens <i>et al</i> , 1978
	NCPPB 1641	<i>Flacourtia indica</i> , United Kingdom	
	Zutra 3/1	<i>Malus</i> sp., Israël	
G8	ICPB TT9		Kerstens <i>et al</i> , 1978 Kerstens <i>et al</i> , 1978 Vogel <i>et al</i> , 2003 Kerstens <i>et al</i> , 1978
	T37	<i>Juglans</i> sp., gall, California USA	
	J-07	1cm <sup>3</sup> of bulk soil, LCSA France	
	6-Mushin	<i>Humulus lupulus</i> , gall, Victoria Australia	
	LMG 75	<i>Euonymus alata</i> cv <i>Compacta</i> , gall, USA	
	LMG 46	<i>Rubus macropetalus</i> , Oregon USA	
	AW137		
	C58	<i>Prunus</i> sp cv Montmorency (cherry), New York USA	
G9	o362	Potting soil, Australia	Kerstens <i>et al</i> , 1978 Kerstens <i>et al</i> , 1978
	o363	Potting soil, Australia	
<i>Agrobacterium</i> sp.	NCPPB 1650	<i>Rosa</i> sp., South Africa	
<i>A. larrymoorei</i>	AF 3.10	<i>Ficus benjamina</i>	Bouzar <i>et al</i> , 1995
<i>A. rubi</i> (G11)	LMG 17935	<i>Rubus ursinus</i> var <i>Loganobaccus</i> , USA	
<i>A. vitis</i>	CFBP 5523		
<i>Rhizobium rhizogenes</i> (G10)	CFBP 2408	Apple tree	Ponsonnet, 1994
	K-84	<i>Prunus persicae</i> , near gall soil, Australia	
	CP85.120	<i>P. leuce</i> , gall, France	

Table 19.

Reference strains representing the diversity of the whole agrobacterial diversity known up to now.



primarily an ecotype with a genomic cohesion resulting from the periodic selections that purges the population of nearly all its diversity at all loci. The diversity of the ecotypes is constrained by periodic selections. Thus, an ecotype is composed by a founder and its descent. In a first time, the descent diverges radially from the founder, forming « star clades », which later diverges hierarchically forming clusters of very closely related strains. As a result, the numerical classification methods based on dichotomy and hierarchical clustering does not support star clades by significant bootstraps values.

We assumed, for the *Agrobacterium* genus, that the genomic species described by Popoff *et al.* (1984), and easily identifiable by AFLP (Mougel *et al.*, 2002) corresponds also to ecotypes. Each of the three major *Agrobacterium* species display a « star » genetic structure. Moreover, these species are assumed to be preferentially associated to particular plant species (Partie II). We hypothesized thus that the speciation in the *Agrobacterium* genus might be at least partly directed by preferential plant-bacterial associations.

For the plant pathologists, the term agrobacteria refers to bacteria which can harbor a Ti plasmid and display actual or potential pathogenicity. However, the term *Agrobacterium* is a genus name, which must be defined according to taxonomic rules whatever its pathogenicity. This leads to long-standing taxonomist quarrels (Young *et al.*, 2001 ; Farrand *et al.* 2003 ; Young *et al.*, 2003). A mid-term can be chosen between the lumpers proposal of Young including all agrobacteria in the genus *Rhizobium* to cope with the *rrs* polyphyly and the recognition of *Agrobacterium* as a definable genus by Farrand. In this paper, we will refer to *Agrobacterium* as the genus which includes biovar 1 members and closely related species, and to *Rhizobium rhizogenes* for biovar 2. The genus *Agrobacterium* is therefore composed of at least nine closely related genomic species (G1 to G9) (Popoff *et al.*, 1984) grouped in the so-called biovar 1, and three other species, *A. larrymoorei* (Bouzar and Jones, 2001), *A. vitis* (formerly known as biovar 3, Ophel and Kerr, 1990), and *A. rubi* (Starr and Weiss, 1943). Among the biovar 1, the genomic species G4 alone corresponds to the *bona fide* species *A. tumefaciens* (type strain ATCC23308 = B6), while C58 which is completely sequenced (Wood *et al.*, 2001 ; Goodner *et al.*, 2001) is the type strain for the genomic species G8, which has not yet received accepted *nomen species*. The genome of C58 is composed of four replicons, one circular

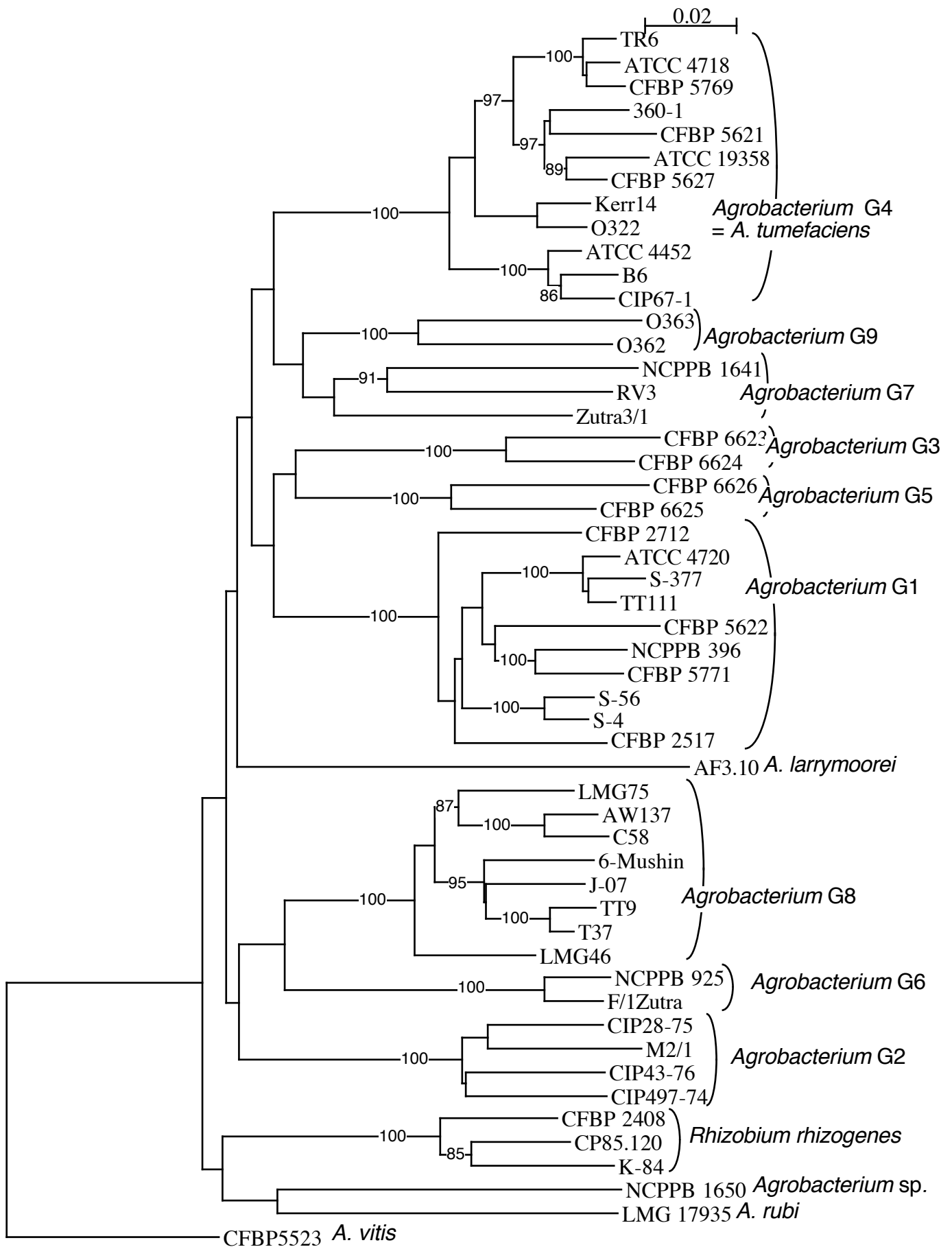


Figure 20: Phylogenetic tree of agrobacterial reference strains, calculated from four different pooled and duplicated analyses with the Neighbor-joining method. Significant bootstraps values, for 100 resampling, are dispalyed.

chromosome (CcC58, 2841420 bp), one linear chromosome (LcC58, 2075560 bp) and two plasmids, the cryptic plasmid (pAtC58, 542779 bp) and the Ti plasmid (pTiC58, 214233 bp). As an indication, the genomic species G1 (type strain TT111) is another genomic species of *Agrobacterium* frequently found in crown gall outbreaks (Parts I and II).

This work aims at establishing, whenever possible, a relationship between the genomic species concept and the ecological species concept, by using the genus *Agrobacterium*, which consists in closely related but well differentiated genomic species. From the AFLP data obtained on reference strains representing all the genus diversity known up to now, and the complete genome sequence of C58, we set up an approach to identify genes and genome regions involved in the delineation of *Agrobacterium* genomic species G8. This new approach consisted in, from experimental AFLP data and the bioinformatics analyses, to link the AFLP fragments, characterising a species, to known genes and functions. From AFLP analyses of the whole *Agrobacterium* genus, we determined the DNA markers specific of each of the three major species G1, G4 (= *A. tumefaciens*) and G8. We then determined by simulating AFLP on the C58 complete genome sequence, which of the predicted fragments are i) detected in C58, and ii) correspond to the consensus DNA markers specific of *Agrobacterium* species G8 to which belongs C58.

### **III. 2. Material and methods**

#### *III. 2. a. Bacterial strains*

The analyzed strains are listed Table 19. The reference strains represents all the species diversity of the genus *Agrobacterium* known up to now. A special emphasis had been put on species G1, G4 and G8, for which the number of reference strains available was the greatest (10, 12 and 8 strains, respectively) and because these three species are the most frequently involved in Crown Gall disease (Part II).

#### *III. 2. b. Molecular analysis*

##### DNA extraction

Strains DNA has been extracted with the DNeAsy TissueKit de Qiagen (Hilden, Germany), following manufacturer's instructions.

<b>AFLP CA</b>		<b>AFLP CC</b>		<b>AFLP CT</b>		<b>AFLP CG</b>	
Fragment size (mean)	Standard deviation	Fragment size (mean)	Standard deviation	Fragment size (mean)	Standard deviation	Fragment size (mean)	Standard deviation
52.10	0.07	48.94	0.12	65.70	0.08	48.16	0.18
65.91	0.23	64.35	0.09	128.12	0.08	55.90	0.09
67.04	0.04	89.22	0.08	150.23	0.07	63.11	0.07
89.31	0.08	161.35	0.12	156.55	0.12	91.53	0.29
108.37	0.05	184.75	0.21	171.60	0.22	94.68	0.25
116.93	0.30	573.97	0.13	186.06	0.16	108.51	0.04
131.91	0.05			191.17	0.12	176.49	0.12
136.59	0.08			204.87	0.17	195.83	0.17
146.81	0.33			246.84	0.63	213.68	0.08
163.08	0.26			386.72	0.28	221.40	0.04
177.57	0.19			525.49	0.12	232.70	0.12
191.68	0.28			528.16	0.19	236.35	0.29
271.39	0.49			589.69	0.23	244.49	0.13
295.09	0.12					250.73	0.11
339.75	0.43					253.91	0.18
418.95	0.21					260.71	0.14
423.55	0.21					299.40	0.16
504.42	0.30					308.22	0.11
525.96	0.18					542.74	0.15

Table 20 (a)

AFLP fragments detected for all strains of *Agrobacterium* species G1

<b>AFLP CA</b>		<b>AFLP CC</b>		<b>AFLP CT</b>		<b>AFLP CG</b>	
Fragment size (mean)	Standard deviation	Fragment size (mean)	Standard deviation	Fragment size (mean)	Standard deviation	Fragment size (mean)	Standard deviation
81.06	0.30	49.13	0.07	62.35	0.06	50.81	0.06
108.40	0.02	89.48	0.06	74.74	0.06	55.74	0.05
144.36	0.05	167.22	0.17	93.73	0.07	71.53	0.08
171.92	0.03	219.43	0.18	148.39	0.25	104.90	0.28
174.63	0.03	271.83	0.08	191.11	0.06	124.80	0.13
178.03	0.05	286.36	0.10	197.20	0.05	131.20	0.06
183.05	0.09	333.43	0.30	201.10	0.03	157.89	0.04
190.48	0.29	372.56	0.37	212.16	0.26	199.50	0.05
210.70	0.15	433.68	0.23	251.25	0.21	204.49	0.11
238.33	0.07	454.53	0.17	435.75	0.15	206.99	0.04
243.79	0.30	518.13	0.08	524.95	0.07	221.38	0.03
263.43	0.26	588.61	0.15	589.77	0.37	245.76	0.38
270.03	0.42					255.65	0.06
379.76	0.33					321.66	0.07
422.31	0.25					430.58	0.14
426.37	0.12						

Table 20 (b)

AFLP fragments detected for all strains of *Agrobacterium* species G4

## AFLP

AFLP used in this study is derived from the Vos *et al.* (1995), using *EcoRI* and *MseI* for genomic DNA restriction. Analyses are conducted on 55ng of genomic DNA. Restriction sites specific adaptators are ligated to the obtained fragments with adaptator *EcoRI* : F1363AdEco+ CTCGTAGACTGCGTACC and F1931AdEco- AATTGGTACGCAGTCTAC, adaptator *MseI* : F1365AdMse+ GACGATGAGTCCTGAG and F1931AdMse- TACTCAGGACTCAT. Primers F1247coreEco GACTGCGTACCAATTC and F1248coreMse GATGAGTCCTGAGTAA, whom sequences correspond to those of the adaptators and restriction sites, are used in a first non-selective PCR. Selective PCR is run with F1248coreMse primer and fluorescently labelled primers containing CoreEco with two discriminant nucleotides added in 3' position (CA : F1598EcoCA-FAM ; CC : F1599EcoCC-HEX ; CG : F1601EcoCG-HEX ; and CT : F1915EcoCT-FAM). In this study we will use the terms Eco+CA/Mse+0, Eco+CC/Mse+0, Eco+CG/Mse+0 and Eco+CT/Mse+0 to refer to the different AFLP analysis performed with respectively CA : F1598EcoCA-FAM ; CC : F1599EcoCC-HEX ; CG : F1601EcoCG-HEX ; and CT : F1915EcoCT-FAM primers. DNA polymerase, nucleotides and buffers required for PCR reaction are provided by the AFLP amplification CoreMix (Perkin-Elmer Applied Biosystems, Foster City, California). AFLP fragments were separated by electrophoresis using a capillary sequencer Megabace 1000 (Amersham Pharmacia Biotech Europe, Orsay, France), which automatically calculate the length of fluorescent fragments by comparison to the MegaBACE ET-900-R size standard (Amersham biosciences, Orsay, France). Genetic profiler V. 1.5 (Molecular Dynamics Inc., Sunnyvale, Californie) was used to display results, according threshold values, and to export data under compatible format for bioinformatic analyses.

For each strains, the four AFLP conditions (Eco+CA/Mse+0, Eco+CC/Mse+0, Eco+CT/Mse+0 or Eco+CG/Mse+0) were performed in duplicates.

Experimental fragments were coded as in the following exemple : the CA109 fragment corresponds to a fragment of 109 pb predicted or obtained with the Eco+CA/Mse+0 AFLP conditions.

<b>AFLP CA</b>		<b>AFLP CC</b>		<b>AFLP CT</b>		<b>AFLP CG</b>	
Fragment size (mean)	Standard deviation	Fragment size (mean)	Standard deviation	Fragment size (mean)	Standard deviation	Fragment size (mean)	Standard deviation
68.22	0.30	82.63	0.08	40.78	0.13	40.66	0.15
95.63	0.26	84.92	0.13	66.76	0.08	50.78	0.05
106.24	0.31	89.26	0.07	88.48	0.08	55.68	0.14
107.95	0.05	213.03	0.13	136.81	0.05	103.92	0.08
121.54	0.06	334.31	0.05	172.72	0.26	108.78	0.05
164.40	0.32	520.08	0.15	266.91	0.54	120.00	0.03
210.24	0.10	574.06	0.15	292.32	0.21	180.46	0.30
231.34	0.24			300.63	0.24	190.69	0.13
233.23	0.05			302.09	0.45	221.38	0.04
244.37	0.05			335.97	0.13	223.53	0.26
463.25	0.11			374.34	0.09	246.23	0.32
507.26	0.15			441.69	0.60	252.77	0.12
627.13	0.11			590.46	0.22	260.56	0.33
						275.38	0.05
						318.09	0.07
						357.18	0.12
						363.03	0.12
						370.14	0.09
						374.54	0.30
						391.18	0.18
						400.29	0.25

Table 20 (c)

AFLP fragments detected for all strains of *Agrobacterium* species G8

### *III. 2. e. Phylogenomic analyses*

The LecPCR, Align2 and Lis programs, used to transform raw data in tabular binary data, and the DistAFLP program, used to calculate evolutionary genome divergences, are described by Mougél *et al.* (2002), and are accessible by free ftp at <http://pbil.univ-lyon1.fr/ADE-4/microb/>. The phylogenomic cladograms are calculated with the Neighbor/UPGMA module of the PHYLIP package (Felsenstein, 1993) (Phylip), and displayed with NJ-Plot (<http://pbil.univ-lyon1.fr/>). Bootstrap values were calculated by using the bootstrap option of DistAFLP, and the Neighbor/UPGMA and Consens modules of the PHYLIP package.

### *III. 2. f. AlignK analysis*

The program AlignK developed by Jean Thioulouse was used to compare the different AFLP analyses of strains belonging to a same species. For one AFLP condition, AlignK calculates how frequent is a fragment of a given size, its average length and the length standard deviation. The program was used to determine consensus fragments detected in all the strains of a given species.

### *III. 2. g. Predictive AFLP*

AFLP fragments were predicted as described by Arnold *et al.* (1999) by looking for DNA restriction sites and selective nucleotides in complete genome sequences. For this purpose, the Digest program was developed by Guy Perrière and was used to perform simulate AFLP on the complete genome sequence of strain C58 (accession numbers : circular chromosome : NC\_003304 ; Linear chromosome NC\_003305 ; cryptic plasmid : NC\_003306 and Ti plasmid : NC\_003308). The program gives the genome location and the predicted length of restriction fragments. This length, plus 27pb for the adaptators, gives the predicted length of the resulting AFLP fragments.

## **III. 3. Results**

### *III. 3. a. AFLP analyses*

Each AFLP analysis yielded 23 to 171 fragments, depending on the analyzed strain and the selective nucleotides. We obtained 69, 45, 76 and 66 fragments in average using Eco+CA/Mse+0, Eco+CC/Mse+0, Eco+CG/Mse+0 and Eco+CT/Mse+0 AFLP conditions, respectively. Except for species G7, the phylogenomic analysis, allowed

AFLP experiment	Circular Chromosome	Linear Chromosome	At Plasmid	Ti Plasmid	Total
AFLP CA	56	50	10	7	123
AFLP CC	27	22	6	4	59
AFLP CT	62	35	12	8	117
AFLP CG	76	50	10	6	142
Total	221	157	38	25	441
% of fragments	50.11	35.60	8.62	5.67	100.00
Genome % of each replicon	50	36.6	9.6	3.8	100.00

Table 21

Number and feature location of predicted AFLP fragments comprised between 30 and 900 bp.

AFLP experiment	Detected		Not detected	Detected	Total
	Predicted	Homoplasic			
CA	47	42.5%	10	4	61
CC	29	6.9%	4	1	34
CT	35	51.4%	22	7	64
CG	61	49.2%	18	4	81

Table 22

Number of AFLP fragments comprised between 30 and 400 pb predicted and detected in C58 strain.



to assign all strains to its correct genomic species, which formed clusters supported by significant bootstrap values (Figure 16).

The three major agrobacterial species, namely G1, G4 and G8, display mean rates of nucleotide substitutions of 5.9%, 5.4% and 5.7%, respectively. The greatest evolutionary genome divergence found among these species were 7.6, 7.7 and 7.8 % in species G1, G4 and G8., respectively. In species G8 this maximum value was found between strain C58 and strain T37.

### *III. 3. b. Consensus AFLP markers of genomic species*

AFLP data were compared simultaneously with AlignK to determine common fragments obtained with all the strains of a given species (Table 20). For the four AFLP conditions tested, the total number of fragments common to a species were 57, 55 and 54 for G1, G4 and G8, respectively. Only rare fragments (9%), such as CA109 and CG221 located inside the highly conserved *rrs* gene, were consistently found in all members of the three species. Fragments common to all members of two species are also rare. These AFLP fragments common to all members of a species are species specific markers involved in the delineation of the species by AFLP.

### *III. 3. c. Predictive AFLP*

A simulation of AFLP was performed on the C58 complete genome sequence. The number of fragments predicted in the four replicons is proportional to the replicon length, indicating that AFLP fragments are evenly distributed on the whole genome (Table 21).

The genes and gene regions corresponding to the predicted AFLP fragments were identified. About 95% of the fragments are located on a gene, or on a gene and the intergenic region, rarely in an intergenic region alone (data not shown).

### *III. 3. d. Detected versus predicted AFLP fragments in strain C58*

In order to establish the correspondance between detected and predicted AFLP fragments, we analyzed the length variability of two fragments corresponding to a unique *EcoRI* site occurring in the *rrs* of G1, G4 and G8 (Mougel *et al.* 2002). As corresponding *rrs* sequences are identical in the three species, the predicted fragment CG221 was found to correspond to detected fragments of equivalent lengths 221.40, 221.38, 221.38 bp in G1, G4 and G8, respectively (Table 20). To the

	AFLP consensus profile	C58 predicted fragment	Replicon	Fragment coordinates	Genomic region	Concerned Gene(s)	Function
CA	68,22	69	pAt	441828 441869	Intergenic spacer	ATU5448 ATU5449 (ibpA)	conserved hypothetical protein heat-shock protein
	95,63	96	Cc	1357976 1358044	Gene	ATU1363	conserved hypothetical protein
	106,24	107	pTi	202825 202904	Gene	ATU6185 (virD5)	virA/G regulated protein
	107,95	1 109 3 109 4 109 2 109	Cc	2517922 2518003	Gene	ATU2547 (SSU 16S rRNA operon1)	SSU 16S ribosomal RNA
			Lc	1042340 1042421	Gene	ATU3937 (SSU 16S rRNA operon3)	SSU 16S ribosomal RNA
			Lc	1304837 1304918	Gene	ATU4180 (SSU 16S rRNA operon4)	SSU 16S ribosomal RNA
			Cc	57277 57358	Gene	ATU0053 (SSU 16S rRNA operon2)	SSU 16S ribosomal RNA
	121,54	123	Cc	521447 521542	Gene and intergenic spacer	ATU0528 (mscL)	large conductance mechanosensitive channel
	164,40	164	Cc	1569922 1570058	Gene and intergenic spacer	ATU1587	hypothetical protein
	210,24	208	Lc	1314896 1315076	Gene	ATU4192	ABC transporter, substrate binding protein [oligopeptide]
	231,34	231	Cc	527114 527317	Gene	ATU0535	conserved hypothetical protein
	233,23	2 232 1 232	pTi	79319 79523	Gene and intergenic spacer	ATU6066 (acd)	acyl-CoA dehydrogenase
			Cc	1043876 1044080	Gene	ATU1050	conserved hypothetical protein
244,37	243 244	Lc	1593505 1593720	Genes and intergenic spacer	ATU4447 ATU4448	ABC transporter, substrate binding protein [sorbitol] ABC transporter, membrane spanning protein [sorbitol/mannitol]	
		Lc	1279838 1280054	Gene and intergenic spacer	ATU4162	two component response regulator	
CC	82,63	83	Cc	2837339 2837394	Gene	ATU2832 (thdF)	tRNA modification GTPase
	84,92	85	Cc	2102686 2102743	Gene	ATU2137	hypothetical protein
	89,26	90	Cc	2542186 2542248	Gene	ATU2571	hypothetical protein
	213,03	213	Cc	1139616 1139801	Gene	ATU1148	conserved hypothetical protein
	334,31	335	CC	2348553 2348860	Gene and intergenic spacer	ATU2375 (exoT)	UDP-hexose transferase

Table 23 (a)

CA and CC AFLP consensus profile for *Agrobacterium* G8 linked with genes and functions

Cc : Circular Chromosome

Lc : Linear chromosome

pAT : At plasmid

pTi : Ti plasmid

The homoplastic fragments are shaded

	AFLP consensus profile	Predicted fragment	Genome location	Genome coordinates	Genomic region	Concerned Gene(s)	Function
CT	40,78	2 41	Lc	1514223 1514236	Gene	ATU4376	conserved hypothetical protein
		1 41	Cc	1510120 1510133	Gene	ATU1520 (znuC)	ABC transporter, nucleotide binding/ATPase protein [zinc]
	66,76	67	Cc	2183189 2183228	Intergenic spacer	ATU2214 ATU2215	conserved hypothetical protein conserved hypothetical protein
	88,48	89	Lc	661263 661324	Gene and intergenic spacer	ATU3605	hypothetical protein
		90	Cc	2025389 2025451	Gene	ATU2061	hydrolase
	136,81	136	Cc	2764723 2764831	Gene and intergenic spacer	ATU2763 (mtrA)	two component response regulator
		137	Cc	463094 463203	Gene	ATU0474	ECF family sigma factor
	172,72	171	Cc	2143577 2143720	Gene	ATU2172	multidrug efflux protein
		174	Cc	299905 300051	Gene	ATU0307 (ugpE)	ABC transporter, membrane spanning protein
	266,91	265	Lc	1349995 1350232	Gene and intergenic spacer	ATU4224 (soxB)	sarcosine oxidase beta subunit
	292,32	NP					
	300,63	298	Lc	967673 967943	Gene and intergenic spacer	ATU3874 (lyxK)	L-xylulose kinase
	302,09	302	pAT	235236 235510	Gene	ATU5239	transcriptional regulator, LacI family
		303	Lc	1252994 1253269	Genes and intergenic spacer	ATU4133 ATU 4134	hypothetical protein hypothetical protein
335,97	336	Cc	1139307 1139615	Gene	ATU1148	conserved hypothetical protein	
374,34	372	Cc			Genes and intergenic spacer	ATU1683	hypothetical protein
						ATU1684	hypothetical protein

Table 23 (b)

CT AFLP consensus profile for *Agrobacterium* G8 linked with genes and functions.

Cc : Circular Chromosome

Lc : Linear chromosome

pAT : At plasmid

pTi : Ti plasmid

NP : Not predicted

The homoplasic fragments are shaded

AFLP consensus profile	Predicted fragment	Genome location	Genome coordinates	Genomic region	Concerned Gene(s)	Function
CG 40,66	39	Lc	1141680 1141691	Intergenic spacer	ATU4025 (fhuC) ATU4026	ABC transporter, nucleotide binding/ATPase protein [ferrichrome] conserved hypothetical protein
	40	Lc	1614572 1614584	Gene and intergenic spacer	ATU4472	hypothetical protein
50,78	51	Cc	382901 382924	Intergenic spacer	ATU0389 (tRNA-Arg) ATU0390 (ubiE)	transcriptional regulator, LysR family ubiquinone/menaquinone biosynthesis
55,68	56	Cc	2839681 2839709	Gene and intergenic spacer	ATU2834	conserved hypothetical protein
103,92	1 104	Cc	1147011 1147087	Intergenic spacer	ATU1154 ATU1155	conserved hypothetical protein conserved hypothetical protein
	2 104	Cc	828653 828729	Gene	ATU0829 (betB)	betaine aldehyde dehydrogenase
108,78	2 108	Lc	1760262 1760342	Gene	ATU4613	glycosyltransferase
	1 108	Cc	2248512 2248592	Intergenic spacer	ATU2276 (braC) ATU2277	ABC transporter, substrate binding protein [branched chain amino acid] conserved hypothetical protein
120,00	120	Cc	2107533 2107625	Gene	ATU2142 (glnA)	glutamine synthetase
180,46	180	Cc	1337901 1338053	Gene	ATU1348	conserved hypothetical protein
190,69	189	Cc	2729384 2729545	Gene and intergenic spacer	ATU2734	dehydrogenase
221,38	5 221	Lc	858754 858947	Genes and intergenic spacer	ATU3775 ATU3776	conserved hypothetical protein hypothetical protein
	3 221	Lc	1042422 1042615	Gene	ATU3937 (SSU 16S rRNA operon3)	SSU 16S ribosomal RNA
	4 221	Lc	1304919 1305112	Gene	ATU4180 (SSU 16S rRNA operon4)	SSU 16S ribosomal RNA
	1 221	Cc	57359 57552	Gene	ATU0053 (SSU 16S rRNA operon2)	SSU 16S ribosomal RNA
	2 221	Cc	2517728 2517921	Gene	ATU2547 (SSU 16S rRNA operon1)	SSU 16S ribosomal RNA
223,53	NP					
246,23	1 245	Lc	79555 79772	Gene and intergenic spacer	ATU3073	aspartate racemase
	2 245	Lc	946624 946841	Genes and intergenic spacer	ATU3856 ATU3857	hypothetical protein hypothetical protein
252,77	254	CC	656376 656602	Gene and intergenic spacer	ATU0661	glycosyltransferase
260,56	NP					
275,38	275	Cc	1472503 1472750	Gene and intergenic spacer	ATU1480	conserved hypothetical protein
318,09	320	Cc	2216741 2217033	Gene and intergenic spacer	ATU2243	conserved hypothetical protein
357,18	1 358	Lc	861803 862133	Gene and intergenic spacer	ATU3704	conserved hypothetical protein
	2 358	Cc	2527749 2528079	Gene and intergenic spacer	ATU2554 (poxA)	lysyl-tRNA synthetase
363,03	363	Lc	782808 783143	Gene	ATU3704	conserved hypothetical protein
	364	Cc	26912 27248	Gene	ATU0025	conserved hypothetical protein
370,14	369	Cc	2608862 2609203	Gene and intergenic spacer	ATU2625 (atpH)	ATP Synthase delta chain
	370	Cc	2569156 2569498	Gene	ATU2591 (redB)	glucosyltransferase
374,54	1 376	Cc	843042 843390	Gene and intergenic spacer	ATU0842	conserved hypothetical protein
	2 376	Cc	1697687 1698035	Gene	ATU1711 (argS)	arginyl-tRNA synthetase
391,18	NP					

Table 23 (c) : CG AFLP consensus profile for *Agrobacterium* G8 linked with genes and functions (Cc : Circular chromosome ; Lc : Linear chromosome ; pAt : At plasmid ; pTi : Ti plasmid ; NP : Not Predicted. The homoplasic fragments are shaded.

contrary, four nucleotide differences occur at positions 522-523 et 533-534 between the *rrs* sequences of G8 and, those of G1 and G4 corresponding to the predicted fragment CA109 (data not shown). These nucleotide differences led to a small (0.42) but visible difference between the estimated lengths of detected fragments :108.37, 108.40 and 107.95 for G1, G4 and G8, respectively (Table 20). As the sequences of genes encoding for proteins are larger than those found in ribosomal genes, we assumed that a detected fragment correspond to a predicted fragment plus or minus 2pb. We also assumed that this uncertainty is increasing with fragment size. We thus decided to limit the comparison between predicted and detected fragments from 30 to 400pb.

The genome location of experimentally detected fragments of strain C58 was determined. Some predicted fragments were not detected and some but rarer fragments were detected but not predicted since they did not fit to the length of a predicted fragment +/- 2pb (Table 22). On the other hand, the correct assignation of a detected fragment to a precise genome location was frequently hampered by the occurrence of several candidates among the predicted fragments (eg homoplasmy).

The detected fragments are evenly distributed on CcC58 and LcC58 chromosomes. Figure 21 indicates the position in centisome on the two main replicons (CcC58 and LcC58) of fragments predicted but not detected in C58 (0), detected but imprecisely located due to homoplasmy on the same or another replicon (1), and detected fragments which were unambiguously located (2).

### *III. 3. e. Consensus AFLP markers: identification and localisation*

Only fragments that could be unambiguously located (2, Figure 21) were used to study the genomic location of G8 species markers. These fragments occurred on the four replicons : 22, 14, 2 and 1 on CcC58, LcC58, pAtC58 and pTiC58, respectively. Figure 21 indicates the location of G8 species markers (2) and those of detected fragments precisely located but that were not found in all G8 strains (1) on the two chromosomes. The homogenous distributions of common and non-common fragments on the chromosomes was tested. With CcC58, fragment occurrences around the origin, oriC, and around the opposite (0.875 to 0.125 centisomes, and 0.375 to 0.625 centisomes, respectively) was compared to intermediate zones (0.125 to 0.375 centisomes and 0.625 to 0.875 centisomes) in order to cope with a possible

	Concerned Gene(s)	Function
Circular Chromosome	ATU0025	conserved hypothetical protein
	ATU0053 (SSU 16S rRNA operon2)	SSU 16S ribosomal RNA
	ATU0307 (ugpE)	ABC transporter, membrane spanning protein
	ATU0389 (tRNA-Arg)	transcriptional regulator, LysR family
	ATU0390 (ubiE)	ubiquinone/menaquinone biosynthesis
	ATU0474	ECF family sigma factor
	ATU0528 (mscL)	large conductance mechanosensitive channel
	ATU0535	conserved hypothetical protein
	ATU0661	glycosyltransferase
	ATU0829 (betB)	betaine aldehyde dehydrogenase
	ATU0842	conserved hypothetical protein
	ATU1050	conserved hypothetical protein
	ATU1148	conserved hypothetical protein
	ATU1154	conserved hypothetical protein
	ATU1155	conserved hypothetical protein
	ATU1348	conserved hypothetical protein
	ATU1363	conserved hypothetical protein
	ATU1480	conserved hypothetical protein
	ATU1520 (znuC)	ABC transporter, nucleotide binding/ATPase protein [zinc]
	ATU1587	hypothetical protein
	ATU1683	hypothetical protein
	ATU1684	hypothetical protein
	ATU1711 (argS)	arginyl-tRNA synthetase
	ATU2061	hydrolase
	ATU2137	hypothetical protein
	ATU2142 (glnA)	glutamine synthetase
	ATU2172	multidrug efflux protein
	ATU2214	conserved hypothetical protein
	ATU2215	conserved hypothetical protein
	ATU2243	conserved hypothetical protein
	ATU2276 (braC)	ABC transporter, substrate binding protein [branched chain amino acid]
	ATU2277	conserved hypothetical protein
	ATU2375 (exoT)	UDP-hexose transferase
	ATU2547 (SSU 16S rRNA operon1)	SSU 16S ribosomal RNA
	ATU2554 (poxA)	lysyl-tRNA synthetase
	ATU2571	hypothetical protein
	ATU2591 (redB)	glucosyltransferase
	ATU2625 (atpH)	ATP Synthase delta chain
	ATU2734	dehydrogenase
	ATU2763 (mtrA)	two component response regulator
	ATU2832 (thdF)	tRNA modification GTPase
ATU2834	conserved hypothetical protein	

Table 24 (a)

Circular chromosome genes corresponding to AFLP consensus profile for *Agrobacterium* species G8

leading-lagging effect. Not very significant difference was found ( $\chi^2=3.504$ , 1 df,  $p=0.061$ ). With LcC58, the homogenous distribution of fragments was tested for two classes corresponding to LcC58 extremities (0,00 to 0.333 centisomes and 0.666 to 1.000 centisomes) compared to LcC58 central part (0.333 to 0.666 centisomes) knowing that *oriC* is located at 0.332. Statistics indicates that the G8 species markers were significantly more frequent in the central part than in the extremities in LcC58 ( $\chi^2=4.416$ ; 1 df ;  $p=0.036$ ).

Homoplasic and non-homoplasic G8 species markers were located and identified, hitting very diverse genetic functions and a lot of hypothetical proteins. Except the ribosomal operon, no house-keeping genes were in the list (Table 24).

### **III. 4. Discussion**

In this study, we found species specific molecular markers for the three three major species G1, G4 (= *A. tumefaciens*) and G8 of the genus *Agrobacterium*. Those markers were obtained by massive arbitrar sampling of the genomes thanks to the AFLP method. According to the recent revision of the bacterial species definition (Stackebrandt *et al.* 2002), such molecular markers are now accepted as relevant phenotypic traits suitable to describe bacterial species. These markers have several advantages in the frame of the description of a species on a genomic basis. The genome sampling is massive with an average per strain of 256 fragments (i.e. 256 different genome regions) for four AFLP conditions. On the other hand, contrary to the multi Locus Sequence Typing approach (Maiden *et al.*, 1988) based on a little number of genes, and generally *a priori* chosen among house-keeping genes, the AFLP markers result from a genome sampling based on the even distribution of one restriction site. *EcoRI* was used in this study because peaks in the profile are regularly distributed, but almost all other endonuclease can be used instead.

The phylogenomic analyses of the reference strains shown results consistant with former diversity analyses performed with other sequencing device (Mougel *et al.*, 2001). However, as already showed by Mougel *et al.* (2002), the species G7 requests however much more AFLP analyses, because it is the *Agrobacterium* species that displays the highest internal diversity. For the other species, generally only one AFLP condition was enough to correctly assign a strain to a species.

	Concerned Gene(s)	Function
Linear Chromosome	ATU3073	aspartate racemase
	ATU3605	hypothetical protein
	ATU3704	conserved hypothetical protein
	ATU3775	conserved hypothetical protein
	ATU3776	hypothetical protein
	ATU3856	hypothetical protein
	ATU3857	hypothetical protein
	ATU3874 (lyxK)	L-xylulose kinase
	ATU3937 (SSU 16S rRNA operon3)	SSU 16S ribosomal RNA
	ATU4025 (fhuC)	ABC transporter, nucleotide binding/ATPase protein [ferrichrome]
	ATU4026	conserved hypothetical protein
	ATU4133	hypothetical protein
	ATU4134	hypothetical protein
	ATU4162	two component response regulator
	ATU4180 (SSU 16S rRNA operon4)	SSU 16S ribosomal RNA
	ATU4192	ABC transporter, substrate binding protein [oligopeptide]
	ATU4224 (soxB)	sarcosine oxidase beta subunit
	ATU4376	conserved hypothetical protein
	ATU4447	ABC transporter, substrate binding protein [sorbitol]
	ATU4448	ABC transporter, membrane spanning protein [sorbitol/mannitol]
ATU4472	hypothetical protein	
ATU4613	glycosyltransferase	
At Plasmid	ATU5239	transcriptional regulator, LacI family
	ATU5448	conserved hypothetical protein
	ATU5449 (ibpA)	heat-shock protein
Ti Plasmid	ATU6066 (acd)	acyl-CoA dehydrogenase
	ATU6185 (virD5)	virA/G regulated protein

Table 24 (b)

Linear chromosome, At and Ti plasmid genes corresponding to AFLP consensus profile for *Agrobacterium* species G8.



With the present work, we propose indeed a reliable and portable method to recognise the three major *Agrobacterium* species even in the absence of databases and phylogenomic analyses. It is possible to identify the genomic species only by looking for the species specific markers, since they are valid traits to accurately assign strains to genomic species. To our knowledge, this is the first time that it is possible to identify bacterial species with such a large number of well defined DNA markers. Moreover, as AFLP profile data bank or at least the present list of selected species specific markers are available on the web, this approach is becoming easily accessible to all laboratories generally without the need for phylogenetic analyses. Moreover, the present offer of AFLP analysis services by several private companies, makes this approach really accessible even to laboratories that have no sequencer.

Using the AFLP gives also the advantage to allow for testing the intra-specific diversity and more diverse are the strain chosen more relevant are the determined species specific markers. As a matter of fact, the species specific markers (Table 20) were selected because they were detected in all the tested strains. Nevertheless, as indicated by factorial analyses (data not shown), there is a continuous gradient of markers from those hitting all the strains to those yielded by a single strain. Among G1, G4 and G8 the maximum diversity between two strains of a given genomic species is about 7.5%. The maximum evolutionary genome divergence value existing between two strains of a given genomic species is 10,5% (in G7, Mougel *et al.*, 2002). Thus the strains analysed for these three species, are not the more diverse but still are diverse enough to ensure significant diversity analyses and relevant marker determination.

It is assumed that common AFLP fragments originate from the same genomic regions in all the strains belonging to a particular species. These regions are thus conserved amongst all members of the species, and the corresponding genes are should belong to the « core of genes » described by Vulic *et al.* (1997), that coevolved and formed the genetic base of a species. We set up a procedure to locate and identify what are those core genes. The predictive AFLP performed on C58 strain permitted us to go further on our analyses and to link the detected AFLP fragments to genomic regions. Arnold *et al.* (1999) has already performed predictive

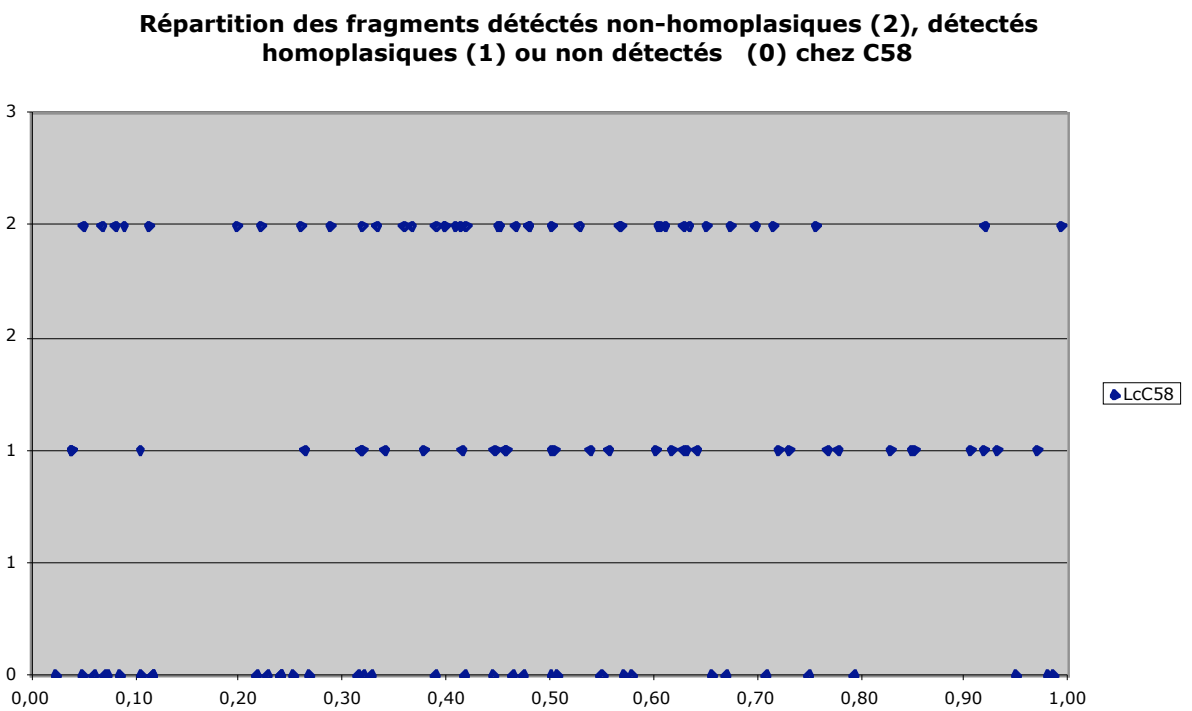
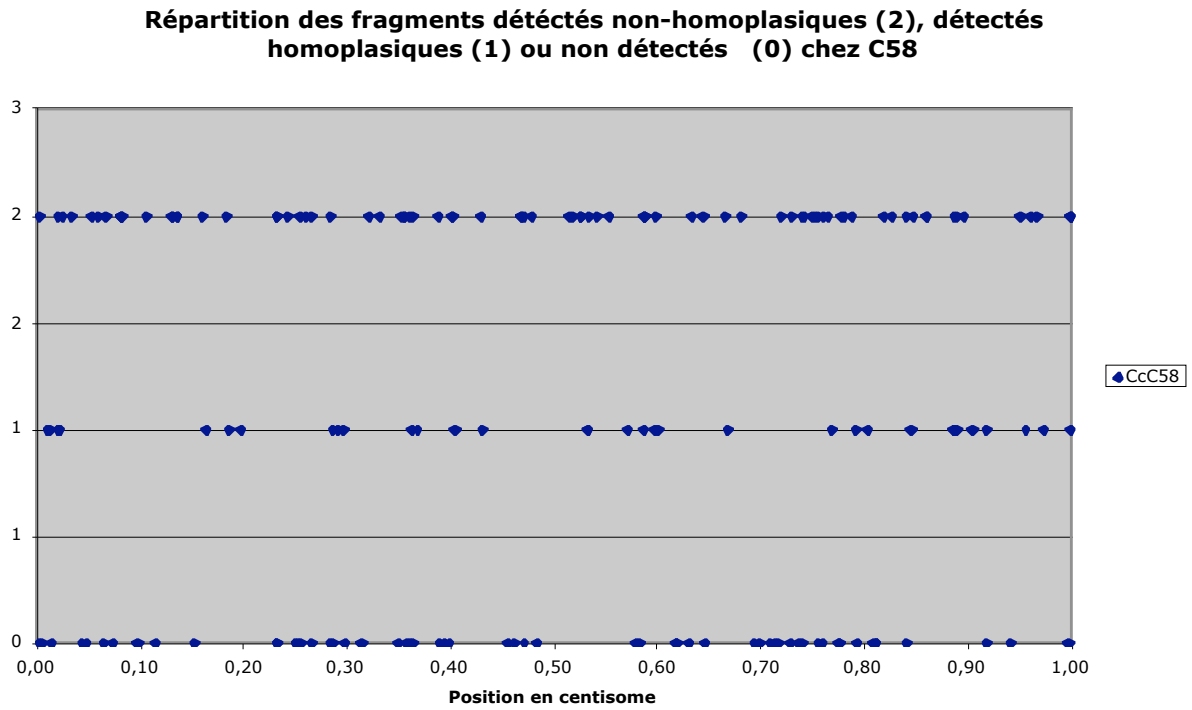


Figure 21.

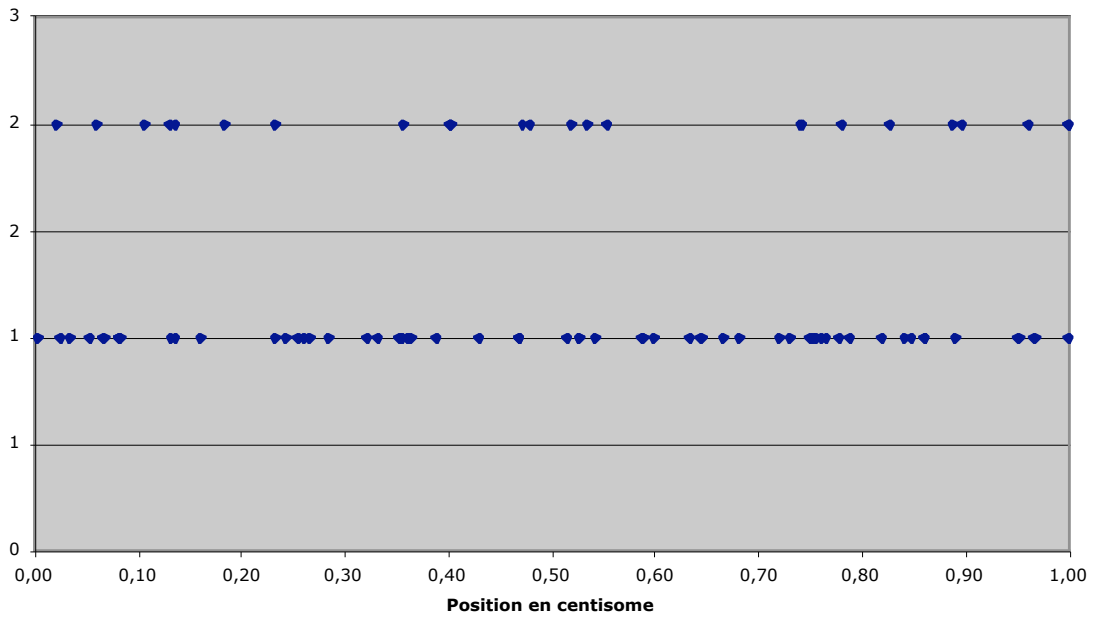
AFLP on the *Escherichia coli* genome. They determined the validity of their AFLP analyses.

In the present work, since strain C58 harbors four replicons, we studied the distribution of predicted and detected AFLP fragments on the whole genome. This distribution is regular for the predicted and detected fragments, showing that the AFLP analysis take in account DNA markers scatters all over the genome. However, not all the predicted fragments, between 30 and 500 pb, were experimentally detected and some other fragments are not predicted (Table 22). It is possible that DNA modifications, especially methylation to which *EcoRI* is sensitive, alter the accessibility to some restriction sites (Reisenauer *et al.*, 1999). Thus, if a restriction site is not accessible for the restriction enzyme, either the associated fragment will disappear from the analyses, or the size of the associated fragment will change, leading to detection of non-predicted fragments. Another cause could be partial digestion by *EcoRI* due to bad buffer. It is preferable to perform digestions in two steps, firstly the *EcoRI* digestion and then the *MseI* (C. Manceau, personal communication). Nevertheless, AFLP is reliable, but we found that the weakest level of homoplasy, of predicted/not-detected and of detected/not-predicted were obtained with the Eco+CC/Mse+0 condition (Table 22), probably in relation with the highest stringency of C in 3' end of selective primers (Sugimoto *et al.*, 2000).

As a result, the detected fragments of C58 were found to be located evenly on the chromosomes (Figure 21), indicating no significative bias for the genome sampling provided by the method.

We then located and identified species specific markers of G8 thanks to AFLP simulation done with its type strain C58, since PCR analyses with relevant primers shown that fragments detected in other G8 strains corresponds to the actual fragments predicted for C58 (results not shown). Due to homoplasy, a large number of fragments could hardly be identified. Nevertheless, the number of unambiguously identified fragments showed that species specific fragments are regularly distributed on the circular chromosome, but showed significantly irregular distribution on the linear chromosome (Figure 22). It looks like if linear chromosome extremities were less conserved among the G8 species than its central part. As if extremities encoded functions more variable from strain to strain. Little is known about the role of the linear chromosome of *Agrobacterium* and examination of genes and functions gives

Répartition des fragments consensus de G8 (2),  
ou seulement détectés (1) chez C58



Répartition des fragments consensus de G8 (2),  
ou seulement détectés (1) chez C58

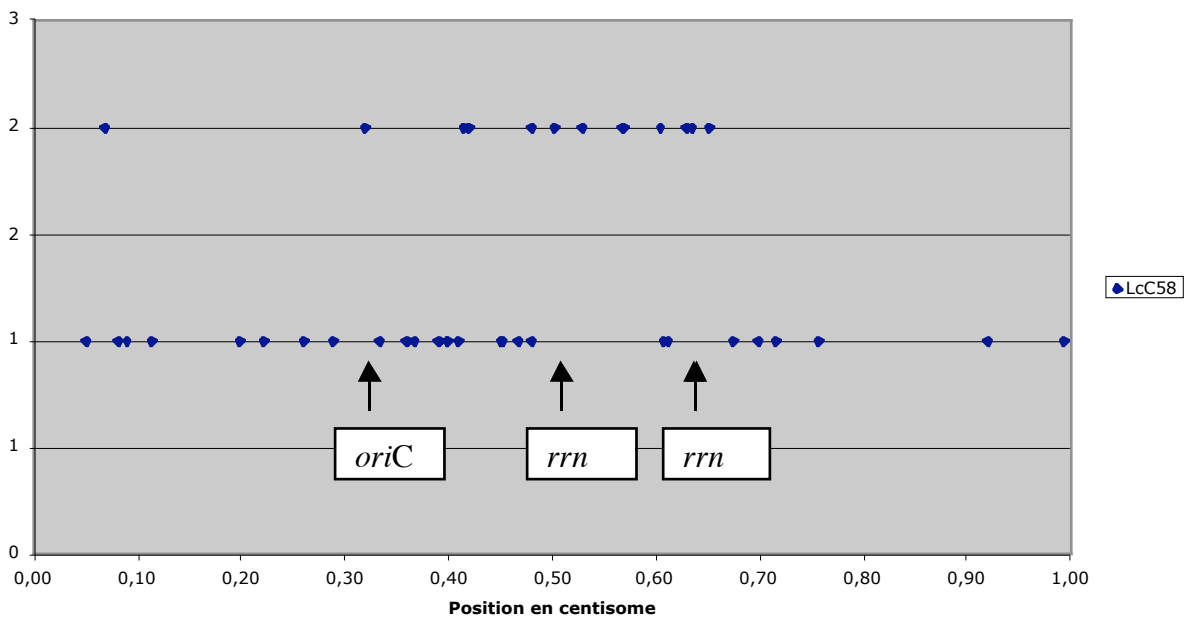


Figure 22.

little information. Nevertheless, it is significant that the two highly conserved ribosomal operons (*rrn*) found on the linear chromosome are actually in its middle part (Figure 22). Moreover, it is known that extremities of the linear chromosomes in *Streptomyces* bear non-essential genes (Volf and Altenbuchner, 1998). It is thus possible that the extremities of the C58 linear chromosome harbour genes involved in strain or sub-ecotype specificity.

Numerous genes were identified (Table 24). To our knowledge, this is the first time that such a large panel of species specific markers were identified. Since these genes were chosen without *a priori*, this provide the first non biased picture of the core of gene of a species. No or rare of them were house keeping genes. This points out what should be the most relevant genes for a MLST analyse of the *Agrobacterium* genus. The MLST (Multi Locus Sequence Typing) is based on the sequencing of about seven genes (Maiden *et al.*, 1998). But these genes are *a priori* chosen amongst house-keeping genes. With our approach it becomes possible to choose the best candidate genes for MLST analysis implied in species definition.

Cohan proposes that the bacterial species correspond to ecotypes, which diverged under ecological pressures (Cohan, 2000, 2001). We observed that in the *Agrobacterium* genus, the species could correspond to ecotypes (Partie II). In the ecotypes the diversity among strains is regularly purged by periodic selection. Thus, any mutation occuring in one of the genes implied in the association between the ecotype and its ecological niche wil provoke periodic selection. Thus these genes are forming the core of the ecotype, thus for Cohan, the core of the species. Knowing these genes, part of the core of the species, could lead to understand which ecological conditions provoked the speciation.

Among the 69 genes concerned by the consensus AFLP fragments, 32 correspond to hypothetical proteins. Thus, it is presently difficult to make firm conclusions about functions involved in G8 speciation from these data. However, the second group (7 out of 69) of consensus fragments correspond to substrate binding protein of ABC transporter. These genes could be implied in plant-bacteria interactions. This will substantiate our hypothesis that plant-bacterial association could have play a major role in the speciation in the genus *Agrobacterium*.

### III. 5. Conclusion

In this study we present a method to assignate rapidly an agrobacterial strain to one of the three *Agrobacterium* major species (*A. sp* G1, type strain TT1111 ; *A. sp* G4, the *bona fide* *A. tumefaciens*, type strain B6 = ATCC23308 ; and *A. sp* G8, type strain C58) by giving lists of consensus AFLP markers detected in all strains of these three species.

We also present an approach combining experimental, on a representative set of strains of the diversity of the analysed species, and predictive AFLP onto complete genome sequence of one strain of these species.

This method allowed us to identify directly the genes, or at least the genomic regions, implied in bacterial species delineation. The question could be asked, whether these genes can point out the ecological reasons responsible for the speciation. For the *Agrobacterium* genus the plant could play a role in speciation. However, more studies are required to establish whether, the core gene detected in this study correspond to actual ecological functions themselves involved in the speciation process, or are conserved just by a « hitch-hiking » phenomenon consecutive to diversity purgation occuring simultaneously at all loci.

## Conclusions et perspectives

# Conclusion et perspectives

## I. AFLP

Ce travail de thèse nous a permis de confirmer la puissance de l'AFLP pour l'étude génomique des populations bactériennes et son utilisation possible comme outil de suivi épidémiologique. A partir des données AFLP, il est possible de construire des bases de données très utiles dans le cadre d'analyses épidémiologiques. Ceci est particulièrement vrai pour la maladie de Galle du Collet, dont les agents pathogènes sont très divers pour une même épidémie.

L'AFLP nous a permis aussi de déterminer, parfois, la qualité des isolats étudiés. Une étude menée sur les agrobactéries pathogènes de l'eucalyptus a montré que les isolats n'étaient en fait pas purs (résultats non présentés), ne permettant pas, de fait, une analyse plus poussée de ces bactéries.

L'analyse simultanée de plusieurs conditions AFLP différentes (*ie* effectuées avec d'autres enzymes de restriction) permettraient d'apporter une plus grande précision des résultats. Ceci est notamment crucial dans le cas d'espèces présentant une très grande variabilité entre les souches, comme l'espèce G7 du genre *Agrobacterium*. L'augmentation de l'information devrait également permettre d'établir les relations phylogénomiques statistiquement valides entre espèces génomiques.

L'utilisation d'autres enzymes de restriction peuvent apporter les données supplémentaires nécessaires. De plus, ceci permettrait d'éviter les biais dus à l'utilisation de l'enzyme *EcoRI*. En effet le site de restriction de cette enzyme est susceptible d'être méthylé (Kahng and Shapiro, 2001), altérant les analyses AFLP.

Une simplification des procédures d'analyse par AFLP pourraient permettre une manipulation plus aisée de la méthode et éventuellement améliorer sa reproductibilité. Clerc *et al.* (1998) utilisent une procédure d'AFLP simplifiée, en utilisant une seule enzyme de restriction, et en visualisant les résultats sur simple gel d'agarose. Là également il faudrait déterminer le profil consensus commun à toutes les souches d'une même espèce. Par contre, cette procédure non-automatisée pose le problème de la détermination fiable de la taille des fragments obtenus.



Toutefois, le recours à des sociétés de services pour l'analyse des AFLP (tout comme les services de séquençages), doit maintenant permettre à tous les laboratoires d'avoir facilement accès à l'AFLP.

## II. Sélection des agrobactéries par la plante

Ces travaux nous ont permis de préciser le rôle des plantes dans la structuration des populations d'*Agrobacterium* dans le sol. Nous avons observé une sélection de certaines espèces d'agrobactéries exercée par les plantes. Par contre, cette sélection n'est pas exclusive et dépend des conditions du milieu. Elle est en effet contrainte par la population des agrobactéries naturellement présentes dans les substrats de culture.

D'un point de vue méthodologique, pour étudier les populations d'agrobactéries il est plus pertinent de multiplier les microcosmes analysés plutôt que le nombre d'isolats par microcosmes. Ce choix méthodologique devrait permettre d'avoir accès à la diversité totale des agrobactéries associées aux plantes analysées.

Il faut surtout faire des essais de piégeage en conditions gnotobiotiques avec un substrat inoculé avec une communauté d'agrobactérie de composition connue, afin de préciser les effets des plantes sur les populations d'agrobactéries.

## III. Spéciation bactérienne

Nous présentons ici, une méthode nouvelle pour déterminer, de façon massive, des marqueurs ADN impliqués dans la définition des espèces bactériennes. De plus, la disponibilité de la séquence complète de la souche type C58 (espèce G8), nous permet de localiser ces marqueurs sur le génome et de les faire correspondre à des gènes et des fonctions identifiés.

Mais surtout, cette approche est applicable à toutes les espèces de bactéries dès que la séquence complète d'un génome de référence est disponible. Il est possible pour toutes ces espèces de déterminer les fragments consensus marqueurs de l'espèce et de déterminer leur répartition sur le génome. Par contre, pour être efficace, cette approche doit être réalisée sur un minimum de souches (huit à dix) de la même espèce. Ces souches doivent être choisies avec soin pour présenter la plus grande diversité possible à l'intérieur de ces espèces. Ici aussi, l'AFLP est utile en première approche car elle permet de déterminer la distance génomique

évolutionnaire existant entre chaque souches analysées et de retenir l'échantillon de soucheprésentant la plus grande diversité.

L'application extensive de cette stratégie originale à diverses espèces bactériennes permettraient de déterminer si il existe des règles générales de répartition des marqueurs ADN impliqués dans la définition des espèces bactériennes. De plus, l'identification des gènes correspondants à ces marqueurs pourraient éventuellement apporter des indications sur les raisons écologiques impliquées dans la spéciation bactérienne.

## Annexes

# Annexe I

## Protocoles d'isolement, purification, tests biochimiques et conservation des agrobactéries

### I. 1. Préparation des extraits de sol et de racines

Après arrachage délicat des plants, ils sont doucement secoués afin de les débarrasser du sol non-adhérent.

Les racines sont ensuite coupées au niveau du collet et mises en agitation avec de l'eau stérile. Cette étape permet de nettoyer les racines et de récupérer le sol qui y adhère : le sol rhizosphérique. Ce sol rhizosphérique est ensuite broyé au waring blender ou au pilon.

Les racines sont rincées une dernière fois et découpées en morceaux avant d'être broyées à l'Ultra-Thurax. Cette étape de broyage permet d'accéder aux bactéries des tissus racinaires.

Le sol nu est traité de la même façon que le sol rhizosphérique, broyé dans de l'eau stérile.

### I. 2. Isolement des agrobactéries

Les extraits de sol et de racines sont ensemencés, par stries d'épuisement, sur les milieux 1A-Te et 2E-Te.

Le milieu 1A-Te permet de sélectionner les *Agrobacterium* de biovar1.

Le milieu 2E-Te permet de sélectionner *Agrobacterium* de biovar2 (*Rhizobium rhizogenes*).

### I. 3. Purification des isolats

Après 6 à 8 jours d'incubation à 28°C, les colonies de morphologie typique d'*Agrobacterium* sont transférées sur milieux MG+Tellurite.

Sur le milieu 1A-Te les *Agrobacterium* forment de grosses colonies rondes, bombées et cuivrées.

Sur le milieu 2E-Te les colonies de *Rhizobium rhizogenes* sont muqueuses, rondes bombées, et vertes.

Sur MG+tellurite les colonies d'*Agrobacterium* sont blanches-crèmes autour et noires au centre, rondes, bombées et parfois muqueuses. Le milieu ne devient jamais fluorescent (dans ce cas il s'agit de bactéries du genre *Pseudomonas*).

Les bactéries sont ensuite réisolées sur MG+tellurite ou MG trois fois de suite. Entre chaque isolement les bactéries sont purifiées 1h00 dans de l'eau stérile agitée.

#### I. 4. Conservation des *Agrobacterium*

En routine les *Agrobacterium* sont conservés à 4°C dans de l'eau stérile ou sur pentes de gélose de milieu LPGA ou MG en tube.

Les isolats sont conservés à -80°C en glycérol 50%. Le contenu d'une boîte de culture est raclé et remis en suspension, par agitation, une heure dans le glycérol 50% avant transfert au congélateur.

#### I. 5. PCR Vir

L'amplification d'une zone du plasmide Ti située entre les gènes *virG* et *virB11* confirme la présence d'un plasmide Ti dans la souche testée.

Programme PCR utilisé :

Hold	30 Cycles			Hold	Hold
95°C	95°C	55°C	72°C	72°C	15°C
7 min	1min	1min	1min	7min	Infini

Amorces utilisées :

F14 : GAACGTGTTTCAACGGTTCA

F749 : GCTAGCTTGGAAGATCGCAC

## Annexe II

### Milieux de culture et d'isolement pour *Agrobacterium*

#### II. 1. MG

D'après Keane *et al.* 1970 (PP178), modifié par Mougel *et al.* 2001 (PP10).

D-Mannitol 5 g.l<sup>-1</sup>, L(-)glutamique acid 2 g.l<sup>-1</sup>, KH<sub>2</sub>PO<sub>4</sub> 0,5 g.l<sup>-1</sup>, NaCl 0,2 g.l<sup>-1</sup>, MgSO<sub>4</sub>, 7H<sub>2</sub>O 0,2 g.l<sup>-1</sup> et extrait de levure 0,5 g.l<sup>-1</sup> et agar 18 g.l<sup>-1</sup>.

Avant de mettre l'agar ajuster le pH à 7,2. avec NaOH en pastilles (4 à 5 pastilles).  
Autoclaver 20' à 120°C.

Avant de couler les boites ajouter éventuellement : tellurite de potassium 200 mg. l<sup>-1</sup>.

#### II. 2. 1A-Te

D'après Brisbane et Kerr 1983 (PP207) modifié par Mougel *et al.* 2001 (PP10)

Spécifique pour *Agrobacterium* Biovar 1

L-Arabitol 3,04g, NH<sub>4</sub>NO<sub>3</sub> 0,16g.l<sup>-1</sup>, KH<sub>2</sub>PO<sub>4</sub> 0,54g.l<sup>-1</sup>, K<sub>2</sub>HPO<sub>4</sub> 1,04 g.l<sup>-1</sup>, MgSO<sub>4</sub> , 7 H<sub>2</sub>O 0,25g.l<sup>-1</sup>, taurocholate de Na 0,29g.l<sup>-1</sup>, Crystal violet 0,2 %, agar 15 g.l<sup>-1</sup>.

Autoclaver 20' à 120°C.

Avant de couler les boites ajouter : tellurite de potassium 80mg.l<sup>-1</sup>.

#### II. 3. 2E-Te

Dérivé de Brisbane et Kerr (Mougel 2001 PP10)

Spécifique pour *Rhizobium rhizogenes* (*Agrobacterium* Biovar 2).

Erythritol 3,05 g.l<sup>-1</sup>, NH<sub>4</sub>NO<sub>3</sub> 0,16g.l<sup>-1</sup>, KH<sub>2</sub>PO<sub>4</sub> 0,54g.l<sup>-1</sup>, K<sub>2</sub>HPO<sub>4</sub> 1,04 g.l<sup>-1</sup>, MgSO<sub>4</sub> , 7 H<sub>2</sub>O 0,25g.l<sup>-1</sup>, taurocholate de Na 0,29g.l<sup>-1</sup>, extrait de levure 0,01 g.l<sup>-1</sup>, Vert Malachite 0,02%, agar 15 g.l<sup>-1</sup>.

Autoclaver 20' à 120°C

Avant de couler les boites ajouter : tellurite de potassium (K<sub>2</sub>TeO<sub>3</sub>) 320mg.l<sup>-1</sup>.

## II. 4. LB - NaCl

Milieu de Luria-Bertani. Maniatis

Bacto-tryptone 10 g.l<sup>-1</sup>, extrait de levure 5 g.l<sup>-1</sup>.

Agar 15 g.l<sup>-1</sup>.

Autoclaver 20' à 120°C

## II. 5. LPGA

Bacto-peptone 5 g.l<sup>-1</sup>, extrait de levure 5 g.l<sup>-1</sup>, glucose 10 g.l<sup>-1</sup>, agar 15g.l<sup>-1</sup>.

Ajuster le pH à 7,2 avant d'ajouter l'agar.

Autoclaver 20' à 120°C

## Annexe III

### Protocoles et solutions pour les tests biochimiques

Les Agrobacterium répondent positivement aux tests de présence d'uréase et de  $\beta$ -glucosidase.

#### III. 1. Uréase

##### III. 1. 1. Milieu Urée-Indole

L-Tryptophane 0,03 g.l<sup>-1</sup>, KH<sub>2</sub>PO<sub>4</sub> 0,01 g.l<sup>-1</sup>, K<sub>2</sub>HPO<sub>4</sub> 0,01 g.l<sup>-1</sup>, NaCl 0,05 g.l<sup>-1</sup>, urée 0,2 g.l<sup>-1</sup>, ethanol 95° 10 ml.l<sup>-1</sup>, rouge de phénol 1% 2,5 ml.l<sup>-1</sup>.

Rouge de phénol 1%: phénosulfonephtaléine 1% dans l'éthanol 95°.

##### III. 1. 2. Préparation :

Faire fondre le Tryptophane à 80°C

Refroidir à 50°C et ajouter les autres produits.

Stériliser par filtration et aliquoter en tube veinoject stériles.

##### III. 1. 3. Test Urée-Indole :

En microplaque répartir 50µl de milieu

Mettre en suspension dedans 1 colonie à tester. Ne pas oublier le témoin négatif

Placer à 28°C

Si la réaction est positive le milieu vire au rose en 1/2h environ

Attention au bout de 2h à 28°C le témoin négatif vire aussi

Attention :

N'utiliser le milieu que s'il est jaune-orangé.

Une fois le tube ouvert l'éliminer

Le milieu Urée-Indole se conserve environ 1 an



## III. 2. $\beta$ -glucosidase

### III. 2. 1. Milieu à l'esculine:

Bacto Peptone 10 g.l<sup>-1</sup>, citrate de fer ammoniacal 1 g.l<sup>-1</sup>, esculine 1 g.l<sup>-1</sup>.

Ajuster le pH à 7,2 puis autoclaver 20' à 120°C.

Répartir en tube veinoject stériles.

### III. 2. 2. Test Esculine :

En microplaque répartir 50 $\mu$ l de milieu

Mettre en suspension dedans 1 colonie à tester. Ne pas oublier le témoin négatif

Placer à 28°C

Si la réaction est positive le milieu vire au noir en 2h environ

Laisser une nuit à 28°C pour laisser le temps à toutes les souches de réagir.

Attention :

N'utiliser le milieu que s'il est gris moiré violet.

Une fois le tube ouvert l'éliminer

Le milieu Esculine se conserve environ 1 an

## III. 3. Production de 3-cétolactose

Les *Agrobacterium* de biovar 1 ont la capacité de produire du 3-cétolactose.

### III. 3. 1. Milieux

#### *III. 3. 1. a. Milieu lactose*

Lactose 10 g.l<sup>-1</sup>, extrait de levure 1 g.l<sup>-1</sup>, agar 20 g.l<sup>-1</sup>.

#### *III. 3. 1. b. Réactif de Benedict*

(pour révélation)

Se prépare en deux solutions

Solution A

Citrate de Sodium 203 g.l<sup>-1</sup>, carbonate de sodium anhydre 1118 g.l<sup>-1</sup>.

Attention:

Dissoudre les poudres par chauffage dans 3/4 du volume d'eau final puis compléter.

Solution B

Sulfate de cuivre 120 g.l<sup>-1</sup>.

Attention:

Dissoudre dans 3/4 du volume d'eau final puis compléter.

Ajouter doucement la solution B à la solution A

### **III. 3. 2. Test de production de 3-cétolactose**

#### *III. 3. 2; a. Culture*

Ensemencer des boîtes de milieu Lactose (en patch, on peut tester plusieurs souches par boîtes) avec les souches à tester.

#### *III. 3. 2; b. Révélation*

Après 2 jours de culture à 37°C recouvrir délicatement le dessus de la boîte avec le réactif de Benedict.

Au bout de 10' (pour les plus rapides) à 2h00, les souches produisant du 3-cétolactose font virer le milieu du bleu au jaune.

## **Annexe IV**

### **Protocole AFLP**

**(Amplified Fragment Length Polymorphism)**

#### **IV. 1. Extraction d'ADN**

Les souches à analysées sont cultivées 48h en milieu LB-NaCl liquide  
L'ADN de ces souches est extrait avec le kit DNeAsy Tissue Kit (Qiagen)

##### **IV. 1. 1. Lyse des cellules**

- 1\* Mettre 2ml de la culture dans un eppendorf
  - 2\* Centrifuger 5 minutes à 13000 rpm et jeter le surnageant
  - 3\* Ajouter 1 ml de TE8 et resuspendre
- Laisser 10 minutes à température ambiante
- 4\* Centrifuger 5 minutes à 13000 rpm et jeter le surnageant
  - 5\* Ajouter par échantillon: 75 µl de Proteinase K (1 mg.ml<sup>-1</sup>)
    - 50 µl de Sarcosyl (5 g/100 ml = 5%)
    - 375 µl d'eau

(Volume total à ajouter par échantillons : 500µl)

- 6\* Placer en agitation horizontale 1h30 sur la paillasse à température ambiante.
- Laisser les tubes immobiles le reste de la nuit.

##### **IV. 1. 2. Extraction d'ADN**

A partir de l'étape 3 de la page 17 du livret explicatif du DNeAsy Tissue Kit

Prévoir un bain-marie à 70°C

- 1\* Dans les cellules lysées sur la nuit :  
Ajouter 200 µl de tampon AL puis vortexer
  - 2\* Incubation 10 minutes à 70°C
  - 3\* Ajouter 200 µl d'Ethanol (96 – 100%) puis vortexer
- Pipeter le mélange dans une minicolonne (fournie)
- 4\* Centrifuger 1 minute à 8000 rpm (si ça suffit pas centrifuger plus longtemps et plus vite. Le mucus des bactéries peut gêner les centrifugations)

Jeter le tube et le liquide

5\* Placer la colonne dans un nouveau tube (fourni) et ajouter 500 µl de tampon AW1

Centrifuger 1 minute à 8000 rpm. Jeter le tube et le liquide

6\* Placer la colonne dans un nouveau tube (fourni) et ajouter 500µl de tampon AW2.

Centrifuger 3 minutes à 13000 rpm. Jeter le tube et le liquide

7\* Elution : Placer la colonne dans un eppendorf 1,5 ou 2 ml neuf (non fourni).

Ajouter 150 µl de tampon AE et incuber 1 minute à température ambiante.

Centrifuger 1 minute à 8000 rpm.

8\* répéter l'étape d'élution avec 100 µl et dans le même tube.

## IV. 2. Dosage de l'ADN

L'ADN des souches est dosé soit par analyse de gel soit en spectrophotométrie puis ramené à 10ng.µl<sup>-1</sup>.

## IV. 3. Digestion - ligation

Pour un échantillon :

ADN (10 ng.µl <sup>-1</sup> )	5,5µl	
Tampon T4 DNA ligase (10X)	1,1µl	
NaCl (0,5 M)	1,1µl	
H <sub>2</sub> O UP Stérile	0,4 µl	
BSA (1mg.ml <sup>-1</sup> )	0,5µl	
Adaptateur Eco (2µM)	1µl	
Adaptateur Mse (20µM)	1µl	
EcoRI (50U.µl <sup>-1</sup> )	0,1µl	→ 5U / échantillon
MseI (50U.µl <sup>-1</sup> )	0,1 µl	→ 5U / échantillon
T4 DNA ligase (5 U.µl <sup>-1</sup> )	0,2 µl	→ 1 U / échantillon

✳Les adaptateurs doivent être dénaturés avant la première utilisation.

Une fois le mélange fait les porter 5' à 95°C (par exemple dans le thermomixer)

Puis les laisser revenir à température ambiante mais très lentement, en 1h au minimum (régler le thermomixer sur 25°C et le laisser refroidir. Cela prend 1h40 ou plus).

Volume final: 11µl (5,5 µl d'ADN + 5,5 µl du mélange)

Placer deux heures à 37 °C. Les tubes doivent être statiques (sinon la ligation ne se fait pas).

Pour arrêter la réaction mettre au frigo

#### IV. 4. PCR pré-sélective

Pour un échantillon :

Produit de la Digestion-Ligation non dilué	4µl
Core Mix	15µl
Amorce F1247CoreMse (10µM)	0,5µl
Amorce F1248CoreEco (10µM)	0,5µl

Volume final: 20µl (4µl d'ADN + 16µl de mix)

La PCR s'effectue en micro-tubes

Programme PCR :

Hold	20 Cycles			Hold
94°C	94°C	56°C	72°C	20°C
2 minutes	20"	30"	2 min	Infini

Vérification de la PCR :

Passage des produits PCR sur gel agarose 1,5% + 8µl BET/100ml

→ Visualiser la qualité de la PCR et la quantité de produits obtenus

On obtient un smear.

Selon l'intensité de ce smear les produits PCR sont dilués entre 10 et 40 fois.

Le séquenceur (à capillaires) étant très sensible à la quantité d'ADN celle-ci doit être la plus faible possible.

Il est donc très important de beaucoup diluer les produits de la PCR pré-sélective, par contre diluer les produits de la PCR sélective (dernière étape avant le séquenceur) ne permet pas, au contraire d'obtenir de bons résultats.

#### IV. 5. PCR sélective

Pour un échantillon :

Produits PCR Préselective dilué	1,5µl
Core Mix	7,5µl
F1247Core Mse (5µM)	0,5µl

Amorce Eco Marquée (1µM)

0,5µl

Volume final: 10µl (1,5µl d'ADN + 8,5µl de mix)

La PCR s'effectue en micro-tubes

Programme PCR :

Hold	10 Cycles			20 Cycles			Hold	Hold
94°C	94°C	66→57°C	72°C	94°C	56°C	72°C	60°C	20°C
2 min	20"	30"	2min	20"	30"	2min	30min	Infini

Les produits de cette PCR sont déposés au séquenceurs.

Le séquenceur pouvant détecter 4 couleurs différentes il est possible de pooler les réactions.

#### IV. 6. Amorces utilisées

##### Adaptateur Eco

5' > 3'

F1363AdEco+ CTCGTAGACTGCGTACC

F1931AdEco- AATTGGTACGCAGTCTAC

##### Adaptateur Mse

5' > 3'

F1365 AdMse+ GACGATGAGTCCTGAG

F1932 AdMse- TACTCAGGACTCAT

##### Amorces non sélectives

5' > 3'

F1247 CoreEco GACTGCGTACCAATTC

F1248 CoreMse GATGAGTCCTGAGTAA

##### Amorces sélectives fluorescentes (marquage en 5')

5' > 3'

F1598 EcoCAFAM GACTGCGTACCAATTCCA Fluorochrome FAM (bleu)

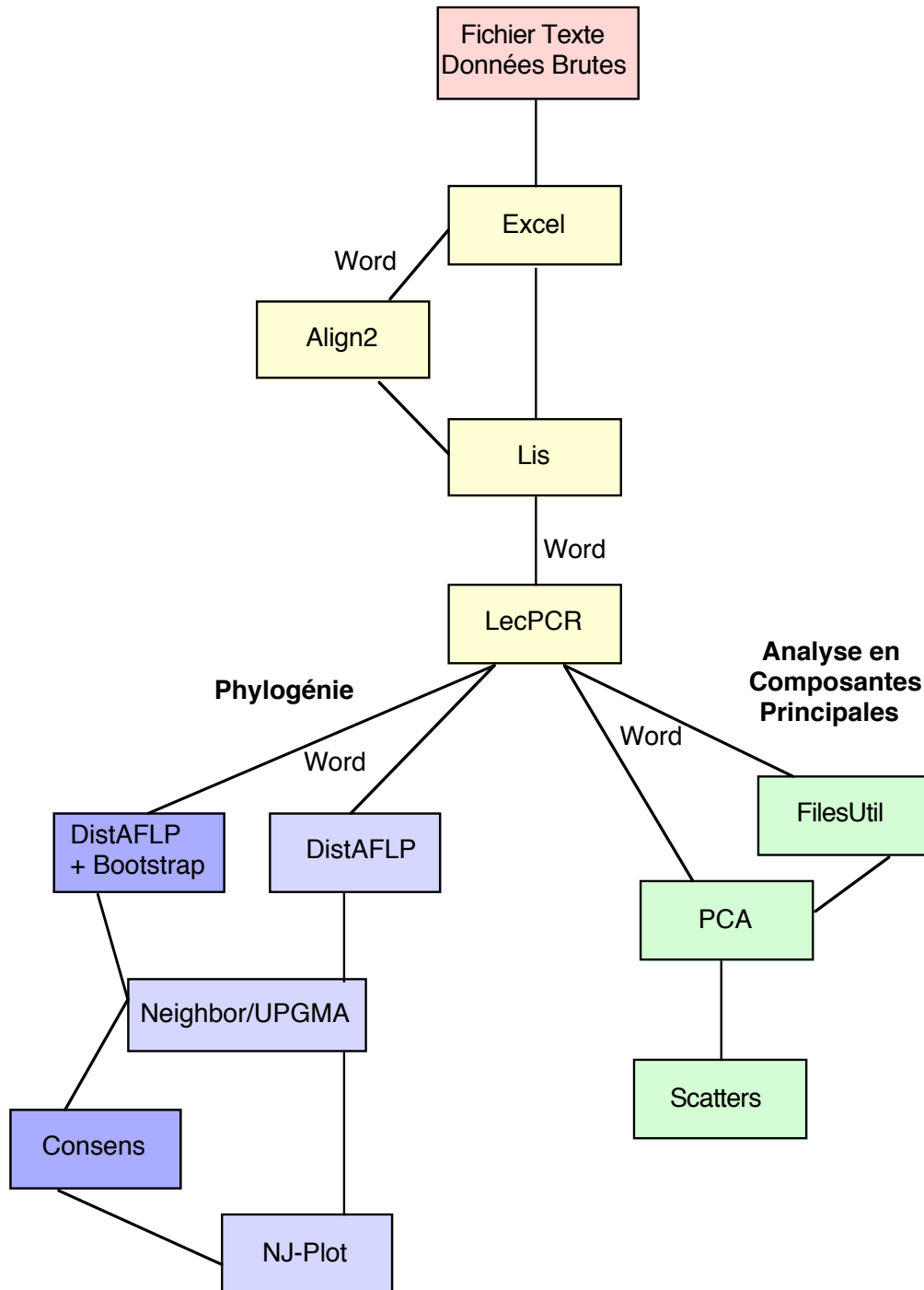
F1599 EcoCCHEx GACTGCGTACCAATTCCC Fluorochrome HEX (vert)

F1915 EcoCTFAM GACTGCGTACCAATTCCT Fluorochrome FAM (bleu)

F1601 EcoCGHEX GACTGCGTACCAATTCCG Fluorochrome HEX (vert)

# Annexe V

## Analyse des données AFLP



## Liste des logiciels utilisés pour l'analyse des résultats AFLP :

**Excel** pour traiter les données brutes et les mettre en forme

**Word** pour mettre les données en forme.

**Align2** pour comparer deux réplifications

**Lis** pour attribuer les fragments à des catégories

**LecPCR** pour transformer les données de tailles en données binaires.

**DistAFLP** pour calculer les divergences des génomes

**Neighbor/UPGMA** pour calculer les arbres phylogénétiques

**Consens** pour avoir les valeurs de bootstrap correspondantes aux embranchements

**NJ-plot** pour visualiser les arbres phylogénétiques

**ADE-4** : logiciel de statistique des populations

**AlignK** : marche dans R. Permet l'analyse simultanée de nombreuses analyses

DistAFLP, Align2 et LecPCR : logiciels disponibles à l'adresse suivante :

<http://pbil.univ-lyon1.fr/ADE-4/microb/>

NJ-Plot et ADE-4 sont disponibles à l'adresse suivante :

<http://pbil.univ-lyon1.fr/>

Neighbor/UPGMA et Consens sont des modules du package PHYLIP disponibles à l'adresse suivante :

<http://evolution.genetics.washington.edu/phylip.html>

## V. 1. Mise en forme

Excel

Mise en Forme

Ouvrir le fichier données brutes (format txt) avec Excel.

Remplacer tous les points par des virgules.

Mise en forme pour Lis

Chaque souche occupe une colonne. La première ligne correspond au nom de la souche et les fragments suivent par ordre croissant de taille.

Attention bien vérifier que les points sont remplacés par des virgules.

Mise en forme pour Align2

Copier la liste, sans le nom de la souche, de fragments

Word

Coller la liste des fragments avec collage spécial texte sans mise en forme.

Enregistrer sous le nom de la souche et en format texte seulement.



Attention bien vérifier que les virgules soient remplacées par des points.

Plus généralement Word sert tout au long de l'analyse pour remettre les données en forme. Toujours enregistrer en format texte seulement.

## V. 2. Comparaisons de duplications

### Align2

Ce logiciel permet de comparer deux analyses, repère les fragments identiques aux deux et leur attribue une valeur moyenne.

A partir d'Excel copier la liste des fragments de la première répétition.

Aller dans Word et faire collage spécial, texte sans mise en forme.

Ne pas mettre le nom de la liste dans le fichier (c'est donc une liste de numeros)

Enregistrer en format texte seulement

Faire de même pour la deuxième liste à comparer.

Ensuite ouvrir Align2

Aller dans Edit puis Préférences

Permet de définir où sont localisés les fichiers à analyser. (il faut qu'ils soient au même endroit)

Rentrer les deux noms de fichier dans Align2 (à l'aide des « mains »)

Cliquer sur aligner. Un nouveau fichier est créé. L'ouvrir avec Word ou Excel.

La première colonne correspond au premier fichier.

Si un fragment est présent dans les deux fichiers le logiciel indique la valeur moyenne entre les deux (Align2 analyse avec un pas de 1). Si le fragment n'est présent que pour un seul des fichiers il indique 0.

## V. 3. Formatage des données

### Lis

Ce logiciel permet d'attribuer une catégorie à chaque fragment. En données brutes la taille des fragments est donnée sous forme décimale.

Lis permet de remplacer ces tailles décimales en nombre entiers et en fonction de toutes les données analysées en même temps.

Dans la page **Données** de Lis :

Coller toutes les données. La première ligne correspond au nom de l'échantillon.

La première colonne de cette feuille (colonne A n'existe pas)

Repérer les coordonnées maximales du tableau, le nombre de souches et le nombre max de fragments

Dans la page **Paramètres** de Lis (paramètres du calcul)

Le Vernier correspond au pas d'analyse. A régler sur 0,75 ou 1

Dans les limites du tableau mettre les coordonnées maxi de la page d'avant.

Dans la page **Calculs** de Lis dans la case A1 dans la formule rentrer les coordonnées maxi de la feuille Données. Puis taper entrer.

Faire de même dans la feuille **Catégories**.

Revenir dans la feuille **Paramètres**.

Pour calculer taper 1 puis entrer dans la case initialisation.

Les données sont transférées dans la feuille calculs.

Ensuite taper 0 dans la case initialisation. Les calculs commencent.

Aller dans la feuille **Calculs**.

Problemes :

Lis Transforme toutes les données de taille de fragments en 0 :

Solution : remplacer les points par des virgules dans la feuille données. Refaire le transfert des données.

Lis refuse de calculer ou parle de formules en boucles :

Aller dans préférences, calcul et permettre les itérations (100 au minimum) ou faire commande + = pour amorcer le calcul. Quand le calcul (au bout des itérations) est fini vérifier que tout est calculé. Sinon relancer le calcul (commande + =).

Une fois les calculs terminés les données sont transférées dans la feuille catégories.

Aller dans la feuille **Catégories**.

Reprendre toutes les données et les copier

Aller dans la feuille **Pour transposer les données**.

Faire collage spécial, valeurs, transposé.

Les données sont en forme. Il n'y a plus qu'à les reprendre (sauf la ligne 1), les copier et aller dans word.

Word

**Mise en forme des données pour Lec PCR.**

Dans un nouveau document faire collage spécial et coller comme texte sans mise en forme. Enregistrer en format texte seulement.

Enlever les 0 en trop (à la fin des données pour chaque souches) :

- Remplacer les tabulations suivies de 0 ("^t0) par rien
- Ensuite remplacer les tabulations (^t) par des espaces

Vérifier les noms des souches.

Les noms ne doivent pas comporter de signes cabalistiques (tels que / - \_ etc...) ou des espaces. De plus ils ne doivent pas dépasser les 10 caractères.

Enregistrer en format texte seulement.

## V. 4. Calculs

### LecPCR

LecPCR transforme les données chiffres en données binaires. Il regarde tous les fragments existants dans le fichier et pour chaque souche il indique la présence du fragment par 1 et son absence par 0. Le résultat de LecPCR dépendra donc des souches analysées en même temps. Il n'est pas possible d'analyser des souches avec LecPCR en plusieurs fois.

Aller dans **edit/préférences** et choisir la localisation du fichier

Case **Input File** : à l'aide de la main, entrer le fichier à analyser.

Case **Output File** : rentrer un nom de fichier de résultat (Nomresultats)

Case **ADE-4 Output File** : mettre 1 pour oui.

Clicker sur OK

Quand les calculs sont terminés quitter et ne pas sauver.

4 fichiers sont créés :

Nomresultats.labf : noms des fragments

Nomresultats.labs : noms des souches

Nomresultats : résultats binaires

Nomresultatsb : résultats binaires en format ADE-4

### V. 4. 1. Pour dessiner des arbres phylogénétiques

#### DistAFLP

Ce logiciel calcule les distances génomiques entre les souches et produit une matrice de distance.

Dans **edit/preferences** localiser le fichier à analyser. Il s'agit du fichier nomrésultats généré par LecPCR

Puis entrer le nom du fichier à analyser dans la case **Input File**.

Case **Restriction Site Length** : entrer le nombre de nucléotides impliqués dans l'AFLP (= le nombre du site de l'enzyme de restriction 1+ le nombre pour l'enzyme 2 + le nombre de nucléotides discriminants)

Ex : AFLP avec MseI et EcoRI et 2 nucléotides discriminants = 4+6+2 = 12

Case **Kvalue** : cette valeur est un facteur de correction pour tenir compte de l'homoplasie (probabilité que dans un même génome deux fragments différents aient la même taille. Il est à noter que l'homoplasie augmente avec la distance génomique. Les valeurs de divergence des génomes sont donc exploitables en intraspécifique ou entre des espèces très proches mais pas pour des espèces éloignées)

Cette valeur doit être estimée (à l'aide des séquences disponibles) pour chaque genre.

Pour *Agrobacterium* **K = 0.89** (attention DistAFLP ne reconnaît que les points et pas les virgules).

Case **Index** : mettre l'indice voulu pour les calculs. (Souvent c'est Dice : 2)

Case **Sqrt or Ln** : mettre 1 (avec l'indice de Dice)

Case **Jukes-Cantor correction** : 1

Case **Sorties ADE-4** : ce n'est pas nécessaire. Donc ne rien mettre.

Cliquer sur OK puis quitter quand les calculs sont terminés. Ne pas sauvegarder.

Si le programme répond qu'il n'a pas assez de mémoire

Vérifier que le fichier source contient bien le nombre de souches, suivi du nombre de fragments sur la première ligne et vérifier qu'il est bien enregistré en format texte seulement.

En utilisant l'indice de Dice DistAFLP produit les 4 fichiers suivants :

Nom.matD

Nom.matDJC

Nom.matDP

Nom.nxy

Le fichier nom.nxy contient les valeurs de divergence des génomes

Le fichier **Nom.matDJC** permet de calculer l'arbre phylogénétique.

Placer celui-ci dans le dossier qui contient le logiciel Neighbor/UPGMA

Neighbor/UPGMA

Le logiciel Neighbor/UPGMA du package Phylip permet de calculer l'arbre phylogénétique à partir des données de DistAFLP.

Taper le nom du fichier à analyser. Attention ce qu'on tape n'apparaît pas. C'est déstabilisant. Bien respecter les majuscules et minuscules.

Si le logiciel n'est pas d'accord vérifier que le fichier à analyser est bien dans le même dossier.

Eventuellement taper **n** puis entrer (arbre UPGMA).

Taper **r** puis entrer (upper triangular data matrix).

Taper **j** puis entrer (randomize input order of species)

Rentrer ensuite un nombre impaire puis taper entrer.

Ensuite taper **y** puis entrer. Les calculs démarrent.

2 fichiers sont créés treefile et outfile

NJ-Plot

Ce logiciel permet de visualiser l'arbre calculé.

Ouvrir le fichier treefile avec.

Puis copier l'arbre et le coller dans Claris pour le sauvegarder.

#### **V. 4. 2. Pour calculer les valeurs de bootstrap**

Après avoir calculé l'arbre revenir à DistAFLP

DistAFLP

Aller dans **Option**

Sélectionner **Bootstrap computation**

Remplir les cases comme précédemment.

Case **number of bootstrap** : indiquer le nombre de bootstrap voulu (100 ou plus)

Lancer les calculs

Neighbor/UPGMA

Ouvrir le fichier avec l'extension .matDJC

En plus des settings habituels faire **m** et entrer la valeur de bootstrap (100) (c'est mieux en UPGMA).

2 fichiers sont créés : outfile et treefile.

Renommer ces fichiers

Consens

Ensuite avec Consens ouvrir le treefile.

Ne pas changer les settings et lancer les calculs. Consens génère un fichier treefile et un fichier outfile. Ouvrir le fichier outfile.

Consens donne un arbre non raciné avec des valeurs de bootstraps. Les valeurs sont à replacer sur l'arbre calculé précédemment.

### V. 4. 3. Pour analyser plusieurs amorçages en même temps

(phylogénies plus complètes)

Pour une même souche il faut coller bout à bout les différents jeux de données en format 0/1 issus de LecPCR.

Ex :

Amorçage CA

1 45

C58

001100110001100001110001001001010000100101101

Amorçage CC

1 38

C58

00111100001110110001110000111110000111

Ces deux jeux de données deviennent :

CA/CC

1 83

C58

0011001100011000011100010010010100001001011010011110000111011000111  
0000111110000111

(En rouge les données CA en noir les données CC.)

Attention mettre les données dans le même ordre pour toutes les souches.

Un conseil : mettre toujours les données dans le même ordre pour tout type d'analyse, des PCR aux fichiers informatiques. Ça facilite les choses.

Ne pas oublier de modifier en haut du fichier le 2° nombre : il correspond aux nombre de caractères (le 1° correspond au nombre de souches).

### V. 4. 4. Pour analyser les données en Analyse en Composantes Principales

A partir des fichiers binaires générés par LecPCR

ADE-4-Compute / One table analysis / PCA

Le module PCA de ADE-4 permet de calculer les ACP

Dans **options** : choisir covariance matrix PCA

Puis **edit/preferences** :localiser le fichier à analyser.

Il s'agit du fichier **nomresultatsb** crée par LecPCR.

Puis cliquer sur OK et c'est parti !

Une fenêtre de résultats s'ouvre. Cliquer sur OK. Un graphique apparaît et le logiciel demande le nombre d'axes à conserver. Les axes les plus importants sont souvent les 2 ou 4 premiers. En fait il faut prendre ceux qui sont placées avant une rupture dans la « pente ».

Une fois le choix fait le logiciel redonne les eigenvalues (valeurs de % d'information attribuées à chaque axe). Noter les valeurs ppour les axes retenus.

PCA génère (dans cette utilisation) 8 fichiers.

Nomb.cpc0

Nomb.cpli

Nomb.cpma

Nomb.cppa

Nomb.cppc

Nomb.cppl

Nomb.cpta

Nomb.cvpv

Pour analyser plusieurs amorçages en même temps :

Il faut fusionner les fichiers binaires générés par LecPCR et faire une liste unique avec les noms de fragments.

Word

Pour les noms de fragments.

Ajouter pour chaque fragment (repéré par sa taille) un nom qui permet de l'identifier.

Ex : CA310 pour le fragment de 310pb détecté avec l'amorçage CA.

Coller les différentes listes à la suite (dans l'ordre) dans un fichier unique.

Enregistrer sous format texte seulement.

ADE-4-Compute / Files utilities / FilesUtil

Ce module de ADE-4 permet de fusionner les fichier binaires ADE-4.

Dans **Option** sélectionner **PasteFiles-SameRow**.

Cette option permet de coller toutes les données pour une même souche comprises dans différents fichiers. Attention toutes les souches doivent être dans le même ordre dans tous les fichiers.

Case **input Bin File 1** : entrer le nom du premier fichier à fusionner.

Puis entrer le nom du deuxième fichier dans la deuxième case...

Attention : entrer les fichiers dans le même ordre que les noms des fragments (ex : CA en premier puis CC puis CT...)

Case **output File** : entrer le nom du fichier de résultats.

Se servir ensuite de ce fichier pour calculer les ACP.

ADE-4-Graph / Scatters / Scatters

Le module Scatters de ADE-4 permet de visualiser les résultats d'ACP.

L'ACP met les données dans un tableau.

En lignes sont portées les souches et en colonnes les fragments.

Pour faire une ACP sur les **souches** il faudra donc utiliser comme fichier source :

**Nomb.cpli**. Pour une ACP sur les **fragments** il faut utiliser le fichier **Nomb.cpc** comme fichier source.

Case XY **coordinate file** : entrer le fichier source.

Case **X-Axis column number** : entrer l'axe à représenter en X

Case **Y-Axis column number** : entrer l'axe à représenter en Y.

Par défaut le logiciel représente les axes 1 et 2.

Case Label file for items : entrer le nom fichier contenant le nom soit des souches soit des fragments. (Nomresultat.labs pour les souches et nomresultats.labf pour les fragments)

Cliquer sur **Draw** et le graphique apparaît.

Aller dans **windows/min et max** pour modifier les limites du graphique.

#### **V. 4. 5. Classification automatique**

Dans cette analyse les groupes de souches sont connus au préalable. Cette analyse permet d'attribuer les fragments AFLP à des groupes de souches précis.

Après une ACP classique une ACP interclasses est effectuée. Puis sur ces résultats on calcule une classification automatique en UPGMA. Ainsi chaque fragment est associé automatiquement à un groupe taxonomique.

##### *V. 4. 5. a. Préparation des fichiers :*

- Le fichier binaire de la matrice de résultats en 01 issu de LecPCR Pour compiler des données (par exemple compiler les données AFLP issues de 4 amorçages)

Il faut le fichier binaire issu de LecPCR pour chacun des amorçages.



Attention : dans ces fichiers il doit y avoir toujours le même nombre de souches et dans le même ordre.

Ouvrir ADE-4 compute.

Dans module selectionner FilesUtilities / FilesUtil

Option : pastefile same row.

Entrer dans l'ordre (grace au bouton set) les fichiers a fusionner

Donner un nom pour le fichier output et valider.

- Les fichiers avec les noms des souches et des fragments

Fichiers .labf et .labs issus de LecPCR (voir fichier analyse des résultats AFLP)

- Préparer un fichier de catégories

Ce fichier comporte dans l'ordre (un n° par ligne) le n° de la catégorie attribuée à chaque souche.

Exemple :

Fichier .labs :	Fichier catégories correspondant :
ATCC4720	1
TT111	1
NCPPB396	1
S377	1
S56	1
S4	1
CFBP5622	1
CFBP2712	1
CFBP2517	1
CFBP5771	1
M2/1	4
CIP497-74	4
CIP28-75	4
CIP43-76	4
CFBP6623	4
CFBP6624	4
B6	2
ATCC4452	2
...	...

Enregistrer ce fichier en format texte.

#### *V. 4. 5. b. Préparation des fichiers de catégorie*

Ouvrir ADE-4compute

- Dans modules sélectionner files utilities / TextToBin

Input file : sélectionner le fichier des catégories

Ce module permet de transformer les fichiers textes en fichiers binaires

- Dans modules sélectionner files utilities / CategVar

Dans option sélectionner read CategFile

Input file : sélectionner le fichier binaire de catégories issu de l'étape précédente

Mettre dans le 2° champs 1 pour complete output (plusieurs fichiers sont créés).

Ce module permet au logiciel de préparer les fichiers de catégories. Ils seront utiles dans les étapes suivantes.

#### *V. 4. 5. c. ACP simple*

Toujours dans ADE-4 compute

- Dans module sélectionner One table analyses / PCA

Option : covariance matrix PCA

Grace au bouton set sélectionner le fichier binaire issu de LecPCR en input file

Valider.

Plusieurs fichiers sont créés dont le fichier en .cpta

Entrer le nombre d'axes à conserver.

#### *V. 4. 5. d. ACP inter*

Toujours dans ADE-4 compute

- Dans modules sélectionner sélectionner One table analyses / Discrimin

Option Initialize : Link Prep

Avec le bouton set sélectionner en statistical triplet (input file) le fichier en .cpta généré par l'ACP simple.

De même sélectionner le fichier de catégories, fichier en .cat

Entrer un nom de fichier output et valider

Cette étape génère plusieurs fichiers don un fichier en .dis

Option between analyses test

En input file sélectionner le fichier en .dis

Entrer un nombre de permutations. Attention ce nombre doit être bien supérieur au nombre total de caractères analysés (par exemple pour 500 caractères 1000 permutations sont nécessaires, pour 2000 caractères 10 000 permutations sont nécessaires). Si le nombre de permutation n'est pas suffisant ADE-4 donne des trucs bizarres.

Valider.

Cette étape permet de tester la significativité de l'analyse. Jusqu'à 5% c'est bon. Au-delà ce n'est pas la peine de continuer l'analyse.

Option between analyses run

En input file sélectionner le fichier en .dis

Entrer le nombre d'axes à conserver.

Cette étape génère plusieurs fichiers

Fichier en .beco contient les informations sur les colonnes (les fragments)

Fichier en .beli contient les informations sur les lignes (les catégories)

Fichier en .bels contient les informations sur les souches

#### *V. 4. 5. e. Classification automatique en UPGMA*

Toujours dans ADE-4 compute

- Dans modules sélectionner Cluster analysis / Cluster

Option Compute distances

En input file sélectionner le fichier en .beco

En type de distance sélectionner la distance Euclidienne (1)

Est créé un fichier en .beco.dist

Option Compute hierarchy distance method

En input file sélectionner le fichier en .beco.dist

Comme algorithme sélectionner UPGMA (2)

Est créé un fichier en .beco.alha

Option prepare convex hulls

Convexe hulls : enveloppes convexes.

En input file sélectionner le fichier en .beco.alha

Entrer le nombre de niveaux hiérarchiques souhaités

A cette étape plusieurs fichiers sont créés dont un fichier en .beco-dend et un en .beco-dend.cat

#### *V. 4. 5. f. Visualisation des résultats*

Ouvrir ADE-4 Graph

- Dans modules sélectionner Scatters / scatterclass

Option : selon le graph désiré

Convex hulls : délimite précisément les groupes

Stars : trace des traits entre le point central du groupe et chacun des individus de ce groupe

Ellipses : trace des cercles autour des groupes

Labels : graphique avec seulement les points

Visualisation des résultats souches

XY coordinate file, sélectionner, avec le bouton set, le fichier en .bels

Category file, sélectionner le fichier en .cat (celui obtenu en début d'analyse à l'étape 2).

Label file for items, sélectionner le fichier .labs

Valider.

Ouvrir Window / Min&Max et redimensionner la fenêtre pour qu'elle soit carrée.

Pour superposer 2 options (par exemple Convex hulls et Star) il faut ouvrir une deuxième fenêtre graphique (Window / New graph window) et faire le deuxième graphique. Le redimensionner de la même façon que le premier. Le menu Edit permet de superposer les deux graphiques dans la même fenêtre.

Visualisation des résultats fragments

Option convex hulls

XY coordinate file, sélectionner, avec le bouton set, le fichier en .beco

Category file, sélectionner le fichier en .beco-dend.cat

Label file for items, sélectionner le fichier .labf

Valider.

Le graphique comporte autant de petits graphiques que de niveaux hiérarchiques.

Pour visualiser un seul de ces graphiques :

Dans Window / Row and column selection

Dans la partie row selection activer le signet File

Sélectionner le fichier en .beco-dend.cat

Entrer le niveau hiérarchique à visualiser dans le cadre suivant

Valider

Redimensionner la fenêtre pour la rendre carrée

Si les graphiques fragments sont illisibles à cause de la présence de nombreux numeros entourés cela veut dire que le nombre de permutations à l'étape 4b (between analysis test de l'ACP interclasse) n'est pas suffisant.

#### *V. 4. 5. g. Obtenir le fichier de résultats*

Ouvrir ADE-4 compute

Dans Modules sélectionner Files utilities / Bintotext

En Binary input file sélectionner le fichier en .beco-dend

En output file donner un nom différent

Valider.

Le fichier résultant donne pour chaque fragment (en ordre, en lignes) le groupe auquel il appartient pour chacun des niveaux hiérarchique (en colonne)

Ouvrir ce fichier dans Excel. Insérer une colonne au début et dans cette colonne copier le fichier .labf

Si le fichier donne des numeros avec décimales et tous les mêmes partout c'est que le nombre de permutations à l'étape 4b n'est pas suffisant.

### **V. 5. Comparaison de plusieurs analyses en même temps**

#### **AlignK**

Format de fichier

Le fichier de données doit s'appeler don.txt

Il doit être sous forme du nom de la souche (sans aucun caractère bizarre) suivi de la suite des fragments. Attention remplacer toutes les virgules par des points.

Puis l'ouvrir avec TextEdit. Vérifier qu'il n'y a rien de bizarre (pas de retour chariot en trop ou autre...) Ne pas se fier à Word.

Calculs

Ouvrir R

Vérifier (dans tools) la localisation du fichier de données  
(faire choose working directory)

Taper la commande :

```
Source( 'alignK.R' )
```

Puis entrer

C'est fait

Si ça ne marche pas remplacer les (') par (« )

Résultats

Le fichier de résultats s'appelle sor.txt

L'ouvrir avec excel

Type de fichier : délimité

Séparateur : espace

Puis dans le fichier remplacer :

Les 0 par rien mais attention en cellules entières

Les points par des virgules (décocher la case cellules entières)

Et voilà y'a plus qu'a s'amuser avec !!!

## Références bibliographiques

## References bibliographiques

- **Alexandre G., Rohr R. Bally R.** (1999). A phase variant of *Azospirillum lipoferum* lacks a polar flagellum and constitutively expresses mechanosensing lateral flagella. *Appl. Environ. Microbiol.* 65: 4701-4704
- **Arnold C., Metherell I., Willshaw G., Maggs A. Stanley J.** (1999). Predictive fluorescent amplified-fragment length polymorphism analysis of *Escherichia coli*: high-resolution typing method with phylogenetic significance. *J. Clin. Microbiol.* 37: 1274-1279
- **Beijerinck M.W. Van Delden A.** (1902). Über die assimilation des freien stickstoffs durch bakterien. *Zentralbl. Bakteriol. Parasitenk. Infektionskr. Hyg. Abdt. li:* 3-43.
- **Beneddra T., Picard C., Petit A., Nesme X.** (1996). Correlation between susceptibility to crown gall and sensitivity to cytokinin in aspen cultivars. *Phytopathology* 86:225-231.
- **Bevan M.** (1984). Binary *Agrobacterium* vectors for plant transformation. *Nuc ac res.* 12: 8711-8721
- **Blears M.J., De Grandis S.A., Lee H. Trevors J.T.** (1998). Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. *J. Ind. Microb. Biotechnol.* 21: 99-114
- **Bojesen A.M., Torpdhal M., Christensen H., Olsen J.E. Bisgaard M.** (2003). Genetic diversity of *Gallibacterium anatis* isolates from different chicken flocks. *J. Clin. Microb.* 41: 2737-2740
- **Bouzar H., Chilton W.S., Nesme X., Dessaux Y., Vaudequin V., Petit A., Jones J.B. Hodge N.C.** (1995). A new *Agrobacterium* strain isolated from aerial tumors on *Ficus benjamina*. *Appl. Environ. Microbiol.* 61: 65-73
- **Bouzar H. Jones J.B.** (2001). *Agrobacterium larrymoorei* sp. Nov., a pathogen isolated from aerial tumours of *Ficus benjamina*. *Int. J. Syst. Evol. Microbiol.* 51: 1023-1026.
- **Boyer F., L'Hour J.** (2001). La genèse. In "La bible". F. Boyer (Eds). Bayard et Media Paul, Paris et Montreal.
- **Brisbane P.G., Kerr A.** (1983). Selective media for the three biovars of *Agrobacterium*. *J. Appl. Bacteriol.* 54: 425-431.
- **Brown J.R.** (2003). Ancient horizontal gene transfer. *Nat. Rev. Genet.* 4: 121-32.



- **Bruce K.D., Hiorns W.D., Hobman J.L., Osborn A.M., Strike P. Ritchie D.A.** (1992). Amplification of DNA from native populations of soil bacteria by using the polymerase chain reaction. *Appl. Environ. Microbiol.* 58: 3413-3416.
- **Chan M.S., Maiden M.C., Spratt B.G.** (2001). Database-driven multi locus sequence typing (MLST) of bacterial pathogens. *Bioinformatics* 17: 1077-1083.
- **Chilton M.D., Drummond M.H., Merio D.J., Sciaky D., Montoya A. L., Gordon M.P., Nester E.W.** (1977). Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of Crown Gall tumorigenesis. *Cell* 11:263-267.
- **Claridge M.F., Dawah H.A. Wilson M.R.** (1997). *Species: the units of biodiversity.* Claridge M.F., Dawah H.A. and Wilson M.R.(Eds). Chapman and Hall, London.
- **Clerc A., Manceau C. Nesme X.** (1998). Comparison of randomly amplified polymorphic DNA with amplified fragment length polymorphism to assess genetic diversity and genetic relatedness within genospecies III of *Pseudomonas syringae*. *Appl. Environ. Microbiol.* 64: 1180-1187
- **Cohan F.M.** (2001). Bacterial species and speciation. *Syst. Biol.* 50: 513-524
- **Cohan F.M.** (2002). What are bacterial species? *Annu. Rev. Microbiol.* 56: 457-487
- **Conn H.J.** (1942). Validity of the genus *Alcaligenes*. *J. Bact.* 44: 353-360.
- **D'Amato E.E., Taylor R.H., Blannon J.C., Reasoner D.J.** (1991). Substrate profile systems for the identification of bacteria and yeasts by rapid and automated approaches. In "Manual of Clinical Microbiology". Balows A., Hausler W.J.J, Herrman K.L., Isenberg H.D. and Shadomy H.J. (Eds). pp:128-136. American Society for Microbiology, Washington DC.
- **D'Angelo C.** (2000). Structure génétique des populations d'*Agrobacterium*: intérêt de l'étude du gène *gyrB* et de l'AFLP. DEA Écologie Microbienne, Université Claude Bernard, Lyon 1.
- **De Cleene M, De Ley J.** (1976). The host range of of Crown Gall. *Botanical Review* 42:389-466.
- **Dessaux Y., Petit A., Farrand S.K., Murphy P.** (1998). Opines and opine-like molecules in plant-*Rhizobiaceae* interactions. In "The Rhizobiaceae: molecular biology of model plant-associated bacteria". Chapter 9:173-198. Spaink H.P., Kondorosi A., Hooykaas P. (Eds). Kluwer Academic Publishers.
- **Duim B., Vandamme P.A., Rigter A., Laevens S., Dijkstra J.R. Wagenaar J.A.** (2001). Differentiation of *Campylobacter* species by AFLP fingerprinting. *Microbiology.* 147: 2729-2737.

- **Falush D., Wirth T., Linz B., Pritchard J.K., Stephens M., Kidd M., Blaser M.J., Graham D.Y., Vacher S., Perez-Perez G.I., Yamaoka Y., Megraud F., Otto K., Reichard U., Katzowitsch E., Wang X., Achtman M., Suerbaum S.** (2003). Traces of human migrations in *Helicobacter pylori* populations. *Science*. 299:1582-1585.
- **Farrand S.K., Van Berkum P.B. Oger P.** (2003). *Agrobacterium* is a definable genus of the family *Rhizobiaceae*. *I.J.S.E.M.* 53: 1681-1687
- **Feil E.J., Maiden M.C., Achtman M. Spratt B.G.** (1999). The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*. *Mol. Biol. Evol.* 16: 1496-1502.
- **Felsenstein, J.** (2004). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
- **Gallie D.R., Hagiya M., Kado C.I.** (1985). Analysis of *Agrobacterium tumefaciens* plasmid pTiC58 replication region with a novel high copy-number derivative. *J. Bact.* 161 :1034-1041.
- **Genetello C., Van Larebeke N., Holsters M., De Picker A., Van Montagu M., Schell J.** (1977). Ti plasmids of *Agrobacterium tumefaciens* as conjugative plasmids. *Nature (London)* 265 :561-563.
- **Geornaras I., Kunene N.F., Von Holy A. Hastings J.W.** (1999). Amplified fragment length polymorphism fingerprinting of *Pseudomonas* strains from a poultry processing plant. *Appl. Environ. Microbiol.* 65: 3828-3833.
- **Gobat J.-M., Aragno M., Matthey W.** (2003). *Le sol vivant*. 2° Ed. Presses Polytechniques et Universitaires Romandes, Lausanne.
- **Grimont P.A.** (1988). Use of DNA reassociation in bacterial classification. *Can J Microbiol.* 34:541-6.
- **Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Quorllo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L., Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F., Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B., Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G., Cielo C. Slater S.** (2001). Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science*. 294: 2323-2328

- **Guan S., Xu R., Chen S., Odumeru J. Gyles C.** (2002). Development of a procedure for discriminating among *Escherichia coli* isolates from animal and human sources. *Aem.* 68: 2690-2698.
- **Guyon P., Petit A., Tempé J., Dessaux Y.** (1993). Transformed plants producing opines specifically promote growth of opine-degrading agrobacteria. *Mol. Plant Microbe Interact.* 6: 92-98.
- **Heulin T., Barakat M., Christen R., Lesourd R., Sutra L., Deluca G., Achouak W.** (2003). *Ramlibacter tataouinensis* gen. nov., sp. nov., and *Ramlibacter henchirensis* sp. nov., cyst-producing bacteria isolated from subdesert soil in Tunisia. *IJSEM*, 53:589-594.
- **Hildebrand E.M.** (1940). Cane Gall of brambles causes by *Phytomonas rubi* n. Sp. *J. Agr. Res.* 61: 685-696.
- **Hooykaas P.J.J., Beijersbergen A.G.M.** (1994). The virulence system of *Agrobacterium tumefaciens*. *Annual Reviews of Phytopathology* 32.
- **Jarraud S., Mougél C., Thioulouse J., Lina G., Meugnier H., Forey F., Nesme X., Etienne J. Vandenesch F.** (2002). Relationships between *Staphylococcus aureus* genetic background, virulence factors, agr groups (alleles), and human disease. *Infect. Immun.* 70: 631-641.
- **Israel D.A., Salama N., Krishna U., Rieger U.M., Atherton J.C., Falkow S., Peek R.M. Jr.** (2001). *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc Natl Acad Sci U S A.* 98:14625-14630.
- **Jiang S.C., Louis V., Choopun N., Sharma A., Huq A. Colwell R.R.** (2000). Genetic diversity of *Vibrio cholerae* in Chesapeake bay determined by amplified fragment length polymorphism fingerprinting. *Appl. Environ. Microbiol.* 66: 140-147
- **Jumas-Bilak E., Michaux-Charachon S., Bourg G., Ramuz M. Allardet-Servent A.** (1998). Unconventional genomic organization in the alpha subgroup of the proteobacteria. *J. Bacteriol.* 180: 2749-2755.
- **Kado C.I.** (1991). Molecular mechanisms of Crown Gall tumorigenesis. *Critical Reviews in Plant Science* 10: 1-32.
- **Kahng, L. S. & Shapiro, L.** (2001). The CcrM DNA methyltransferase of *Agrobacterium tumefaciens* is essential, and its activity is cell cycle regulated. *J Bacteriol* 183, 3065-3075.
- **Keane P.J., Kerr A., New P.B.** (1970) Crown Gall on stone fruits: identification and nomenclature of *Agrobacterium* isolates. *Aust. J. Biol. Sci.* 23:585-595.

- **Kerstens K., De Ley J.** (1984). Genus III. *Agrobacterium* Conn 1942. In "Bergey's manual of systematic bacteriology". Krieg N.R. & Holt J.G. (Eds). Williams and Wilkins, Baltimore, USA. I: 244-254.
- **Kerstens K., De Ley J., Sneath P.H.A., Sackin M.** (1973). Numerical taxonomic analysis of *Agrobacterium*. J. Gen. Microbiol. 78: 227-239.
- **Kiredjian M.** (1979). Le genre *Agrobacterium* peut-il être pathogène pour l'homme? Med. Mal. Infect. 9: 233-235.
- **Koekman B.P., Hooykaas P.J., Schilperroort R.A.** (1982). A functional map of the replicator region of the octopine Ti plasmid. Plasmid 7: 119-132.
- **Koeleman J.G.M., Stoof J., Biesmans D.J., Savelkoul P.H.M. Vanderbroucke-Grauls C.M.J.E.** (1998). Comparison of amplified ribosomal dna restriction analysis, random amplified polymorphic DNA analysis, and amplified fragment length polymorphism fingerprinting for identification of *Acinetobacter* genomic species and typing of *Acinetobacter baumannii*. J. Clin. Microb. 36: 2522-2529
- **Krimi Z., Petit A., Mougel C., Dessaux Y. Nesme X.** (2002). Seasonal fluctuations and long-term persistence of pathogenic populations of *Agrobacterium* spp. In soils. Appl. Environ. Microbiol. 68: 3358-3365
- **Kuhn I., Huys G., Coopman R., Kersters K. Janssen P.** (1997). A 4-years study of the diversity and persistence of coliforms and *Aeromonas* in the water of a swedish drinking water well. Can. J. Microbiol. 43: 9-16
- **Kuipers E.J., Israel D.A., Kusters J.G., Gerrits M.M., Weel J., Van Der Ende A., Van Der Hulst R.W., Wirth H.P., Hook-Nikanne J., Thompson S.A. Blaser M.J.** (2000). Quasispecies development of *Helicobacter pylori* observed in paired isolates obtained years apart from the same host. J. Infect. Dis. 181: 273-282
- **Levin B.R. Bergstrom C.T.** (2000). Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 97: 6981-6985.
- **Loper J.E., Kado C.I.** (1979). Host-range conferred by the virulence-specifying plasmid of *Agrobacterium tumefaciens*. J. Bact. 139:591-596.
- **Ludwig W.** (1999). The role of rRNA as a phylogenetic marker in the context of genomics. USFCC Newsl. 29: 2-6.
- **Maiden M.C., Bygraves J.A., Feil E., Morelli G., Russell J.E., Urwin R., Zhang Q., Zhou J., Zurth K., Caugant D.A., Feavers I.M., Achtman M. Spratt B.G.** (1998). Multilocus sequence typing: a portable approach to the identification of

clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. 95: 3140-3145

- **Majewski J.** (2001). Sexual isolation in bacteria. Fems. Microbiol. Lett. 199: 161-19.
- **Majewski J. Cohan F.M.** (1999). DNA sequence similarity requirements for interspecific recombination in *Bacillus*. Genetics. 153: 1525-1553
- **Majewski J., Zawadzki P., Pickerill P., Cohan F.M. Dowson C.G.** (2000). Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. J. Bacteriol. 182: 1016-1023
- **Maynard-Smith J., Smith N.H., O'rourke M. Spratt B.G.** (1993). How clonal are bacteria? Proc. Natl. Acad. Sci. U. S. A. 90: 4384-4388.
- **Mayr E.** (1942). Systematics and the origin of species from the viewpoint of a zoologist. Columbia University Press. New York.
- **Mogilner N., Zutra D., Gafny R. Bar-Joseph M.** (1993). The persistence of engineered *Agrobacterium tumefaciens* in agroinfected plants. Mol. Plant. Microbe. Interact. 6: 673-675.
- **Montoya A.L., Chilton M.D., Gordon M.P., Sciaky D., Nester E.W.** (1977). Octopine and nopaline metabolism in *Agrobacterium tumefaciens* and crown gall tumor cells: role of plasmid genes. J Bacteriol. 129: 101-107.
- **Moore L.W., Allen J.** (1986). Controlled heating of root-pruned dormant *Prunus* spp. seedlings before transplanting to prevent Crown Gall. Plant Dis. 70:532-536.
- **Moore L.W., Kado C.I. Bouzar H.** (1988). *Agrobacterium*. In laboratory guide for identification of plant pathogenic bacteria: 16-36. Nw s.c.h.a.a.d.,aps press, saint paul, minnesota.
- **Mougel C.** (2000). Structure génétique des populations d'*Agrobacterium* spp. : Effet sélectif de la plante et implication dans la diffusion conjugative du plasmide Ti.Thèse. Université Claude Bernard, Lyon 1.
- **Mougel C., Cournoyer B. Nesme X.** (2001). Novel tellurite-amended media and specific chromosomal and Ti plasmid probes for direct analysis of soil populations of *Agrobacterium* biovars 1 and 2. Appl. Environ. Microbiol. 67: 65-74
- **Mougel C., Teyssier S., D'angelo C., Groud K., Neyra M., Sidi-Boumedine K., Cloeckert A., Pelouille M., Baucheron S., Chaslus-Dancla E., Jarraud S., Meugnier H., Forey F., Vandenesch F., Lina G., Etienne J., Thioulouse J., Manceau C., Robbe P., Nalin R., Briolay J. Nesme X.** (2001). Experimental and

theoretical evaluation of typing methods based upon random amplification of genomic restriction fragments (AFLP) for bacterial populations genomics. *Genet. Sel. Evol. Suppl.* 1: s319-s338.

● **Mougel C., Thioulouse J., Perriere G. Nesme X.** (2002). A mathematical method for determining genome divergence and species delineation using AFLP. *Int. J. Syst. Evol. Microbiol.* 52: 573-586.

● **Mylvaganam S., Dennis P.P.** (1992). Sequence heterogeneity between the two genes encoding 16S r RNA from the halophilic archaeobacterium *Haloarcula marismortui*. *Genetics* 130: 399-410.

● **Nesme X.** (2001). Integrated Control of Crown Gall in Mediterranean Countries (CGMED). Final Report. European Union Contract number : ERBIC18CT970198.

● **Nesme X., Michel M.F., Digat B.** (1987). Population heterogeneity of *Agrobacterium tumefaciens* in galls of *Populus* L. from a single nursery. *Appl. Env. Microb.* 53: 655-659.

● **New P.B., Kerr A.** (1972). Biological control of Crown Gall: field measurements and glasshouse experiments. *J. Appl. Bacteriol.* 34:233-236.

● **Ochman H., Lawrence J.G. Groisman E.A.** (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature.* 405: 299-304

● **Ophel K. Kerr A.** (1990). *Agrobacterium vitis* sp. Nov. For strains of *Agrobacterium* biovar3 from grapevines. *IJSB.* 40: 236-241

● **Owen R.J., Ferrus M. Gibson J.** (2001). Amplified fragment length polymorphism genotyping of metronidazole-resistant *Helicobacter pylori* infecting dyspeptics in England. *Clin. Microbiol. Infect.* 7: 244-253

● **Palys T., Nakamura L.K. Cohan F.M.** (1997). Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. *Int. J. Syst. Bacteriol.* 47: 1145-1156.

● **Petit A., Delhaye S., Tempé J. Morel G.** (1970). Recherche sur les guanidines des tissus de Crown Gall. Mise en évidence d'une relation biochimique spécifique entre les souches d'*Agrobacterium* et les tumeurs qu'elles induisent. *Physiologie Végétale* 8: 205-213.

● **Petit A., Tempé J.** (1978). Isolation of *Agrobacterium* Ti-plasmid regulatory mutants. *Mol. Gen. Genet.* 167:147-155.

- **Picard C., Ponsonnet C., Paget E., Nesme X., Simonet P.** (1992). Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. *Appl. Env. Microb.* 58: 2717-2722.
- **Pionnat S., Keller H., Hericher D., Bettachini A., Dessaux Y., Nesme X. Poncet C.** (1999a). Ti plasmids from *Agrobacterium* characterize rootstock clones that initiated a spread of Crown Gall disease in mediterranean countries. *Appl. Environ. Microbiol.* 65: 4197-4206.
- **Pionnat S., Keller H., Hérliche D., Bettachini A., Dessaux Y., Nesme X., Poncet C.** (1999b). Ti plasmids from *Agrobacterium* characterize rootstock clones that initiated an outspread of crown gall disease in Mediterranean countries. *Appl. Environ. Microbiol.* 65:4197-4206.
- **Pirnay J.P., De Vos D., Cochez C., Bilocq F., Pirson J., Struelens M., Duinslaeger L., Cornelis P., Zizi M., Vanderkelen A.** (2003). Molecular epidemiology of *Pseudomonas aeruginosa* in a burn unit: persistence of multidrug-resistant clone and a silver sulfadiazine-resistant clone. *J. Clin. Microb.* 41: 1192-1202.
- **Ponsonnet C.** (1994). Structure génétique des populations d'*Agrobacterium*. Apport à l'écologie des plasmides Ti. Thèse. Université Claude Bernard, Lyon 1.
- **Ponsonnet C., Nesme X.** (1994). Identification of *Agrobacterium* strains by PCR-RFLP analysis of pTi and Chromosomal regions. *Arch. Microbiol.* 161: 300-309.
- **Popoff M.Y., Kersters K., Kiredjian M., Miras I. Coynault C.** (1984). Taxonomic position of *Agrobacterium* strains of hospital origin. *Ann. Microbiol. (paris).* 135a: 427-442.
- **Puvanesarajah V., Schell F.M., Stacey G., Douglas C.J., Nester E.W.** (1985). Role for 2-linked-β-D-glucan in the virulence of *Agrobacterium tumefaciens*. *J. Bact.* 164: 102-106.
- **Raio A., Zoina A., Moore L.W.** (1997). The effect of solar heating of soil on natural and inoculated agrobacteria. *Plant Pathol.* 46:320-328.
- **Rayssiguier C., Thaler D.S. Radman M.** (1989). The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature.* 342: 396-401.
- **Reisenauer A., Kahng L.S., McCollum S., Shapiro L.** (1999). Bacterial DNA methylation: a cell cycle regulator? *J Bacteriol* 181 :5135-5139.

- **Riker A.J., Banfield W.M., Wright W.H., Keitt G.W. Sagen H.E.** (1930). Studies on infectious hairy root of nursery apple trees. *J. Agr. Res.* 41: 507-540.
- **Roberts, M. S., And F. M. Cohan.** (1993). The effect of DNA sequencedivergence on sexual isolation in *Bacillus*. *Genetics* 134:401–408.
- **Salama N., Guillemin K., Mcdaniel T.K., Sherlock G., Tompkins L. Falkow S.** (2000). A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc. Natl. Acad. Sci. U. S. A.* 97: 14668-14673
- **Sawada H., Ieki H., Oyaizu H. Matsumoto S.** (1993). Proposal for rejection of *Agrobacterium tumefaciens* and revised descriptions for the genus *Agrobacterium* and for *Agrobacterium radiobacter* and *Agrobacterium rhizogenes*. *Int. J. Syst. Bacteriol.* 43: 694-702.
- **Schena M., Heller R.A., Theriault T.P., Konrad K., Lachenmeier E., Davis R.W.** (1998). Microarrays: biotechnology's discovery platform for functional genomics. *Trends. Biotechnol.* 16: 301-306.
- **Schleifer K.H., Stackebrandt E.** (1983). Molecular systematics of prokaryotes. *Annu Rev Microbiol.* 37:143-87
- **Schneider S., Roessli D., Excoffier L.** (2000). Arlequin ver 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- **Shen P. Huang H.V.** (1986). Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. *Genetics.* 112: 441-457.
- **Skerman V.S.D., McGowan V., Sneath P.H.A.** (1980). Approuved list of bacterial names. *IJSB* 30: 225-420.
- **Smith E.F. Townsend C.O.** (1907). A plant-tumor of bacterial origin. *Science (new york).* 25: 671-673.
- **Sneath P.H., Sokal R.R.** (1962). Numerical taxonomy. *Nature* 193: 855-860.
- **Stachel S.E., Messens E., Van Montagu M, Zambrisky P.** (1985). Identification of the signal molecules produced by wounded plants cells that activate T-DNA transfer in *Agrobacterium tumefaciens*. *Phytochemistry* 27:2781-2785.
- **Stackebrandt E., Frederiksen W., Garrity G.M., Grimont P.A., Kämpfer P., Maiden M.C., Nesme X., Rossello-Mora R., Swings J., Truper H.G., Vauterin L., Ward A.C. Whitman W.B.** (2002). Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* 52: 1043-1047



- **Stackebrandt E. Goebel B.M.** (1994). Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.* 44: 846-849
- **Starr M.P. Weiss J.E.** (1943). Growth of phytopathogenic bacteria in a synthetic asparagin medium. *Phytopathology.* 33: 314-318.
- **Sugimoto N., Nakano M. and Nakano S.I.** (2000). Thermodynamics-Structure Relationship of Single Mismatches in RNA/DNA Duplexes. *Biochemistry* 39 : 11270-11281.
- **Tenover F.C., Arbeit R.D., Goering R.V., Mickelsen P.A., Murray B.E., Persing D.H., Swaminathan B.** (1995). Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J. Clin. Microb.* 33: 2233-2239.
- **Teyssier-Cuvelle S., Mougel C. Nesme X.** (1999). Direct conjugal transfers of Ti plasmid to soil microflora. *Mol. Ecol.* 8: 1273-1284.
- **Teyssier-Cuvelle S., Oger P., Mougel C., Groud K., Farrand S., Nesme X.** (2004). A highly selectable and highly transferable Ti plasmid to study conjugal host range and Ti plasmid dissemination in complex ecosystems. *Microbial Ecology* 48:10-18.
- **Thomashow M.F., Nutter R., Montoya A.L., Gordon M.P., Nester E. W.** (1980). Integration and organization of Ti plasmid sequence in Crown Gall tumors. *Cell* 19:729-739.
- **Torsvik V., Goksoyr J., Daae F.L.** (1990). High diversity in DNA of soil bacteria. *Appl. Environ. Microbiol.* 56:782-787.
- **Urbanczyk H., Suzuki K., Yoshida K. Kondo K.** (2003). Physical and gene maps of *Agrobacterium* biovar 2 strains and their relationship to biovar 1 chromosomes. *Microbiology.* 149: 3035-3042.
- **Urwin R. Maiden M.C.J.** (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends in microbiology.* 11: 479-487
- **Valentine I.** (2003). *Agrobacterium tumefaciens* and the plant: the David and Goliath of modern genetics (update on *Agrobacterium*-mediated transformation of plants). *Plant physiology.* 133: 948-955
- **Vancanneyt M., Lombardi A., Andrighetto C., Knijff E., Torriani S., Bjorkroth K.J., Franz C.M.A.P., Foulquie Moreno M.R., Revets H., Devuyst L., Swings J., Kersters K., Dellagio F. Holzapfel W.H.** (2002). Intraspecies genomic groups in

*Enterococcus faecium* and their correlation with origin and pathogenicity. Appl. Env. Microb. 68: 1381-1391.

- **Van Larebeke N, Engler G., Holsters M., Van den Elsacker S., Zaenen I., Schilperrort R.A., Schell J.** (1974). Large plasmid in *Agrobacterium tumefaciens* essential for Crown Gall-inducing ability. Nature 252:169-170.
- **Van Valen, L.** (1976). Ecological species, multispecies, and oaks. Taxon 25:233–239.
- **Vogel J., Normand P., Thioulouse J., Nesme X. Grundmann G.L.** (2003). Relationship between spatial and genetic distance in *Agrobacterium* spp. In 1 cubic centimeter of soil. Appl. Environ. Microbiol. 69: 1482-1487.
- **Volff J.N., Altenbuchner J.** (1998). Genetic instability of the *Streptomyces* chromosome. Mol. Microbiol. 27: 239-246.
- **Vos P., Hogers R., Bleeker M., Reijans M., Van De Lee T., Hornes M., Frijters A., Pot J., Peleman J., Kuiper M.** (1995). AFLP: a new technique for DNA fingerprinting. Nucleic acids res. 23: 4407-4414
- **Vulic M., Dionisio F., Taddei F. Radman M.** (1997). Molecular keys to speciation: DNA polymorphism and the control of genetic exchange in enterobacteria. PNAS. 94: 9763-9767
- **Wayne L.G., Brenner D.J., Colwell R.R., Grimont P.A.D., Kandler O., Krichevsky M.I., Moore L.H., Moore W.E.C., Murray R.G.E., Stackebrandt E., Starr M.P. Truper H.G.** (1987). Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. IJSB. 37: 463-464
- **Whitman W.B., Coleman D.C. Wiebe W.J.** (1998). Prokaryotes: the unseen majority. Proc. Natl. Acad. Sci. U. S. A. 95: 6578-6583.
- **Woese C.R.** (1987). Bacterial evolution. Microb. Revs. 51: 221-271
- **Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P., Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida Jr N.F., Woo L., Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee Sr D., Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C., Kutuyavin T., Levy R., Li M.J., McClelland E., Palmieri A., Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D., Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M., Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.Y., Dolan M., Chumley F., Tingey S.V., Tomb J.F. Milton P.G., Maynard V.O. And Nester W.E.** (2001). The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. Science. 294: 2317-2323

- **Young J.M., Kuykendall L.D., Martinez-Romero E., Kerr A. Sawada H. (2001).** A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* conn 1942 and *Allorhizobium undicola* De Lajudie *et al.* 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*. IJSEM. 51: 89-103
- **Young J.M., Kuykendall L.D., Martinez-Romero E., Kerr A. Sawada H. (2003).** Classification and nomenclature of *Agrobacterium* and *Rhizobium* - a reply to farrand *et al.* (2003). IJSEM. 53: 1689-1695.
- **Zambrisky P.C. (1992).** Chronicles from the *Agrobacterium* plant-cell DNA transfer story. Annu. Rev. Plant Physiol. Plant Mol. Biol. 43:465-490.
- **Zhu J., Oger P.M., Schrammeijer B., Hooykaas P.J., Farrand S.K. Winans S.C. (2000).** The bases of Crown Gall tumorigenesis. J. Bacteriol. 182: 3885-3895.
- **Zhu P., Van der Ende A., Falush D., Brieske N., Morelli G., Linz B., Popovic T., Schuurman I.G., Adegbola R.A., Zurth K., Gagneux S., Platonov AE., Riou J.Y., Caugant D.A., Nicolas P., Achtman M. (2001).** Fit genotypes and escape variants of subgroup III *Neisseria meningitidis* during three pandemics of epidemic meningitis. Proc. Natl. Acad. Sci. USA. 98: 5234-5239

---

Sélection d'écotypes bactériens pathogènes et non-pathogènes par la plante en relation avec la différenciation en espèces génomiques chez *Agrobacterium* spp.

Ce travail avait pour but : 1) de vérifier l'existence d'une relation entre espèces génomiques et écotypes associés aux plantes chez *Agrobacterium* spp.; 2) d'identifier les zones génomiques impliquées dans la différenciation en écotypes et, partant, en espèces génomiques dans ce taxon. Nous avons d'abord cherché et caractérisé au niveau génomique (par AFLP) des écotypes non-pathogènes et pathogènes (hébergeant un plasmide Ti), dans différents biotopes. Nos résultats montrent que la plante sélectionne des écotypes spécifiques au sein de la communauté des agrobactéries du sol. De plus, il apparaît que c'est parmi les écotypes spécifiques d'une plante donnée que sont recrutées certaines agrobactéries pathogènes isolées des tumeurs. L'AFLP prédictive réalisée sur le génome de la souche C58 nous a ensuite permis d'identifier les fragments AFLP caractéristiques de l'espèce G8. Les gènes et fonctions ainsi révélées pourraient effectivement concerner des relations plantes-bactéries.

---

Plant selection of pathogenic and non-pathogenic bacterial ecotypes linked with genomic species differentiation in *Agrobacterium* spp.

The aims of this work were : 1) to verify the existence of relations between genomic species and plant-associated ecotypes in *Agrobacterium* genus ; 2) to determine which genomic zones are implied in the ecotype differentiation and thus in the genomic species differentiation in this taxa. Firstly, we determined and characterised, at the genomic level (by AFLP), non-pathogenic and pathogenic (harbouring a Ti plasmid) ecotypes in different biotopes. Our results show that the plant select specific ecotypes among the agrobacterial soil community. Moreover, some pathogenic strains, isolated from crown galls, are recruited among these particular ecotypes. Secondly, predictive AFLP performed on strain C58 genome, permitted us to identify the AFLP fragments characteristic of genomic species G8. The genes and functions revealed this way could actually be implied in plant-bacteria relationships.